A theoretical study of charge confinement in quantum dots: Modelling the SnO2 charge writing process.

Williams, Owen Leyton

A Theoretical Study of Charge Confinement in Quantum Dots: Modelling the SnO_2 Charge Writing Process

By

Owen Leyton Williams

Submitted to the University of Wales in fulfilment of the requirements for the Degree of Doctor of Philosophy

University of Wales, Swansea

August 2007

Preface

Declarations

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed

O. L. Williams 30/08/07

This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by footnotes giving explicit references and a bibliography is appended.

Signed

O. L. Williams 30/08/07

I hereby give consent for my thesis, if accepted, to be available for photocopying and inter-library loan, and for the title and summary to be made available to outside organisations.

Signed

O. L. Williams 30/08/07

Acknowledgements

I would like to thank both my supervisor, Professor Steve Wilks, for his encouragement and guidance, and Professor Paul Rees for his advice during the course of this work.

I am greatly indebted to all my family for all their support and tolerance. To my parents and sister – June, Chris and Helen – thank you for all the baby-sitting; to my children - Christopher, Tom and Sam - I thank you for many an interesting diversion, endless good cheer and those essential Lego breaks; and finally and most importantly, I would like to thank my wife, Laura, for her unfaltering patience, belief and tea, and without whom this would not have been possible.

Abstract

A suite of models is constructed to facilitate the simulation of the SnO_2 charge writing process. In particular, at dimensions where the semiconductor band bending does not fully evolve, this entails the self-consistent solution of the non-linear Poisson equation and the Kohn-Sham equations at non-zero temperature, with the charge in the occupied surface states also self-consistently reconciled with the fundamental electron density generating the confining potential. In this way, a full quantum mechanical treatment of the discrete eigenstates of the quantum dot, inclusive of electron-electron effects, is made, and a Tip-QD-Substrate tunnelling model developed.

This work favourably conforms with observed experimental measurements, not only satisfying the recorded data on the ratios of surface state densities far better than existing models, but also offers a tentative explanation for some of the hitherto unsatisfactorily explained sensitivity behaviour of polycrystalline gas sensors on the decrease of the grain radii.

It models the charging of a spherical 4nm radius nanocrystal well, with the calculated *I-V* characteristic clearly exhibiting indications of the Coulomb blockade effect in good agreement with experiment. The calculated maximum electron complement of one nanocrystal of between 81 and 87 injected electrons with a modal potential difference interval between charge transfer events of 0.065V, is in excellent concordance with the experimentally inferred population of 86 elections, charge storage events occurring at intervals of 0.07V.

Contents

Chapter 1 Introduction

There is much interest currently in the properties and behaviour of nanoscaled semiconductors, as the reduction of their physical dimensions can correspond to significant alteration in their electronic properties and the manifestation of unusual and exciting characteristics. The smaller the particle, the greater the surface area to volume ratio, and the more biologically and chemically reactive the particle becomes. The better its solubility, the greater its ability to penetrate membranes, and the larger its capacity to permeate through media. The force of gravity becomes less important with its position replaced by surface tension and electrostatics. Most importantly of all, as the size of the particle decreases, the physics governing its nature moves away from the classical rules of the macroscopic world, and into the bizarre and wonderful quantum realm and all the amazing phenomena that this entails.

New developments in this emerging field of nanotechnology have enabled the use of charge writing to pattern substrates, where charge from the tip of a scanning tunnelling microscope (STM) is transferred into the discrete energy levels of a quantum dot, formed by the confinement of a semiconductor in all three spatial dimensions.

At the University of Wales Swansea's Multidisciplinary Nanotechnology Centre (MNC), the Semiconductor Interface Laboratory (SIL) have taken a SnO_2 nanocrystalline film, a surface traditionally used in gas sensing, and used it to successfully demonstrate charge writing, see Figure 1-1, electrons being injected into the quantised eigenstates of the individual quasi-spherical grains, the stored charge remaining stable for many weeks.

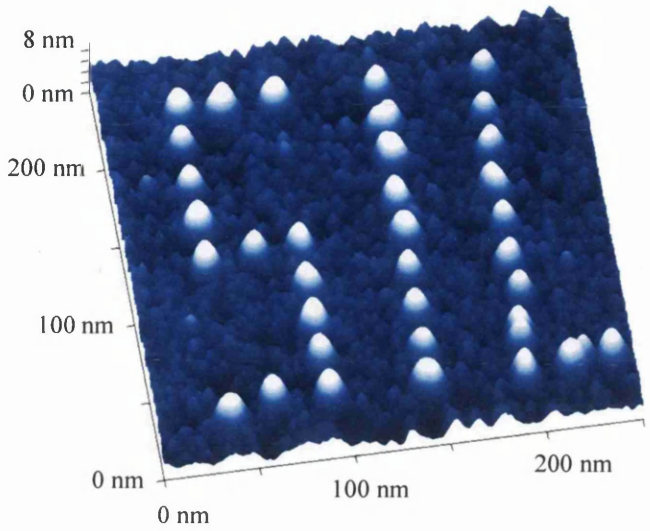


Figure 1-1 250nm^2 STM scan of a nanopatterned SnO_2 polycrystalline surface by the Semiconductor Interface Laboratory (SIL) of the Multidisciplinary Nanotechnology Centre (MNC). Charge points average 15nm in diameter and 8nm in height and were formed by biasing an STM tip for 100μs at -6V.

This process has many implications across many fields; from the nanoelectronic – for instance data storage (both digital and analogue) and computing with quantum cellular automata (QCA), to the biomedical, with the possibility of nanoscale catalysis and molecular docking.

While some progress has been made through the experimental research in understanding this new and novel process, the comprehension of the underlying physics behind this charge writing and electron storage phenomenon is comparatively in its infancy.

This theoretical work develops a suite of models to facilitate the simulation of this procedure in order to obtain a clearer understanding of the physical process. In its most complex form, at dimensions where the semiconductor band bending does not fully evolve, this involves the self-consistent solution of the non-linear Poisson equation and the Kohn-Sham equations, with inclusion of the effects of the surface states also self-consistently reconciled with the fundamental electron density generating the confining potential. These theoretical models allow the electronic structure of the spherical quantum dots to be assessed in some detail, and permit preliminary simulations of the charging process itself, illustrating clearly the effects of Coulomb blockade and single electron charging.

The theoretical simulations match pleasingly with the observed experimental measurements, and perhaps most importantly, offer a surprising hypothesis regarding the behaviour of the ionised donor density at small dimensions. This proposal not only satisfies the observed data of the ratios of surface state densities far better than existing models, but also offers a tentative explanation for some of the hitherto unsatisfactorily explained sensitivity behaviour of polycrystalline gas sensors on the decrease of the grain radii.

1.1 Introduction to Charge Writing

In the constant pursuit of technological advancement, recent decades have seen the continuing decrease in size of mircoelectrical components in an effort to maximise processing power. In latter years, this physical diminution of the constituent semiconductors to the nanoscale has lead to significant alterations of their electronic properties, and a drift away from their bulk characteristics. Spatial confinement of the semiconductor gives rise to discrete energy levels and allowed electron states. On each reduction of dimensionality, the energy and state density becomes more precisely defined, giving rise to improved charge carrier transport and optical properties. This realisation of practical semiconductor quantisation heralds a revolution in solid-state physics.

The study of matter in this nanometer regime, the field of nanotechnology, shadows under its auspices many new and exciting areas of development, encompassing not only the science of electronics, but with far reaching implications for the disciplines of engineering and medicine also.

The key to progress is felt by many [1,2,3] to lie in the self-assembly of nanoscale devices and structures, with the critical restriction on development, the difficulties in selectively positioning and orientating objects reproducibly on the nanoscale. Already DNA has been used to create self-assembling periodic nanostructures [4] and self-assembled molecules called rotaxanes have been manufactured which are able to flip between two stable states and have potential as switches in future molecular based computers [1]. Charge writing offers the possibility of efficient and programmable self-assembly [5] by patterning a surface with localised charge to act as 'docking' sites for particular polarised particles. This pattern, repeatable to a high degree of accuracy, might allow the precision orientation of the desired nano-objects, enabling the repetitive fabrication of complex structures.

Work conducted by the Multidisciplinary Nanotechnology Centre at Swansea University, has demonstrated this charge writing on a polycrystalline SnO_2 surface composed of quasi-spherical

grains of radius 4nm, see for example Figure 1-1. Electrons were injected into the nanocrystals[i] using the tip of a scanning tunnelling microscope (STM) under ultra-high vacuum conditions with a spatial resolution of 15nm [7]. In Figure 1-1, the STM tip was biased at -6V for 100μs for each writing event, the resultant charged point encompassing two to three of the 4nm radii grains, each 'cluster' protruding around 8nm proud of the surface. Stored in a vacuum for more than three weeks, the confined electrons still remain localised in the charge injected sites.

The potential of this technique for the future is considerable.

Bio-chemically, the patterning of the surface could be used to orientate polar species and catalyse chemical reactions, or even provide a means of self-assembly as discussed above.

Electronically, digital data storage devices can be envisaged, sites charged or uncharged corresponding to the binary '1' or '0' states. Through this, computing with quantum cellular automata (QCA) [8,9,10] would also be feasible, digital logic functions performed by arrays of quantum dot cells. Analogue data storage may even be possible exploiting the Coulomb staircase of the charging events; this Coulomb blockade has been shown by several groups [11,12], including on these R = 4nm SnO_2 nanocrystals by the MNC's charge writing experimental team [6].

Whatever the application, it is clear that the smaller the resolution of the nanopatterning technique, the more efficient and versatile the application. This resolution depends on both the nature of the tip and the size of the charge confining nanocrystal. Whilst charge writing for selective absorption has been achieved to a degree by Mesquida *et al.* [13], creating nucleation sites for silica nanocrystals, the spatial resolution was of the order of 1μm. Nanocrystalline SnO_2 films on the other hand, besides being both comparatively cheap and simple to manufacture, offer a much smaller spatial resolution; the 4nm radii grains of Figure 1-1 reducible in principle to grains of radii 1nm. Indeed, it appears that the MNC is the only group to be researching the nanopatterning of nanocrystalline SnO_2 films through STM [14].

---

[i] Surface modification by the tip, such as material deposition, has been ruled out, along with the possibility of the charge being stored in the oxide layer of the silicone substrate [6], leaving the only feasible alternative that of charge storage within the discrete energy states. This is supported by the height dependence of the charged point on the bias of the STM tip and the ability to erase after writing the observed features with a positively biased STM tip [7].

The task of modelling this charge writing process and the details of the electron confinement are non-trivial. SnO_2 has received some attention in the literature due to its use in the field of gas sensing, and some work has been done [15,16,17] on the nature of the band bending caused by the depletion of electrons from the conduction band through the formation of surface states. Of particular relevance is the behaviour of the semiconductor when the grains become too small to possess a distinct region un-depleted of charge carriers. In this case, the curvature of the conduction (and valence) band flattens and the depth of the potential well, formed between the conduction band at the surface and at the centre of the grain, diminishes. To model this, Poisson's equation must be solved, including not only the effects of the ionised donor vacancies within the charge density but the, often-neglected, mobile charge carriers. In this way Poisson's equation is non-linear and requires a numerical solution.

The limitations of this existing model from literature are considerable for the smaller dimensions of grain as no inclusion is made of the effects of quantisation. The discrete energy spectrum is not calculated, and so, consequentially, the potential is not self-consistent with any eigenstates that would be generated by it. Furthermore, it is proposed within this work that when the spatial dimensions are such that a region un-depleted of electrons does not develop at the grain centre, then the two boundary values employed in these papers are no longer automatically consistent with each other and the remainder of the parameters defining the un-depleted system. The consequences of this appear to be quite profound.

Naturally, there are further works to be found in the literature that offer some, but not all, of the desired building blocks of the charge writing model on other materials and in other symmetries.

For instance, in Martí *et al.* [18], the use of R = 3.9nm $In_{0.58}Ga_{0.42}As$ spherical quantum dots within room temperature solar cells is discussed. The authors present both analytical and numerical models of the potential, but although a self-consistent solution of the non-linear Poisson equation with the Schrödinger equation is mentioned, they use the spherical Bessel function (see Section 3.3) of a square potential well to approximate the wave functions. The dots have one energy level only, at which the Fermi level is set; although the occupancy of this level is determined with Fermi-Dirac statistics. An analytical charge balancing type calculation is used to assess the doping of the $Al_{0.4}Ga_{0.6}As$ in which the QDs are imbedded, though considerable approximations are made. No inclusion is made of electron-electron interation forces or of interface (surface) states.

For 2-dimensional circular quantum dots of radii ~12.5nm to 300nm, Mucucci *et al.* [19] investigates the capacitance and behaviour of the chemical potential for discrete energy levels, inclusive of electron-electron effects including use of the KLI-approximation for exchange favoured in this work. However, the applicability of this publication to this treatise is unfortunately limited. Without surface states and assuming integer occupancy of its energy levels (and so independent of the chemical potential), with its basic confining potential formed from the assumption of a uniform background positive charge, its methods are more suggestive than directly beneficial.

Torsti *et al.* [20] offers a more immediately useful model, applied to cylindrically symmetric Na quantum dots formed between a monolayer of Na and a Cu substrate at 1200 K. Although the basic confining potential is again formed from a uniform positive background charge, the discrete energy levels are calculated self-consistently with regard to the electron-electron interaction effects through the Kohn-Sham equations. The fractional occupancy of the energy levels is calculated with the Fermi-Dirac statistics employed here and surface states are discussed with regard to their effects in the monolayer and substrate. Electron-electron effects are however only considered through the local density approximation (LDA).

The theoretical models developed in this work all have the same aim of minimising the complexity of each individual approximation whilst still retaining meaning in its results. In simulating the band bending behaviour, at their simplest, the depletion approximation of literature [15] is used for large grains and has an analytical solution. At their most complex, the Poisson-Kohn-Sham-Charge Balance (P-KS-CB) method developed herein, applied to small dimensions where the band bending is not fully formed, self-consistently reconciles the non-linear Poisson equation (and so including the influence of the mobile charge carriers) and the Kohn-Sham equations. The exchange effects are approximated with the KLI potential [21] and the correlation potential with an LDA method [22]. Calculated at elevated temperatures, the fractional occupancy of the discrete energy levels is given by Fermi-Dirac statistics.

Most importantly, and to the knowledge of the author not employed anywhere else, this P-KS-CB method incorporates a procedure to ensure full consistency between the calculated surface state densities and the self-consistent charge density. This ensures that the two non-linear Poisson equation boundary values are always consistent with each other and the remainder of the defining

parameters. It also ensures that the energy minimisation procedure behind the Kohn-Sham method retains its meaning.

The Kohn-Sham method maps the interacting many bodied problem onto an auxiliary single-particle problem. It is based on the variational principle and minimises the total energy of the system with respect to the electron density and the constraint of constant particle number. In this way the electron density, or ground state density, is that of the many-bodied system. Strictly, the single particle eigenstates have no physical meaning and only form part of the mathematical construction devised to yield the correct density. Yet they can provide good approximations to physical systems (see for example [23,24,25,26,27]), and have mathematically been shown to be the excitation energies to zeroth order in the electron-electron interaction [28]. Consequentially, they are used as approximations of the discrete energy levels of the system throughout this work[ii]. Now, some authors (see for example Kohnanoff [29]), feel that fractional occupancies of the energy levels cannot be used within the variational scheme, contrary to what is implied by the work of Ref.'s [21, 30] and in particular Torsti *et al.* [20]. For fractional occupancy dependent on the electron density (for example (indirectly) through the Fermi-Dirac distribution), one obvious problem is that the 'constraint' used to minimise the total energy functional is now itself a function of the electron density. It is suggested in this work that the extra self-consistency step, which ensures that the two non-linear Poisson equation boundary conditions are consistent with each other and the remaining defining parameters, also implies that the Kohn-Sham minimising electron density is the 'true' electron density of the many bodied system, despite its density dependent constraint.

This P-KS-CB methodology offers a rather surprising prediction regarding the behaviour of the ionised donor density as the radius of the nanometric grains diminishes. Interestingly, this appears not only to satisfy the experimentally measured surface state density ratio of 4nm to 15nm radii grains [31] considerably better than the existing models, but seems also to offer the beginnings of an explanation for some of the hitherto unsatisfactorily explained sensitivity behaviour of polycrystalline gas sensors (see Sections 2.4 and 6.1.3).

[ii] Chapter 7 discusses a method of accounting for the electron self-energy using Green's functions, providing a means to evaluate the true electron addition and removal energies of the many body system for future work.

At a radius of 4nm the charge written nanocrystals of Wilks *et al.* [6] require the full P-KS-CB method to simulate their electron structure. Representing the STM-nanocrystalline film-substrate system as two tunnelling junctions and treating the charged grain in isolation (i.e. not considering charge leakage into neighbouring grains), then through a consideration of the free energy changes [32] with the tunnelling rate estimated through a Fermi Golden Rule approach [33], the P-KS-CB data can be used to simulate the charge storage with the grains. The models generate the characteristic Coulomb staircase charging pattern, and very favourably calculate the maximum complement of injected electrons and the voltage interval between charging events in comparison to experiment.

This treatise is novel in the contribution it makes to address the experimental SnO_2 data from this new and exciting field of charge writing. Its work is original in its effort to combine the confining potential generated from the non-linear Poisson equation at non-zero temperature self-consistently with the Kohn-Sham equations also balancing the charge on the grain to ensure the consistency of the occupied surface states and charge density.

1.2 Introduction to Chapters

Chapter 1 is a chapter of introductions. It introduces the field of nanotechnology and discusses the importance of charge writing. It offers a brief introduction to the aims of the models developed herein and an indication of the worth of their results. It discusses some of the pertinent solid-state physics basics necessary to an understanding of the problem and introduces the effects of nanoscale confinement and the quantisation of the energy spectrum.

Chapter 2 discusses the nature of surface states and their effects on the conduction and valence bands. It introduces the spherically symmetric, non-linear, complete charge density model of the literature, and develops the numerical techniques necessary to solve this form of Poisson's equation. These methods are then tested and compared against published SnO_2 gas sensing work and surface state density data. Depletion widths and the effects on the conduction band if the grain is too small to possess a non-depleted region are discussed. Conduction through gas sensing films is also covered and the effectiveness of the complete charge density model at small radii is considered.

Chapter 3 considers the discrete eigenstates formed within the potential well of the SnO_2 nanocrystals in the absence of electron-electron effects. It introduces the basic quantum mechanics necessary to understand the forms of the angular and radial components of the wave functions. It develops finite difference numerical techniques to ascertain the orthonormal eigenstates of an arbitrary radial potential, and evaluates their accuracy both by comparison with the analytical solutions of the spherically symmetric square potential well, and by comparison with an alternative tight-binding approach. In addition, this Chapter also introduces the concept of self-consistency and develops an iterative scheme to achieve this between the solutions of the non-linear Poisson equation and Schrödinger's equation.

Chapter 4 is devoted to the effects and modelling of electron-electron interactions. It discusses how the single-particle techniques of the previous Chapter can be adapted to simulate the true multi-particle system. Through simpler approaches, density functional theory (DFT) is introduced, and the Kohn-Sham method, where the single-particle eigenstates minimise the energy of the full many-body system, is developed. Some DFT approaches to the effects of

electron exchange and correlation are briefly covered, with the main focus on a local density approximation (LDA) to correlation, and the Krieger, Li and Iafrate (KLI) approximation to the exact exchange. The numerical techniques necessary to simulate these last two potentials and the Coulomb potential for the spherically symmetric problem are also developed.

Chapter 5 is the most thoughtful of the Chapters. It considers the issues involved in determining a unique potential, or rather unique for a given methodology, to satisfy the non-linear Poisson equation. It proposes that for small radii it is not necessarily certain that the same values of the defining parameters n_d, E_f, T, Q, m^* and ε corresponding to the barrier height S_b for large radii also correspond to a value of zero for the second Poisson equation boundary value, the derivative of the potential at the grain centre. It suggests that in this way the uniqueness of the non-linear Poisson potential for a given method, or form of electron density, is ensured. Significantly, it is suggested that if the surface barrier height is to be kept constant along with E_f, T, Q, m^*, and ε over a range of grain radii, then the ionised donor density cannot remain constant. It offers some validation of this theory, mainly from discussions of published work and indications of discrepancy arising therein through not considering this issue. A numerical technique is described and implemented to remove this issue of non-consistency. It is demonstrated how this method can be used to determine the movement of the Fermi level on the addition of electrons to a quantum dot at non-zero temperatures, where the fractional occupancy of the discrete energy levels is a function of the Fermi level and the charge density.

Chapter 6 applies the models and numerical techniques evolved over the preceding Chapters to modelling the SnO_2 grains as used in the experimental systems. Firstly, from the measured values of the conduction band on the grain surface, the ionised donor density of the 4nm radius nanocrystals are determined, their surface density compared with 15nm radius grains in excellent agreement with that experimentally measured. The calculated bulk ionised donor density and surface state density are also in qualitative agreement with that generally found. This ionised donor density behaviour is then used to offer a tentative explanation for the observed sensitivity increase of polycrystalline gas sensing films particularly below R = 10nm and R = 3nm. Secondly, this 4nm n_d is used to model the movement of the Fermi level with the incremental increase of charge stored within the grain, this data then used in a tunnelling model to simulate the charge writing process. Compared with experiment, favourable results are again produced with respect to the maximum quantity of stored electrons, current magnitude, observable Coulomb staircase characteristic, and the voltage interval between charging events.

Chapter 7 provides a summary of the models used and developed over the previous Chapters and presents a synopsis of their results. It discusses the validity of the work and the main avenues to be pursued for future development of the simulations. It focuses on quasiparticle methods, surrounding the individual electrons with a positive polarisation cloud and so offering an improved assessment of the true many-electron eigenstates than the Kohn-Sham single-particle eigenstates alone, and discusses the simulation of the quantum dot charging and the tunnelling current. It proposes several routes through which the sophistication of these calculations can be improved, and briefly mentions the challenges involved in taking into account the effects of the neighbouring grains and the sinter neck connections.

1.3 Introduction to Solid State Physics

To achieve a good understanding of semiconductors, it is essential to consider the three main forms of solid matter – metal, semiconductor, and insulator – even in a brief synopsis of this nature. It is sensible then to begin with a basic introduction to metals.

1.3.1 Free Electron Model

One of the simplest models of metals is that of Sommerfeld [34]. Here the solid metal is seen as a lattice, or crystal, made up of the nuclei and core electrons of the constituent atoms, stripped of their valence electrons. These free 'valence' electrons form the conduction electrons of the metal, and are free to move through the whole structure formed by the lattice of ions. This electron sea, or free Fermi gas, is treated with Fermi-Dirac statistics; an energy distribution, which, unlike its classical counterpart, the Maxwell-Boltzmann distribution, accounts for the Pauli exclusion principle, precluding identical fermions (particles with half-integer spin e.g. electrons) from occupying the same point in space.

The Fermi-Dirac distribution, illustrated in Figure 1-2 and given by

$$f(E) = \frac{1}{1 + e^{(E-\mu)/k_B T}} \tag{1.1}$$

yields the probability that an orbital of energy E will be occupied in an ideal electron gas in thermal equilibrium. μ is called the chemical potential and is defined as the point at which $f(E) = 0.5$ for $E = \mu$.

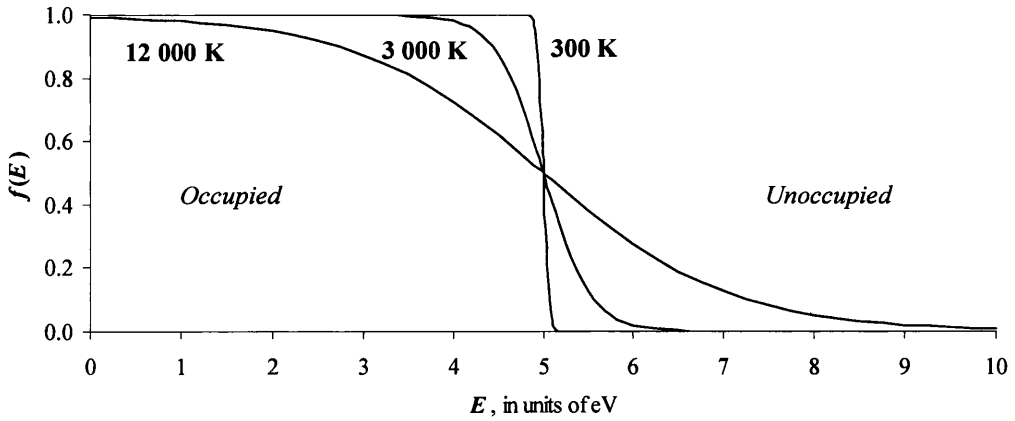


Figure 1-2 Fermi-Dirac distribution function at 300 K, 3 000 K, and 12 000 K. The chemical potential is set at 5eV.

At absolute zero, the chemical potential is equal to the Fermi energy or Fermi level, E_f, defined as the energy of the highest occupied orbital at $T = 0$ K. However, for the rest of this work, the convention of Blakemore [35] and much of the work of literature will be followed and the chemical potential will generally be referred to as the Fermi energy for all temperatures.

The Sommerfeld model assumes that there are no electron-electron interaction forces and averages the potential of each of the ion cores over the whole crystal and sets this potential to zero. As such, the time-independent Schrödinger equation (TISE) can be written

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)\psi_k = E_k\psi_k \tag{1.2}$$

E_k and ψ_k denoting its eigenvalues and eigenvectors respectively. Due to the lattice structure formed by the ion cores, the wave function itself is required to be periodic in x, y and z. This in mind, if the free electron gas is taken to be confined to a 'box' of side length L and volume L^3 containing N electrons, then the wave functions must satisfy the boundary conditions

$$\begin{aligned}\psi_k(x+L,y,z)&=\psi_k(x,y,z)\\ \psi_k(x,y+L,z)&=\psi_k(x,y,z)\\ \psi_k(x,y,z+L)&=\psi_k(x,y,z)\end{aligned} \tag{1.3}$$

The wave functions are therefore given by the travelling plane wave

$$\psi_k(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} \tag{1.4}$$

with the x components of the wave vector $\mathbf{k}$ satisfying

$$k_x = 0,\pm\frac{2\pi}{L},\pm\frac{4\pi}{L},\ldots \tag{1.5}$$

since $e^{ik_xL}=1$ from Eq. (1.3), and similarly for k_y and k_z. The corresponding eigenvalues E_k follow from Eq. (1.2):

$$E_k = \frac{\hbar^2k^2}{2m} \tag{1.6}$$

this is often referred to as the dispersion relation of the wave function along its direction of propagation.

In solid-state physics, it is often useful to treat problems not in conventional space, but in a reciprocal space defined by the orthonormal vector set k_x, k_y, k_z, called **k**-space. In this new space, points of equal energy form surfaces, with that surface corresponding to E_f known as the Fermi surface. The Fermi surface acts to separate the occupied and unoccupied states (strictly speaking only at $T = 0$ K, but the situation is little different for finite temperature [35]). In the free electron

model the Fermi surface is a sphere of radius k_f, see Figure 1-3, where $E_f = \hbar^2 k_f^2/2m$ from Eq. (1.6).

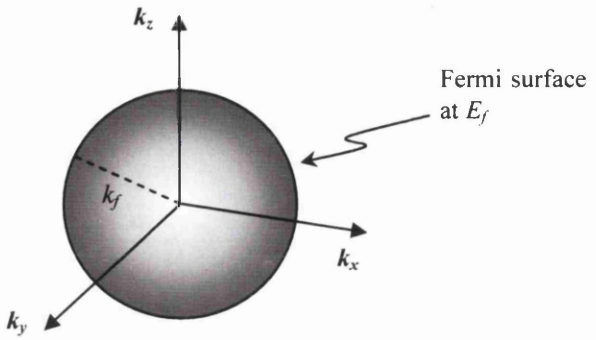


Figure 1-3 Representation of the Fermi sphere in the k-space of the free electron model. Occupied orbitals fill the sphere up to its radius k_f. Strictly, this is for $T = 0$ K although there not a great deal of difference for finite temperature [35].

There are two electrons resident on each allowed wave vector or k point within the Fermi sphere, and each k point has a volume of $(2\pi/L)^3$ in conventional space. The sphere has a volume of $4\pi k_f^3/3$ and so contains N electrons, where N satisfies

$$N = 2\frac{4\pi k_f^3}{3}\left(\frac{L}{2\pi}\right)^3 = \frac{L^3}{3\pi^2}k_f^3 \tag{1.7}$$

Consequentially, the Fermi level can be expressed as

$$E_f = \frac{\hbar^2}{2m}\left(\frac{3\pi^2 N}{L^3}\right)^{2/3} \tag{1.8}$$

from Eq.'s (1.6) and (1.7).

Defining the number of orbitals per unit energy as the density of states, $D(E)$, such that

$$N = \int_0^\infty f(E)D(E)dE \tag{1.9}$$

then

$$D(E) = \frac{dN}{dE} \tag{1.10}$$

The total number of orbitals of energy $\leq E$ is given by

$$N = \frac{L^3}{3\pi^2}\left(\frac{2mE}{\hbar^2}\right)^{3/2} \tag{1.11}$$

from Eq. (1.7), and so, for the three dimensional free electron model

$$D(E) = \frac{L^3}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2} E^{1/2} \tag{1.12}$$

The volume L^3 is conventionally set to 1.

The Sommerfeld model gives good insight into many of the properties of metals (such as conductivity (thermal and electrical), magnetic susceptibility and heat capacity), but fails to describe some of their basic properties, such as the polarity of the charge carriers in the Hall coefficient. The next evolutionary step, as it were, in the consideration of the nature of solid matter is to take into account the periodic lattice potential and its influence on the conduction electrons. This model is often called the nearly free electron model.

1.3.2 Nearly Free Electron Model

The lattice or crystal of ions through which the conduction band electrons travel is taken to be a structure formed from repeated identical units, known as unit cells. The repeat length of the unit cell making up the crystal is known as the lattice parameter a; for gallium arsenide (GaAs) the lattice parameter is a = 0.565nm. A primitive cell is a minimum volume cell which can be constructed in different ways according to different conventions, but contains only one lattice point, see Kittel [36] for details, and the Brillouin zone is defined as the reflection of the Wigner-Seitz primitive cell into the reciprocal lattice. The first Brillouin zone will have particular importance latter in this section and can be envisaged as [36]:

> "*... the smallest volume enclosed by the perpendicular bisectors of the reciprocal lattice vectors drawn from the origin*"

see for instance Figure 1-4.

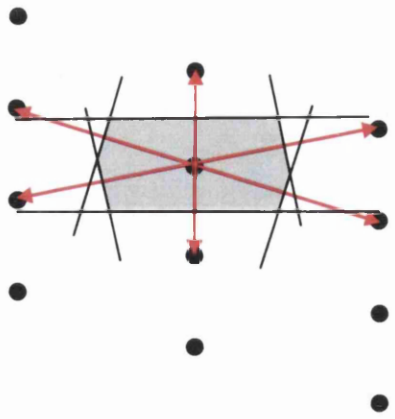


Figure 1-4 Construction of the first Brillouin zone (shaded region) for an oblique 2-dimensional reciprocal lattice.

The solution to the TISE for the periodic potential $V(\mathbf{r})$ formed from these repeating cells is given by the Block function (1.13)

$$\psi_k(\mathbf{r}) = u_k(\mathbf{r})e^{i\mathbf{k}.\mathbf{r}}$$

see Kittel [36] for proof. It is the product of the two functions – the 'unit cell' function $u_k(\mathbf{r})$ having the periodicity of the lattice (and so the same in each unit cell throughout the lattice) and the plane wave 'envelope' function $e^{i\mathbf{k}.\mathbf{r}}$.

In one dimension, this wave function ψ_k forms the solution to the Kronig-Penney model, which can be used to give a remarkable insight into the fundamental difference between insulators and conductors; the existence of forbidden energy gaps in the energy spectrum.

The Kronig-Penney model assumes that the 1D periodic potential can be modelled as a periodic array of square wells as illustrated in Figure 1-5.

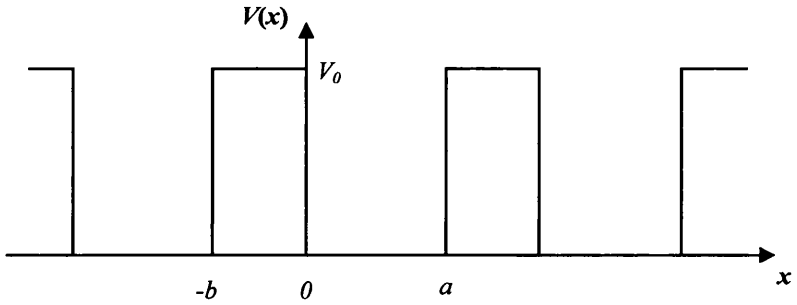


Figure 1-5 1-dimensional Kronig-Penney potential

The period of the potential is $a + b$, and the TISE for the regions $0 < x < a$ and $-b < x < 0$ can be written

$$\begin{aligned} \frac{d^2\psi_k}{dx^2} + \frac{2m}{\hbar^2}E_k\psi_k = 0 \qquad & 0 < x < a \\ \frac{d^2\psi_k}{dx^2} + \frac{2m}{\hbar^2}(E_k - V_0)\psi_k = 0 \qquad & -b < x < 0 \end{aligned} \tag{1.14}$$

Letting

$$\begin{aligned} \alpha^2 &= \frac{2m}{\hbar^2}E_k \\ \beta^2 &= \frac{2m}{\hbar^2}(E_k - V_0) \end{aligned} \tag{1.15}$$

and assuming that $E_k < V_0$, then from the Block theorem i.e. $\psi_k(\mathbf{r}) = u_k(\mathbf{r})e^{i\mathbf{k}.\mathbf{r}}$, Eq. (1.14) becomes

$$\begin{aligned} \frac{d^2u_k}{dx^2} + 2ik\frac{du_k}{dx} + \left(\alpha^2 - k^2\right)u_k = 0 \qquad & 0 < x < a \\ \frac{d^2u_k}{dx^2} + 2ik\frac{du_k}{dx} - \left(\beta^2 + k^2\right)u_k = 0 \qquad & -b < x < 0 \end{aligned} \tag{1.16}$$

with the solutions

$$\begin{aligned} u_{1k} &= Ae^{i(\alpha-k)x} + Be^{-i(\alpha-k)x} \qquad 0 < x < a \\ u_{2k} &= Ce^{(\beta-ik)x} + De^{-(\beta-ik)x} \qquad -b < x < 0 \end{aligned} \tag{1.17}$$

The boundary conditions of the wave functions require that

$$\begin{aligned} u_{1k}(0) &= u_{2k}(0) \\ u_{1k}(a) &= u_{2k}(-b) \end{aligned} \tag{1.18}$$

and from the continuity of ψ_k and $d\psi_k/dx$ that

$$\begin{aligned} \left.\frac{du_{1k}}{dx}\right|_{x=0} &= \left.\frac{du_{2k}}{dx}\right|_{x=0} \\ \left.\frac{du_{1k}}{dx}\right|_{x=a} &= \left.\frac{du_{2k}}{dx}\right|_{x=-b} \end{aligned} \tag{1.19}$$

It can be shown, using the method of Merzbacher [37], that this implies that

$$\frac{\beta^2 - \alpha^2}{2\alpha\beta}\sinh\beta b \sin\alpha a + \cosh\beta b \cos\alpha a = \cos k(a+b) \tag{1.20}$$

For convenience, let $V_0 \rightarrow \infty$ and $b \rightarrow 0$ in such a way that the product $V_0 b$ remains finite (finite square barriers $\rightarrow$ delta functions), then Eq. (1.20) becomes

$$P\,\mathrm{sinc}\,\alpha a + \cos\alpha a = \cos ka \tag{1.21}$$

where

$$P = \frac{mV_0 ba}{\hbar^2} \tag{1.22}$$

This implies that Eq. (1.21) can only be satisfied for the values of αa for which $P\,\mathrm{sinc}\,\alpha a + \cos\alpha a$ lies between ±1, see Figure 1-6. For the other values of the energy there are no Block function solutions to the wave equation, and so forbidden regions, or gaps, arise in the energy spectrum.

It can be seen from Figure 1-6 that the width of the allowed energy bands increases with increasing αa (with increasing energy) since the magnitude of $P\,\mathrm{sinc}\,\alpha a$ decreases.

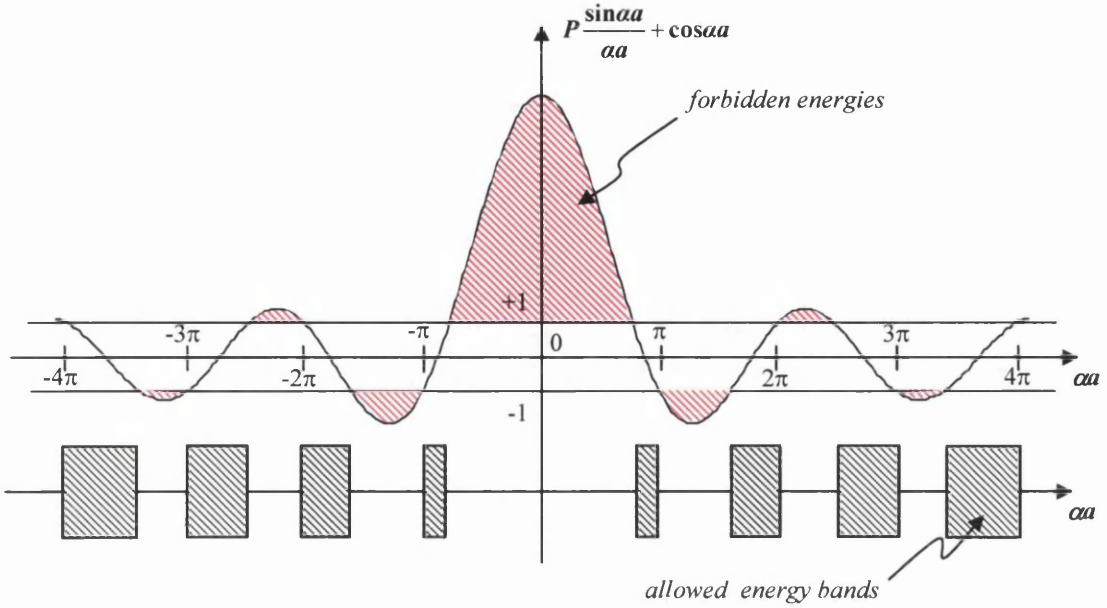


Figure 1-6 Plot of $P\text{sinc}\alpha a + \cos\alpha a$ for $P = 2\pi$. The forbidden values of the energy (red hashed regions) are given by those ranges of αa where the function exceeds ±1. The allowed energy ranges, or permissible energy bands, are marked on the second axes as hashed blocks.

If P increases then the 'binding energy' of the electron can be seen to increase. For example, if $P \rightarrow \infty$ then $\sin\alpha a$ must tend to 0, implying that $\alpha a = \pm n\pi$ where $n = 0, 1, 2, 3...$ and therefore from Eq. (1.15),

$$E_n = \frac{\hbar^2 n^2 \pi^2}{2ma^2} \tag{1.23}$$

the familiar expression of the energy levels of a particle confined within a quantum box.

At the other limit, $P \rightarrow 0$ then $\cos\alpha a = \cos ka$ implying that $\alpha a = ka$, and therefore Eq. (1.15) becomes Eq. (1.6)

$$E_k = \frac{\hbar^2 k^2}{2m}$$

the continuous energy spectrum of the free electron model.

From Eq. (1.21) it is possible to deduce the onset of the energy gaps at

$$k = \frac{n\pi}{a} \qquad n = 0, \pm 1, \pm 2, \ldots \tag{1.24}$$

from $\cos ka = \pm 1 \Rightarrow ka = n\pi$. These k values define the boundaries of the Brillouin zones – the first zone extends from $-\pi/a$ to π/a, the second from $-2\pi/a$ to $-\pi/a$ and π/a to $2\pi/a$ and so forth.

Due to the periodic nature of Eq. (1.21), if k is replaced by $k+2\pi n/a$ where n is an integer, then Eq. (1.21) remains unchanged. This motivates the use of the reduced wave vector, limited to the region

$$-\frac{\pi}{a} \le k \le \frac{\pi}{a} \tag{1.25}$$

In this way, the extended zone representation of the Energy-wave vector relationship can be reduced to the reduced zone representation of the reduced wave vector, see Figure 1-7.

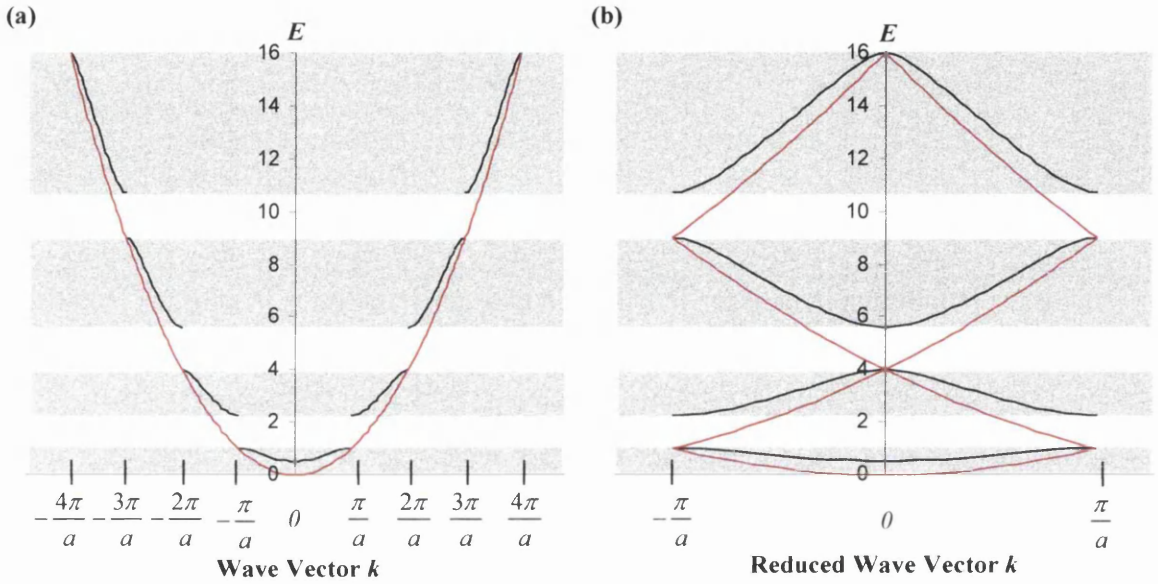


Figure 1-7 Plot of Energy vs. (a) Wave vector k (b) Reduced wave vector k for the Kronig-Penney model with $P = 3\pi/2$ (black line) and the free electron model (red line). Energy in units of $\hbar^2\pi^2/2ma^2$. The shaded bands represent the allowed energy bands of the Kronig-Penney model; needless to say, the whole spectrum of E is allowed in the free electron model.

The shaded regions represent the bands of allowed energies, separated from each other by the forbidden energy gaps.

As would be anticipated, on departing from the Kronig-Penney 1-dimensional case the 2- and 3-dimensional Brillouin zones become more complex, see Blakemore [35], and surfaces of constant energy in **k**-space depart considerably from the sphere of Figure 1-3, see Kittel [36]. Nonetheless, this 1-dimensional model is sufficient to introduce the very important concept of allowed and forbidden energy regions, if in a rather idealised manner.

The velocity, or rather group velocity[i], of an electron at energy E_k moving through the crystal lattice in real space can be written

$$\boldsymbol{v}_g = \frac{d}{d\mathbf{k}}\left(\frac{E_k}{\hbar}\right) = \frac{\nabla_k E_k}{\hbar} \tag{1.26}$$

using $dE_k = \nabla_k E_k \cdot d\mathbf{k}$, and so its acceleration, when subjected to an external electric field for instance, is given by

$$\begin{aligned} \boldsymbol{a} &= \frac{d\boldsymbol{v}_g}{dt} \\ &= \frac{\nabla_k}{\hbar}\frac{dE_k}{dt} \end{aligned} \tag{1.27}$$

The rate of change of the energy of the electron is equal to the scalar product of the force on the electron $-e\boldsymbol{E}$ and $\boldsymbol{v}_g$ its velocity i.e.

$$\frac{dE}{dt} = -e\boldsymbol{E} \cdot \boldsymbol{v}_g \tag{1.28}$$

therefore

$$\boldsymbol{a} = -\frac{e}{\hbar^2}\nabla_k \nabla_k E_k \cdot \boldsymbol{E} \tag{1.29}$$

On comparison with Newton's 2nd law, the tensor $\nabla_k \nabla_k E_k / \hbar^2$ must have the dimensions of $(\text{mass})^{-1}$. This introduces the effective mass tensor for an electron subject to a periodic potential

$$\left[m_{ij}\right] = \frac{\hbar^2}{\dfrac{\partial^2 E_k}{\partial k_i \partial k_j}} \tag{1.30}$$

and it can be seen that the effective mass of the electron is inversely proportional to the curvature of the band. This quantity reflects the variation in the propagation of the electron wave function along the different planes of the lattice. For this work it is assumed that while the main semiconductor of interest, SnO_2, is anisotropic (the electron mass is dependent on the direction of wave propagation) by defining the 'density of states' electron mass

$$m_D^* = \left(m_x^* m_y^* m_z^*\right) \tag{1.31}$$

referred to here simply as m^*, an adequate description of the effective mass along the radial axis of the polycrystalline[ii] spherical grains is provided. For isotropic semiconductors, the effective

---

[i] The group velocity is the velocity of energy propagation in the medium and is defined as $\partial\omega/\partial\mathbf{k}$ where ω is the angular frequency.

electron mass is equal to the density of states effective mass. It is implicitly assumed throughout the rest of this section that semiconductor refers to an isotropic semiconductor, and that the dispersion relation Eq. (1.6)

$$E_k = \frac{\hbar^2 k^2}{2m^*}$$

is a satisfactory approximation, handily allowing the effective mass to be considered independent of energy. This is a better approximation near the band maxima and minima.

If the allowed energy bands are either full or empty, then no electrons are free to move if an electric field is applied, and the crystal behaves like an insulator. If one or more of the bands is only partially filled, then the electrons are free to move and the crystal is deemed a metal. Between these two extremes lie semiconductors (and semimetals) – see Figure 1-8.

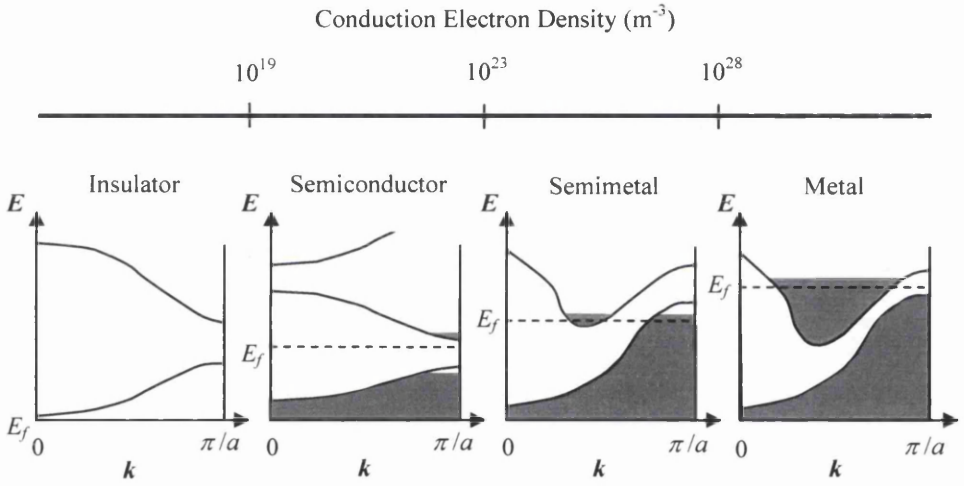


Figure 1-8 Schematic representation of occupied states and band structure of insulators, semiconductors, semimetals and metals at a low, but finite, temperature. Approximate charge carrier concentrations are also given, although the semiconductor range can be extended via doping.

Generally, 'the' conduction band refers to the lowest energy empty, or conduction, band, and 'the' valence band refers to the highest energy fully occupied, or valence, band. The band gap is defined as the energy difference between the lowest point of the conduction band and the highest point of the valence band.

[ii] A polycrystal is an object composed of randomly oriented crystalline regions. Polycrystalline materials usually result when a substances solidifies rapidly with crystallisation commencing at many nucleation sites.

For metals and semimetals the conduction band and valence band can be seen as overlapping (no band gap), as schematically illustrated in Figure 1-9. If the band overlap is small with only few states involved, then the material is treated as a semimetal, see Kittel [36] for a more precise and detailed treatment.

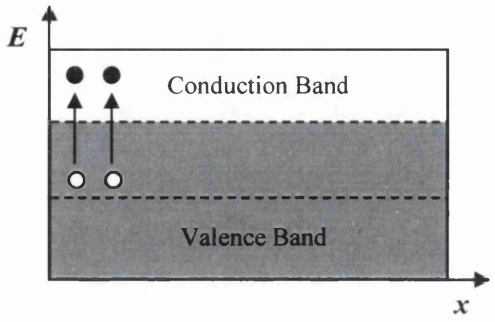


Figure 1-9 Schematic representation of metal (or semimetal) conduction and valence bands in conventional space. Note band overlap. Electrons easily excited into the conduction band.

For insulators however, the band gap is large (>5eV), and as such, electrons cannot be excited, thermally or otherwise, into the conduction band, see Figure 1-10. If the insulator is excessively heated, then the material breaks down.

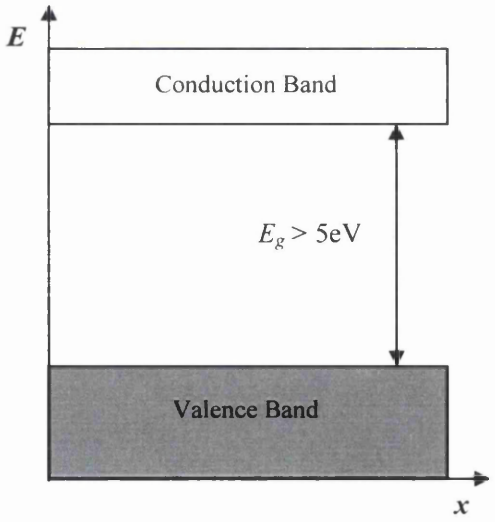


Figure 1-10 Schematic representation of the conduction and valence bands of an insulator in conventional space. Bands seperated by wide band gap, E_g. Electron transitions between bands prohibited.

Semiconductors have an intermediate band gap ($0 < E_g < 5eV$) and it is possible that electrons can be promoted by light or heat, from the valence band into the conduction band leaving behind a positively charged 'hole' in the valence band, see Figure 1-11. Both the electrons and their holes correspond to electrical conductivity.

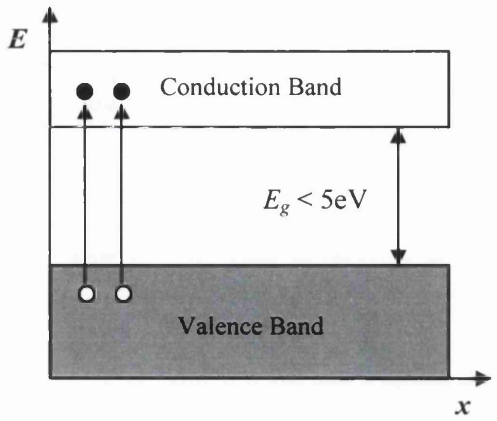


Figure 1-11 Schematic representation of the conduction and valence bands of a semiconductor in conventional space. The band gap, E_g, between the bands is such that electrons can be promoted from the valence band into the conduction by thermal excitations or photons. $0 < E_g < 5eV$.

If the conduction band has a minimum in reciprocal space at the same value of **k** as the valence band maximum as illustrated in Figure 1-12, then the semiconductor is referred to as a direct band gap semiconductor since a direct optical electron transition between bands is possible through the absorption of a photon.

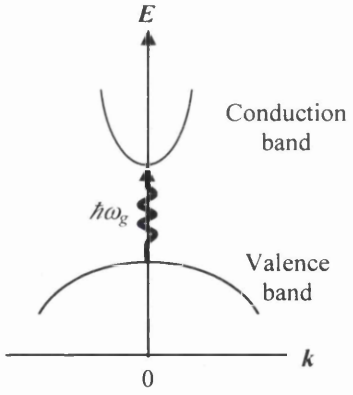


Figure 1-12 Schematic of a direct band gap. Conduction band minima and valence band maxima located at the same point in k space. The photon is of energy $E_g = \hbar\omega_g$ (black wavy line).

For an indirect semiconductor on the other hand, the conduction band maxima and valence band minima are widely separated in **k**-space, see for instance Figure 1-13, and both a photon and a phonon[iii] are required to optically promote an electron from the valence band into the conduction band.

[iii] A quantum of oscillation in a crystal lattice made to vibrate via heat or sound waves.

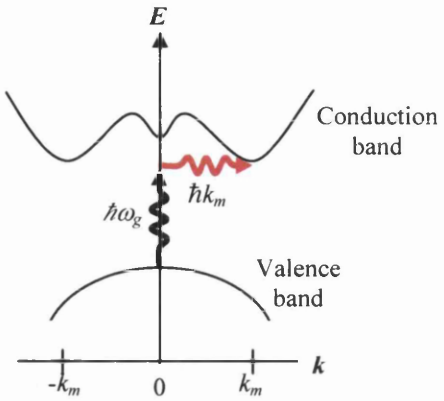


Figure 1-13 Schematic of an indirect band gap. Conduction band minima and valence band maxima not located at the same point in k space, but separated by k_m. The photon is of the energy $E_g = \hbar\omega_g$ (black wavy line) and phonon of momentum $\hbar k_m$ (red wavy line).

If a semiconductor is pure and crystallographically perfect, then the density of electrons in the conduction band, n_0, and the density of holes in the valence band, p_0, will be equal (intrinsic semiconductor). The Fermi level is then placed at the centre of the band gap, see Blakemore [35]. However, if dopant atoms or flaw states are present in the semiconductor for instance (extrinsic semiconductor), it is possible that one kind of charge carrier will dominate, and the position of the Fermi level will alter, see Figure 1-14.

In an n-type extrinsic semiconductor, negative mobile charge carriers (electrons) of density n dominate the electronic conduction; donor atoms (for example impurities in the crystal[iv]) or flaws are present which become positively charged, releasing electrons into the conduction band (although these electrons may also of course become trapped within another flaw). In a p-type extrinsic semiconductor, acceptor atoms[v] or flaws are present which become negatively charged, accepting electrons from the valence band, and the conduction is dominated by the positive charge carriers (holes) of density p.

[iv] For instance if a lattice composed of Group IV elements e.g. Si, is doped with Group V donor atoms e.g. As, then each impurity atom occupies a Si lattice point. Each dopant atom forms 4 covalent bonds with its neighbours, but has a 'spare' valence electron which can be promoted to the conduction band, leaving the donor 'atom' behind as an ionised state. These donor impurities are represented as localised states below the conduction band.

[v] For instance if a lattice composed of Group IV elements e.g. Si, is doped with Group III acceptor atoms e.g. P, then each impurity atom 'accepts' an excited electron from the valence band in order to complete its 4 covalent bonds. The positive hole left behind in the electron's wake remains weakly bound to the excess negative charge. These acceptor impurities are represented as localised states above the valence band.

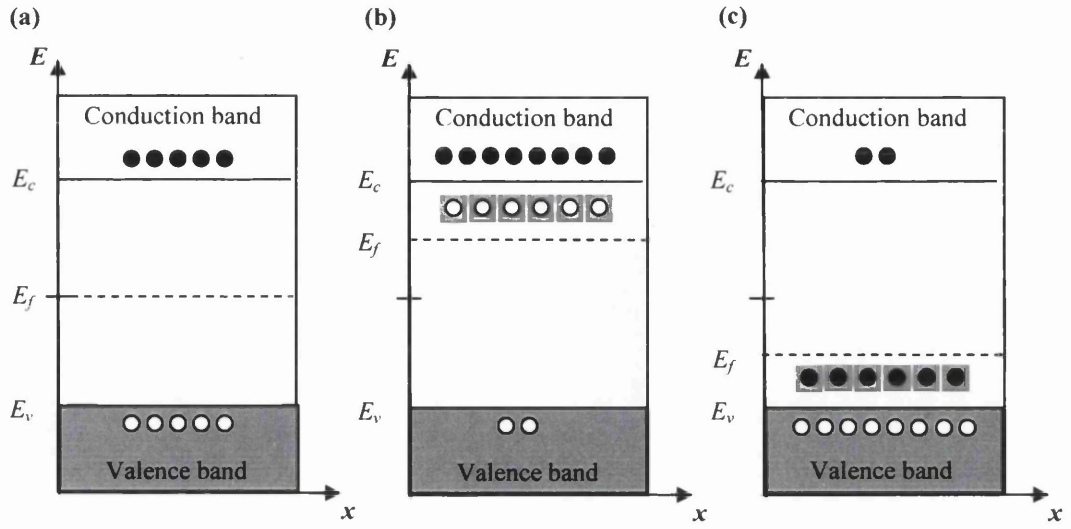


Figure 1-14 Schematic representation of Fermi level for (a) intrinsic (b) *n*-type and (c) *p*-type semiconductors.

Assuming that the conduction band has a single energy minimum at the centre of the Brillouin zone $E = E_c$, and that the effective mass is energy independent, then in the same way that Eq. (1.12) was calculated, it is possible to derive the corresponding equation for the density of states within a semiconductor

$$D(E) = \frac{1}{2\pi^2}\left(\frac{2m^*}{\hbar^2}\right)^{3/2}\left(E - E_c\right)^{1/2} \tag{1.32}$$

for $E \geq E_c$. At a temperature T there is at equilibrium a unique energy distribution and Fermi level for those electrons thermally excited into the conduction band. The probability that a state of energy E is occupied is given by the Fermi fractional occupancy factor $f(E)$ of Eq. (1.1), and therefore the total electron density is given by

$$\begin{aligned} n &= \int_{E_c}^{\infty} f(E)D(E)dE \\ &= \frac{1}{2\pi^2}\left(\frac{2m^*k_BT}{\hbar^2}\right)^{3/2}\int_{E_c/k_BT}^{\infty}\frac{\sqrt{(E-E_c)/k_BT}}{1+e^{(E-E_F)/k_BT}}d(E/k_BT) \end{aligned} \tag{1.33}$$

which can be written

$$n = N_c\mathcal{F}_{1/2}\left(\frac{E_f - E_c}{k_BT}\right) \tag{1.34}$$

using the Fermi-Dirac integral,

$$F_j(y_0) = \int_0^{\infty}\frac{y^j dy}{1+e^{y-y_0}} = \Gamma(j+1)\mathcal{F}_j(y_0) \tag{1.35}$$

see Appendix I, Approximation of the Fermi-Dirac Integral, with the gamma function $\Gamma(3/2)=\sqrt{\pi}/2$, see Appendix I, Gamma and Beta Functions, and defining the effective density of conduction band states, N_c, as

$$N_c = 2\left(\frac{m^* k_B T}{2\pi\hbar^2}\right)^{3/2} \tag{1.36}$$

The Fermi-Dirac integral is not trivial to evaluate, and in practice, substantial approximations are made if the density belongs to either of two important limiting cases, see Figure 1-15.

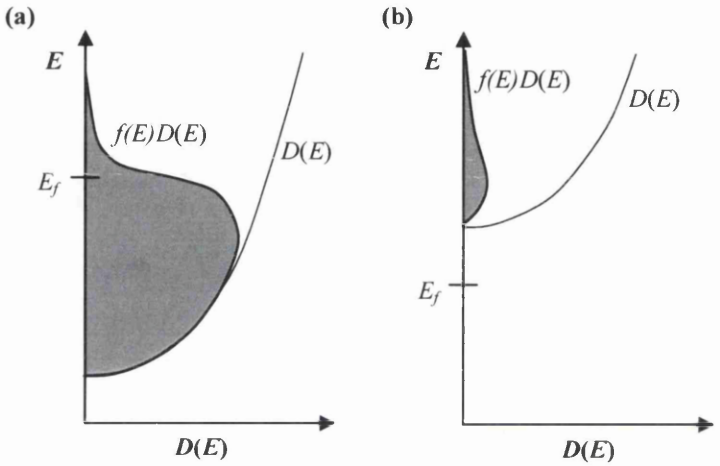


Figure 1-15 Representation of the density of states, $D(E)$, for two limiting cases, (a) degenerate electron density and (b) non-degenerate (or classical) electron density.

If the Fermi level lies at least ~$2k_BT$ into the conduction band, when the temperature is small and the conduction electrons numerous, then E_f - E_c » k_BT and the conduction electron gas is classified as degenerate[vi]. In the asymptotic limit of large positive y_0 the Fermi-Dirac integral becomes [35]

$$F_j(y_0) \approx \frac{y_0^{j+1}}{j+1}\left[1+\frac{\pi^2 j(j+1)}{6y_0^2}+\ldots\right] \qquad y_0 \gg 1 \tag{1.37}$$

and so the metallic-like total electron density can be written as

$$n = \frac{1}{3\pi^2}\left(\frac{2m^*(E_f - E_c)}{\hbar^2}\right)^{3/2} \tag{1.38}$$

with a Fermi energy of

[vi] The non-degenerate/degenerate terminology is not to be confused with state degeneracy (more than one quantum state for a given energy).

$$E_f = E_c + \frac{\hbar^2}{2m^*}\left(\frac{3\pi^2}{n}\right)^{2/3} \tag{1.39}$$

When the Fermi level is at least ~$2k_BT$ below the bottom edge of the conduction band, when total electron density is very small or the temperature very high, and only a small fraction of the band states are occupied, then the conduction electron gas is referred to as non-degenerate (or classical). In this case E_c is taken to be substantially greater than E_f and the asymptotic from of F_j for large negative y_0

$$F_j(y_0) \approx \Gamma(j+1)e^{y_0} \qquad y_0 < -2 \tag{1.40}$$

can be used [35], so that

$$n = N_c e^{(E_f - E_c)/k_BT} \tag{1.41}$$

and

$$E_f = E_c - k_BT \ln\left(\frac{N_c}{n}\right). \tag{1.42}$$

Naturally, an equivalent set of conductive hole relationships can also be developed. However, since the main semiconductor of interest for this work, SnO_2, is naturally *n*-type once annealed due to the presence of oxygen vacancies[vii] which act as impurities in the crystal (see also Section 2.1), a detailed description of hole behaviour is redundant, and so the reader is referred to the main references of this section [35,36].

As important as this synopsis of solid state physics was, as will be detailed in the next section, the bulk semiconductor equations of this section have in general little relevance to the simulation of the quantised energetic behaviour of a 4nm radius SnO_2 spherical grain, other than as first estimation.

Nevertheless, it is certainly evident from such equations as Eq. (1.39) and Eq. (1.42) (aside from

[vii] In compound semiconductors (e.g. Tin dioxide) the crystal lattice forms out of precise ratios of the constituent atoms. Crystals with the nominal atom ratios are termed stoichiometric. Some defects, such as vacancies (where an atom is missing from its usual site in the lattice), are termed stoichiometric defects, the crystal now departing from the nominal ratio of the two atoms. The lattice of SnO_2 forms out of O^{-2} anions and Sn^{+4} cations. Annealing of the material seems to encourage oxygen vacancies (see for example Samson and Fonstad [96]), freeing up to 2 electrons to be promoted into the conduction band.

the constituent parts of their derivations) that a fundamental connection exists between the Fermi level and the density of electrons in the conduction band. This illustrates, perhaps more than a purely verbal explanation, the importance of always understanding the relationship and behaviour of the two if the population of conduction band electrons is altered, as occurs during the process of charge writing for example. These two equations, although not even applicable in the dimensions where quantisation dominates, exemplify the complexity of the inter dependence of the two quantities and emphasise the inappropriateness of merely placing the Fermi level at the energetic position of the lowest empty orbital in this non-zero temperature range.

1.4 Introduction to Quantum Confinement

The work of the preceding section, Section 1.3, was dedicated to the basics of solid-state physics for bulk materials. The main focus of this thesis however is not that of the bulk medium, but of the quantum realm. This section is a simple introduction to how the behaviour of the conduction electrons change as their degrees of motion are curtailed in idealised rectangular semiconductor. It offers a brief look at their response as the physical dimensions of their lattice are reduced, and their charge carriers are confined in one (a quantum well), two (a quantum wire) and finally, all three (a quantum dot) dimensions, as illustrated in Figure 1-16. It presents a simplified glimpse at the quantisation of the electrons' energy spectrum and discusses the density of states for each reduced dimensionality.

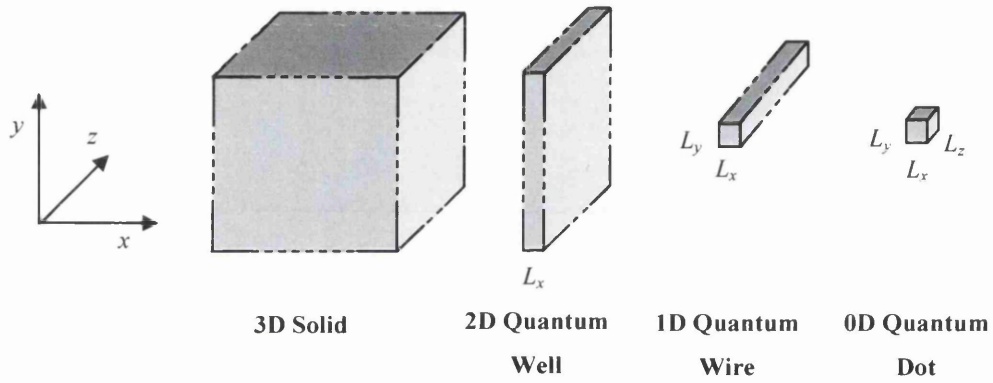


Figure 1-16 Representation of Quantum Confinement. The dimensionality of the structure represents its number of degrees of freedom. L_x denotes the diameter of the quantised region in the x direction, and similarily for the y- and z-axes.

In the bulk solid, the plane-wave wave function is free to propagate along any of its three axes. Its energy spectrum is continuous, not discrete, with a plethora of states for any given energy, as can be seen from the standard, bulk, density of states; see for instance Figure 1-17.

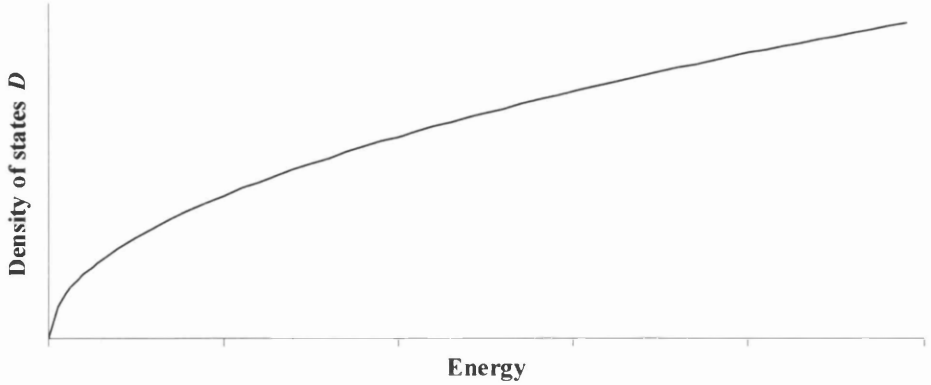


Figure 1-17 Density of states for a bulk semiconductor.

In contrast, in a quantum well, confining the electrons in one spatial dimension, along the x-axis for example, causes the formation of quantum states along that direction of motion. For the idealised case of an infinitely deep rectangular potential in this one dimension, let

$$\begin{aligned} V(x) &= 0 \qquad |x| < L_x \\ &= \infty \qquad \text{otherwise} \end{aligned} \tag{1.43}$$

then ψ must therefore equal zero for $|x| > L_x$, implying that $\psi(L_x) = \psi(-L_x) = 0$. Inside the well

$$\frac{d^2\psi}{dx^2} + \frac{2m}{\hbar^2} E\psi = 0 \tag{1.44}$$

If $E < 0$ then Eq. (1.44) can be expressed as

$$\frac{d^2\psi}{dx^2} - \alpha^2\psi = 0 \tag{1.45}$$

with

$$\alpha^2 = \frac{2m}{\hbar^2}|E| \tag{1.46}$$

and it is evident that the boundary conditions cannot be satisfied for a linear combination of the standard solutions $e^{\alpha x}$ and $e^{-\alpha x}$. This implies therefore that $E > 0$. Setting

$$\beta^2 = \frac{2m}{\hbar^2} E \tag{1.47}$$

then

$$\frac{d^2\psi}{dx^2} + \beta^2\psi = 0 \tag{1.48}$$

with the solution $\psi = \sin kx$. From the boundary conditions it is clear that $\sin kL_x = 0$, implying that $kL_x = n_x\pi$, $n_x = 1, 2, 3 \ldots$. Therefore

$$E_{n_x} = \frac{\hbar^2 n_x{}^2 \pi^2}{2mL_x{}^2} \tag{1.49}$$

compare with Eq. (1.23).

The energy dispersion relation for the two unconfined directions, y and z, can be given to first approximation by the free electron dispersion relation, thus

$$E_{k_y k_z} = \frac{\hbar^2}{2m^*}\left(k_y^2 + k_z^2\right) \tag{1.50}$$

The total energy of each state for the quantum well depicted in Figure 1-16 is therefore

$$E_x = \frac{\hbar^2\pi^2 n_x^2}{2m^* L_x^2} + \frac{\hbar^2}{2m^*}\left(k_y^2 + k_z^2\right) \qquad n_x = 1,2,3\ldots \tag{1.51}$$

In the two dimensional **k**-space spanned by k_y and k_z, the Fermi surface is a circle of radius k_f centred on the origin (where $k_y = k_z = 0$). The circle has an area of πk_f^2 and with two electrons per k point of area $(2\pi)^2$ then the total number of electrons contained within the Fermi circle is

$$N = 2\frac{\pi k_f^2}{(2\pi)^2} = \frac{k_f^2}{2\pi} \tag{1.52}$$

The number of orbitals per unit energy, $D_{k_yk_z}(E_{k_yk_z})$, otherwise known as the density of states of the yz-space is

$$D_{k_yk_z}(E_{k_yk_z}) = \frac{dN}{dE_{k_yk_z}} = \frac{1}{2\pi}\frac{d}{dE_{k_yk_z}}\left(\frac{2m^*E_{k_yk_z}}{\hbar^2}\right) = \frac{m^*}{\pi\hbar^2} \tag{1.53}$$

an expression independent of energy. For the total 'two dimensional' density of states, it must be realised that each quantum state i.e. each discrete eigenstate of E_{n_x}, has a state density of $m^*/\pi\hbar^2$ in the yz-plane. Therefore, the total 2D density of states $D_x(E_x)$ is

$$D_x(E_x) = \frac{m^*}{\pi\hbar^2}\sum_{n_x}\theta(E_x - E_{n_x}) \tag{1.54}$$

using the Heaviside step function $\theta(E_x - E_{n_x})$ (introduced in Appendix III, Introduction to Green's Functions). $D_x(E_x)$ is illustrated in Figure 1-18.

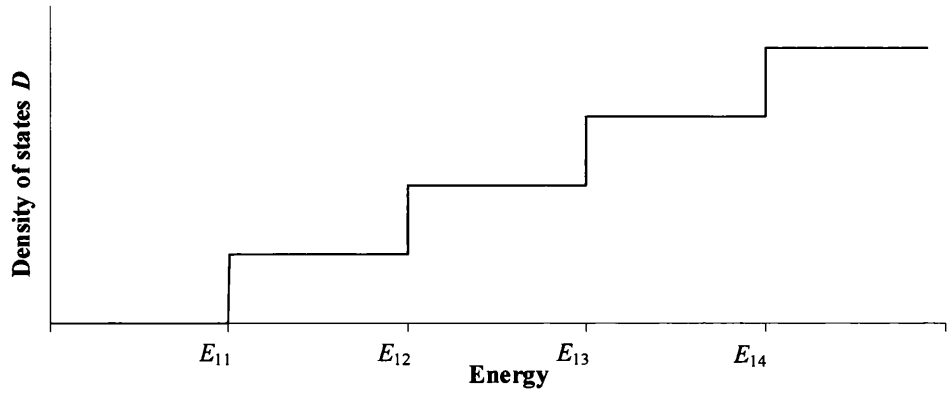


Figure 1-18 Density of states for a 2-dimensional quantum well. Continuous energy spectrum along y- and z-axes, confined along x-axis only

For a quantum wire, carrier confinement is in two dimensions, say the x- and y- axes, and as such, the travelling plane-wave wave function is only able to propagate along the z-axis. In this way

$$E_{xy} = \frac{\hbar^2\pi^2}{2m^*}\left(\frac{n_x^2}{L_x^2} + \frac{n_y^2}{L_y^2}\right) + \frac{\hbar^2k_z^2}{2m^*} \qquad n_x, n_y = 1,2,3\ldots \tag{1.55}$$

analogous to the case of the quantum well.

To access the density of states a little more thought is required, the Fermi surface now a single k point, k_f, necessarily equal to k_z and so of 'volume' 2π. Representing the **k**-space volume 'within' this Fermi surface as V_{k_z}, then

$$\begin{aligned} D_{k_z}(E_{k_z}) &= \frac{dN}{dE_{k_z}} \\ &= \frac{2}{2\pi}\frac{dV_{k_z}}{dE_{k_z}} \\ &= \frac{1}{\pi}\frac{dV_{k_z}}{dk_z}\frac{dk_z}{dE_{k_z}} \\ &= \frac{1}{\pi\hbar^2}\sqrt{\frac{m^*}{2E_{k_z}}} \end{aligned} \tag{1.56}$$

as dV_{k_z}/dk_z must equal 1. Each quantised state has a state density of Eq. (1.56), therefore the total 'one dimensional' density of states is

$$D_{xy}(E_{xy}) = \frac{1}{\pi\hbar^2}\sum_{n_x n_y}\sqrt{\frac{m^*}{2(E_{xy}-E_{n_x n_y})}}\theta\left(E_{xy}-E_{n_x n_y}\right) \tag{1.57}$$

as illustrated in Figure 1-19.

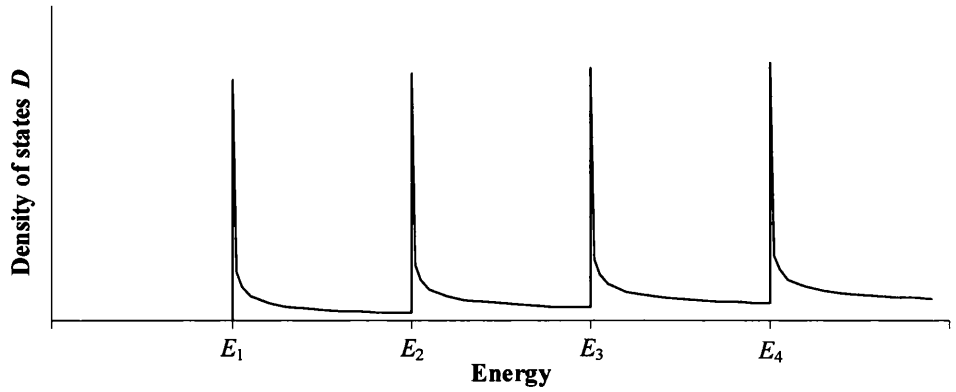


Figure 1-19 Density of states for a 1-dimensional quantum wire. Continuous energy spectrum along z-axes only, confined along x- and y-axes.

For the quantum dot, the electrons are confined in all three dimensions and no 'free' propagation is permissible. In this way the quantised total energy is

$$E_{xy} = \frac{\hbar^2\pi^2}{2m^*}\left(\frac{n_x^2}{L_x^2}+\frac{n_y^2}{L_y^2}+\frac{n_z^2}{L_z^2}\right) \qquad n_x, n_y, n_z = 1,2,3\ldots \tag{1.58}$$

and the 'zero dimensional' density of states is described by a series of delta functions at the energy locations of the discrete eigenstates. Thus

$$D_{xyz}(E_{xyz}) = 2\sum_{n_x n_y n_z} \delta\left(E_{xyz} - E_{n_x n_y n_z}\right) \tag{1.59}$$

as can be seen in Figure 1-20.

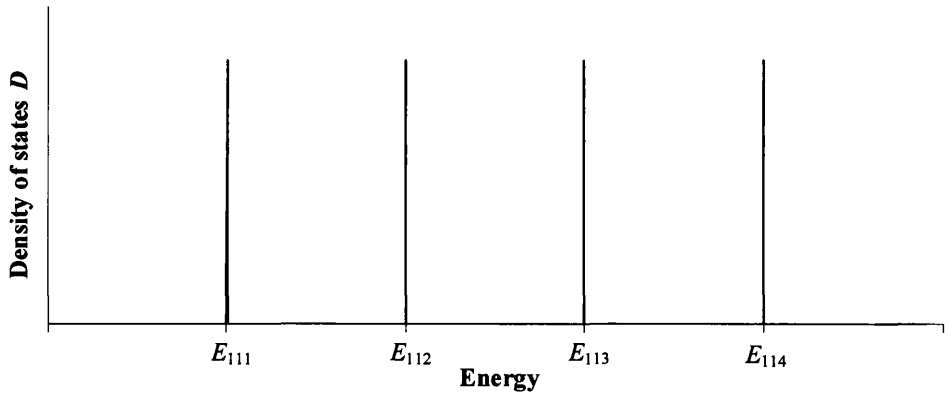


Figure 1-20 Density of states for a 0-dimensional quantum dot. Confined along all three axes. No continuous energy spectrum – electrons only supported in the discrete energy levels.

On each reduction of dimensionality, the energy and state density becomes more precisely defined, giving rise to enhanced charge carrier transport and optical properties.

Quantum wells pave the way for semiconductor lasers of far greater efficiency than the traditional diode laser, the location of its charge carriers more concentrated. Furthermore, merely by varying the depth and width of the well, the wavelength of the emitted light can be precisely tuned.

Quantum wires offer vastly superior conductivity and lower weight to their macroscopic counter parts. It is even feasible that bundles of nanowires could be manipulated to form a macroscopic 'rope', electrons readily tunnelling between individual nanosized quantum wire strands, transferring the quantum characteristics into the every day world.

Quantum dots can be used to form semiconductor lasers superior not only to their bulk versions, but to even quantum well lasers. They could be used to form the next generation of super-efficient photovoltaic cells [38], their spatial confinement increasing the efficiency with which the energy of incident photons can be converted to electrical energy. They could act as biological sensors, biological tags in the detection of tumours, LEDs, have roles in quantum cryptography

and quantum computing, as well as all the previously mentioned uses, from data storage to chemical catalysis, and a plethora of applications in addition to these.

It is indeed true to say that quantum confinement heralds a revolution in solid state physics that will have far reaching consequences.

1.5 Conclusions

This Chapter introduces the topic of charge writing and discusses the aims of this work and the rational behind it. It highlights the achievement of the Semiconductor Interface Laboratory in demonstrating charge writing on a SnO_2 nanocrystalline film, and indicates that while progress is being made in understanding this novel procedure through experiment, the knowledge of underlying physics behind this phenomenon is comparatively in its infancy.

It offers a guide to the theories used and developed herein and their achievements. It informs the reader of the tasks necessary to simulate the electronic characteristics of the nanocrystals at dimensions where the semiconductor band bending does not fully evolve, requiring the self-consistent solution of the non-linear Poisson equation and the Kohn-Sham equations, with inclusion of the effects of the surface states, also self-consistently reconciled with the electron density.

It presents a brief introduction to the basics of solid-state physics required to understand the work developed within this treatise, and introduces the consequences of confining the degrees of motion of the electrons within the semiconductor on the energy spectrum. It discusses the accompanying density of states for each reduced dimensionality and degree of quantisation.

Chapter 2 Surface States and Band Bending

This Chapter begins the development of the models of this work. It introduces surface states and their band bending effects, and lays down the simplest model of the conduction band of a small nanometric SnO_2 grain and the numerical techniques necessary to solve it. It also contains lengthy diversions into the topic of gas sensing, the customary application of these polycrystalline tin dioxide surfaces, both as a means to compare the effectiveness of the developed numerical methods with results from literature, and for the insights into the electronic behaviour of nanometric particles this subject affords; invaluable in assessing the most efficient routes forward in the charge writing simulation.

Section 2.1 discusses the nature of surface bonds, their formation and their effect on the conduction and valence bands of a semiconductor. The concepts of depletion regions and surface barriers are introduced, and a more quantitative description of the effects of surface charge and its effects is outlined through the conventional model of a metal-semiconductor interface. Fermi level pinning is introduced and the fundamentals of surface gas absorption are covered. The sensitivity of films of nanometric grains to reducing gases is then discussed, touching on sintering and possible conductivity mechanisms.

Section 2.2 introduces the non-linear Poisson equation and the boundary conditions of the complete charge density model of literature with which to simulate the band bending at the gas-semiconductor interface.

In Section 2.3 numerical techniques to solve the non-linear Poisson equation are outlined. The Taylor series expansion method is developed and its effectiveness with respect to alternative methods discussed. The basic concepts of computational error are introduced. To solve the

Poisson equation of Section 2.2 with its relevant boundary conditions, both the bisection and the shooting methods, to be used in conjunction with the Taylor series expansion technique, are developed, and their general efficiency with respect to alternative methods discussed.

Section 2.4 applies the developed numerical techniques to particular cases of SnO_2 nanocrystals and compares the results with those found in the literature. Through this, the depletion width concept is covered in more detail and the effects on the shape of the conduction band if the grain is too small to possess a non-depleted region introduced. This flattening of the curvature of the conduction band for small grain sizes is then considered in terms of conduction through a gas sensing film, and a very simple model for the sensitivity of such films (purely illustrative of trends and not to be considered exact) when the conduction process is dominated by transport over Schottky-like barriers is developed. The experimentally observed increase in sensitivity of gas sensing films on diminution of grain size, and in particular below $R = 10$ nm, is discussed both in terms of the conduction band movement and also with regard to the effect of surface states on Fermi level unpinning and the movement of the surface barrier height.

Section 2.4 presents a summary of the Chapter; its important results and their implications for the development of the charge writing model.

2.1 Surface States and the Bending of the Conduction Band Bottom

During the introduction to solid state physics of Section 1.2, one important facet of semiconductor crystals was neglected – the presence and effect of surface states. Surface states form, as may be expected from the name, at the surface of the semiconductor lattice, created by 'dangling' bonds from the crystal. These are the free bonds of the surface semiconductor atoms, available due to the disruption of the regular lattice repeating pattern caused by the physical edge of the material. These atoms, or rather ions, are no longer surrounded on all sides like their compatriots in the 'bulk' of the semiconductor as illustrated in Figure 2-1.

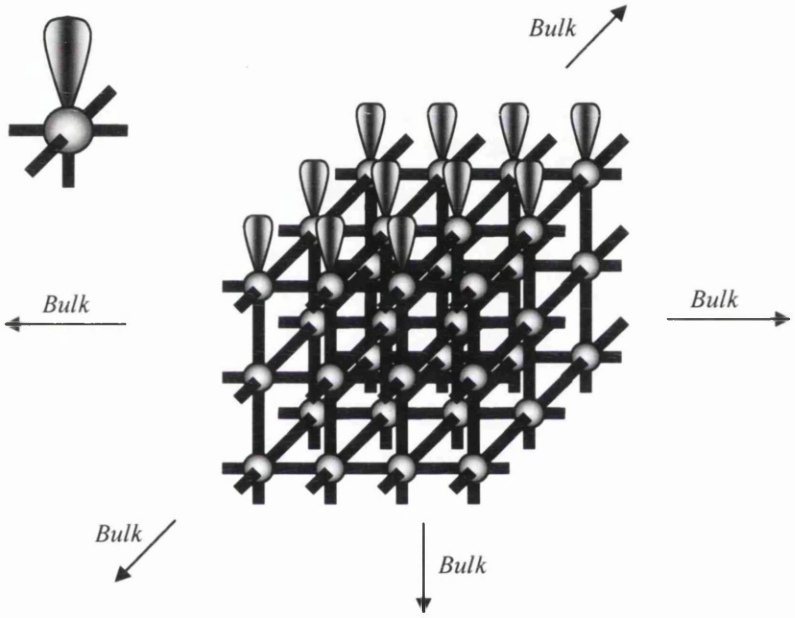


Figure 2-1 Schematic illustration of 'dangling' bonds (denoted by electron cloud) at the surface of a lattice.

Electrons promoted from the valence band and donor levels can become 'trapped' in these surface states, and as such, electrons can be thought of as draining from the conduction band into these states, the conduction and valence bands bending as a result. When equilibrium is established, a surface barrier and a positively charged spatial region are formed – see Figure 2-2, and a neutrality level can be defined from the surface value of the conduction band and the lowest occupied surface state. The Fermi level now lies within the surface states.

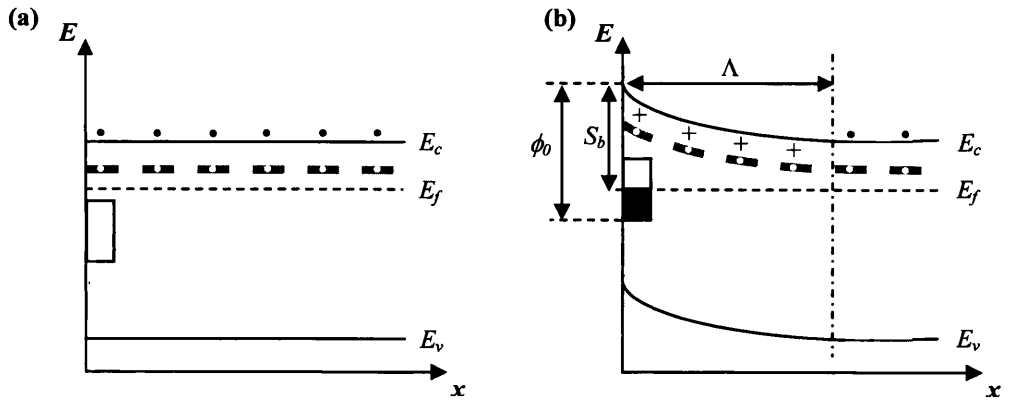


Figure 2-2 Illustration of the (a) flat band model of conduction and valence bands (b) surface state model in electrical equilibrium, the electrons having transferred from the conduction band into the surface states forming a surface barrier S_b and a charged region depleted of electrons Λ (depletion width). The Fermi level now lies within the surface states, and a charge neutrality position, ϕ_0, can be defined as the energy gap between the lowest surface state and the position of the conduction band at the surface of the grain. To correspond to the semiconductor of particular interest in this work, SnO_2, the semiconductor depicted here is n-type.

The effect and density of these intrinsic states can be enhanced or passivated by surface interactions with the adjacent medium (unless, of course, the crystal is in a vacuum); be that as a surface exposed to reacting gases, as in the case of the gas sensor, or in metal-semiconductor or heterojunction interfaces. In the latter, states can be induced through disorder in the lattice at the junction, and in all by chemical bonding.

To see the effect of these 'interaction' surface states alone, neglecting the intrinsic states of the semiconductor, then as described by Tung [39][i], in a modern reformulation of the work of Bardeen [40], consider a metal-semiconductor interface where the work function of the metal is Φ_M and χ_s the electron affinity of the semiconductor, as depicted in Figure 2-3. Let the band gap of the semiconductor at the interface have an induced surface density per unit area per electron volt of N_{ss} with a charge neutrality position of ϕ_0. The total charge at the surface of the semiconductor from these states per unit area is then $eN_{ss}(S_b - \phi_0)$ from Figure 2-3, where S_b is known as the Schottky barrier height[ii].

[i] To be fair, it should perhaps be noted that Tung's paper goes on to suggest that at the metal-semiconductor interface, it is not the presence of the traditional surface states of Bardeen that causes the formation of the Schottky barrier height, but rather polarised chemical bonds. In this work, this point is largely insignificant as the metal-semiconductor only appears in order to introduce the surface states of the gas-semiconductor interface, the presence of which are clearly accepted throughout the published literature.

[ii] As ϕ_0 is greater than S_b in this case, the net charge in the surface states is negative, and as such, N_{ss} act as 'acceptor' states here.

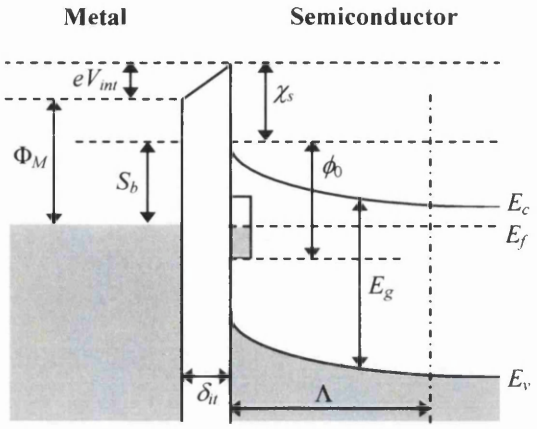


Figure 2-3 Schematic of metal-semiconductor interface bandbending

This charge and its associated image charge in the metal, separated by a short distance δ_{it}, forms a dipole across the junction, creating a small voltage drop across the interface. The barrier height, or specifically, the Schottky barrier height is then

$$S_b = \Phi_M - \chi_s - (S_b - \phi_0)e^2 \frac{N_{ss}\delta_{it}}{\varepsilon_{it}} \tag{2.1}$$

ε_{it} being the dielectric constant of the interface layer, and the interface voltage drop given by

$$V_{it} = -(S_b - \phi_0)e\frac{N_{ss}\delta_{it}}{\varepsilon_{it}} \tag{2.2}$$

i.e.

$$S_b = \Phi_M - \chi_s + eV_{it} \tag{2.3}$$

Defining the gap state parameter γ_{gs} as

$$\gamma_{gs} = \frac{1}{1 + e^2 \frac{N_{ss}\delta_{it}}{\varepsilon_{it}}} \tag{2.4}$$

then Eq. (2.1) can be re-arranged to yield

$$S_b = \gamma_{gs}(\Phi_M - \chi_s) + (1 - \gamma_{gs})\phi_0 \tag{2.5}$$

From this it can be seen that when N_{ss} is very large, say in the limit $\gamma_{gs} \rightarrow 0$, then $S_b \rightarrow \phi_0$. This means that the barrier height S_b is actually independent of the type of metal, and is a function purely of the nature of the surface states. Physically, this can be interpreted as the electron transfer between the semiconductor and the metal when the two interface not coming, as would be expected, from the semiconductor conduction band (the Fermi level situated at the edge of the populated states in the metal and in the band gap or conduction band for the semiconductor, assuming it is *n*-type at least) but from the surface states. Conventionally this is described as

Fermi level pinning, since the position of the Fermi level is fixed relative to the surface position of the conduction band in this limit.

At the other extreme, when $\gamma_{gs} \rightarrow 1$, when the surface charge is small, then $S_b \rightarrow \Phi_M - \chi_s$ and the barrier height is described by the traditional Schottky-Mott model of the metal-semiconductor interface. Any small changes in Φ_M, corresponding to contact with different metals for example, will be reflected in S_b.

The smaller the dimensions of the semiconductor, the greater the surface area to volume ratio of the sample and the larger the effect of these surface states. This has been exploited to the advantage of gas sensing; there has been much interest in the literature for example in the gas sensing abilities of SnO_2 - Ref. [41] for instance focuses on the increase of sensitivity of nanocrystalline tin dioxide gas sensing films over coarser grained films.

The increased reactivity of the 'exposed' surface of tin dioxide constituent atoms increases the likelyhood of gaseous atoms and molecules from the surrounding atmosphere being absorbed at the SnO_2 surface. Weak Van der Waals dipole-dipole interactions can cause physisorption, a very weak reaction and one unlikely to lead to any charge transfer between the gas particle and the semiconductor. Some gaseous species however can form a strong chemical bond with the surface, known as chemisorption, with charge transfer between absorbent (that which absorbs – the SnO_2 crystal) and the absorbate (the absorbed substance – the gas species). Ionosorption is specifically chemisorption with electron transfer to/from the conduction band, the absorbed gas acting as a surface state on the semiconductor donating/accepting electrons and as such, altering the surface electronic behaviour.

Oxygen species (e.g. O^-, O_2^-) can be ionosorbed onto the tin dioxide surface, where they act as surface acceptor states. These create, or enhance, the surface barrier and cause a depletion layer (a spatial region depleted of charge carriers) to form, which can penetrate deeply into the nanocrystal. Exposure now of the gas sensor to a reducing gas, such as carbon monoxide (CO), acts to the reverse, releasing electrons into the nanocrystal and lowering the surface barrier by removing the ionosorbed surface oxygen.

It should be noted also that any charged chemisorbed surface complex will induce a dipole

moment, as in the case of the metal-semiconductor interface, causing a voltage drop across the surface, V_{il}.

The sensitivity of a gas sensor is the ratio of its resistance in air and its resistance after exposure to the target gas. The SnO_2 gas sensing films comprise of a layer of polycrystalline grains through which a current is passed, electrons travelling across the interconnecting grain boundaries. Indeed, without interconnections, any current flow would only be possible through the surface states in physical contact, quantum mechanical tunnelling or thermionic emission. The interconnections form during one crucial phase of the nanoparticles manufacture – annealing. Mentioned in Section 1.3, annealing (heating and gradually cooling) encourages the formation of oxygen vacancies, but also creates sinter necks – the above-mentioned narrow joins at the grain boundaries. Kennedy *et al.* [42] report that the sinter neck growth can be described by

$$\frac{dR_{sin}}{dt} = kR^{n-m+1}R_{sin}^{-n} \tag{2.6}$$

where R_{sin} is the radius of the sinter neck, R the radius of the particle (strictly before sintering, but assumed to be indistinguishable from is sintered size here) and the m and n exponents are related to the type of diffusion forming the neck (lattice diffusion $m = 3$, $n = 3.78$, surface diffusion $m = 4$, $n = 5$). The constant k is dependant on several factors such as the surface tension, the diffusion coefficient of the material and the temperature. Integrating, and imposing the boundary condition that at $t = 0$ no sinter neck exists, then

$$R_{sin} = \left((n+1)kR^{n-m+1}t\right)^{\frac{1}{n+1}} \tag{2.7}$$

If it is assumed that the sintering parameters and conditions, including duration of annealing, are maintained constant over a range of radii, then it can be inferred from Eq. (2.7) that the larger the grain, the larger the sinter neck, but the smaller the ratio (R_{sin}/R)[iii]. Therefore, the smaller the grain the larger the effect of the annealing.

Regardless of neck size, the sintering procedure vastly improves the charge transport process, Ref [42] reporting that for 35nm diameter samples the resistance before sintering of 700 kΩ drops to 1 kΩ afterward. The annealing procedure and the formation of the sinter necks can be considered to allow conductance through three general mechanisms [43], represented in Figure 2-4.

[iii] To see this, arbitrarily choose $m = 3$, $n = 3.78$ and set tk to 1×10^{-6}, then for $R = 1$nm, $R_{sin} = 0.448$nm and so $R_{sin}/R = 0.448$, and for $R = 100$nm, $R_{sin} = 2.49$nm and therefore $R_{sin}/R = 0.025$.

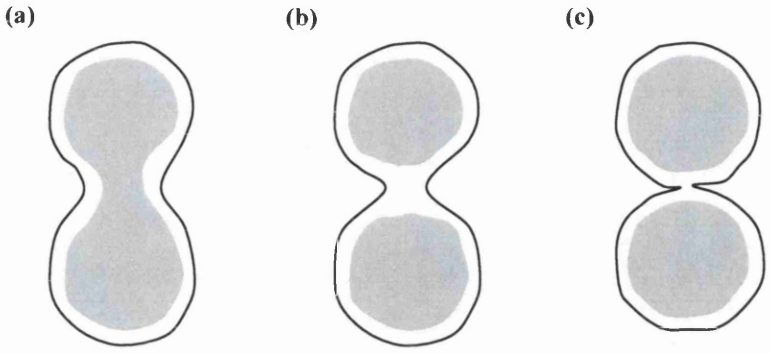


Figure 2-4 Illustration of the three conductance models for charge transport between two sintered grains (a) Open neck (b) Closed neck and (c) Schottky barrier

The first, Figure 2-4(a), has a junction described as 'open' necked. The grain conductivity is largely that of the region un-depleted of charge carriers in the centre of the neck and would be determined by the energy required to promoted electrons from the donor vacancies (referred to as the activation energy) and the area of the 'effective' channel connecting the grains. The second scenario, illustrated in Figure 2-4(b), is a closed neck junction. The depletion zones of the grains overlap, isolating the non-depleted regions of each grain. This is a higher resistance case than Figure 2-4(a) and can be caused by less complete sintering or surface state depletion of electrons. The conductivity would be a function of the activation energy of the surface states and the occupancy of those surface states. The third diagram, Figure 2-4(c) is that of a Schottky-like barrier, charge transport having to occur 'over' the surface barrier.

The dominant process would depend on the size of the constituent particles, the degree of sintering and the type of film. Thick (or porous) films (>1μm) would allow all three mechanisms, while only the first two would be prevalent in thin film sensors[iv] [43]. Conduction through the films is a complicated process however, with more than one 'type' of contact in evidence. Structural inhomogeneities may even offer lower resistances 'paths' through the film and fine porosity may limit the penetration of gas into the sensor, defining domains, such that conduction between these domains forms the dominant transport process. Consequentially, an exact model of the conduction process is complex indeed, if not impossible.

[iv] Although it is suggested by Beekmans [44] that, even in open/closed neck conductance, intergrain barriers may have a role to play. It is proposed that the poor alignment of the crystal lattice between adjacent grains will give rise to a barrier within the sinter neck.

Nonetheless, for all three mechanisms it is clear that the greater the charge in the surface states the greater the resistance of the film, the surface states acting to deplete the grain of its mobile charge carriers, and for the Schottky mechanism, also forming a surface barrier which the electrons must overcome to conduct. As such, the smaller the grain, the greater the influence of the surface states and the more sensitive a gas sensor made up of such small grains becomes.

Furthermore, it has been reported [41,42,45,46,47] that the sensitivity of the sensing film increases with the decrease of grain size, particularly below a diameter of ~20nm [41,42], at a rate greater than that which would be expected from the increase in surface area to volume ratio alone, and it has been suggested that this might be evidence of a certain amount of Fermi level unpinning [31]. To understand this, consider that as the volume of the grain diminishes, then although the surface area to volume ratio increases, the actual density of occupied surface states must decrease to maintain the neutrality of the grain. Thus from Eq. (2.4), as the surface charge density decreases γ_{gs} will increase. Examination of Eq. (2.5) reveals that

$$\gamma_{gs} = \frac{\delta S_b}{\delta \Phi_M} \tag{2.8}$$

ergo, as γ_{gs} increases, so does the dependence of the surface barrier height on the work function of the environmental gas. If the dominant conductance mechanism in the film is over the Schottky-like barriers, then the sensitivity of the semiconductor sensor will also increase. As the barrier height is now able to change, then evidentially, the Fermi level also has more freedom to move in response to the ionosorption of the target gas and it is said to become 'unpinned'.

To summarise then, surface states arise from the free or 'dangling' bonds of surface ions of a lattice. These states can be enhanced by chemical bonds; for example by chemisorbed oxygen species bonding with the surface of a tin dioxide nanocrystal. Electrons can be thought of as draining from the conduction band into these surface states, giving rise to a surface barrier and a spatially charged, or depleted, region extending into the semiconductor. For nanoparticles of semiconductor, if the sinter neck between adjoining grains is sufficiently small, then transport across this surface barrier is the dominant conduction process between the grains. As such, the sensitivity of a gas sensing film composed of such grains is dependent upon the response of this barrier to changes in the environmental gasses. It can be seen that the lower the density of surface states the freer the Schottky-like barrier height becomes to move. This is often referred to as the unpinning of the Fermi level. Indeed, regardless of the exact conduction mechanism, it is the surface states which, as a general rule, control the sensitivity of the sensor. The smaller the

grains, the greater the surface area to volume ratio and the larger the effect these states have over the electronic behaviour of the grains.

A detailed treatment of the band bending cause by these surface states in the nanoparticle regime follows in the next two sections.

2.2 The Complete Charge Density Model (CCDM)

Inherent in any theoretical model of a physical system such as this, there is some level of approximation. Here it is assumed that the quasi-spherical grains are completely spherical and uniform enough in composition that experimentally measured quantities such as S_b and n_d have meaning for the model. In this way, a complete charge density model (CCDM), similar to that of Malagú *et al.*[15], can be applied to one spherical grain and its properties can be taken as representative of those of all the nanoparticles on the film. In keeping with the literature, any effects of strong sintering are neglected, and electron transfer between grains is taken to be dominated by the Schottky-like surface barrier.

As mentioned previously, SnO_2 is an *n*-type semiconductor due to the presence of oxygen vacancies in the lattice structure. It has a wide direct band gap, between ~3.57eV [48] and 3.86eV [49] [i,ii], and as such it is practical to represent its charge density as

$$\rho(r) = e\left(n_d - N_c \mathscr{F}_{1/2}\left(\frac{E_f - v_p(r)}{k_B T} \right) \right) \tag{2.9}$$

using the expression for the bulk electron density in the conduction band, Eq. (1.34), and where the bottom of the conduction band is given by the potential $v_p(r)$, see Figure 2-5. n_d represents the ionised density of donors at the operating temperature of the sensor, and the Fermi level is chosen as the energetic minimum of the system i.e. $E_f = 0$. The potential itself will follow from the solution of Poisson's equation (see Section 5.2 for a derivation of this)

$$\nabla^2 v_p(r) = \frac{e\rho(r)}{\varepsilon_0 \varepsilon_r} \tag{2.10}$$

although, being a variable of the charge density, Eq. (2.10) is a non-linear equation, complicating the matters of its solution; this will be addressed in Section 2.3. The usual Poisson 'potential' φ of Section 5.2 is related to $v_p(r)$ by $\varphi(r) = - v_p(r)/e$ (the potential $\varphi(r)$ not to be confused with the charge neutrality position ϕ_0).

---

[i] The difference in the measured values of the band gap probably due to stoichiometric issues as the number of oxygen vacancies may vary between manufacturing processes [49].

[ii] Maffeis *et al.* [31] report a surface band gap of ~2.5eV on their nanocrystalline particles, suggesting perhaps the influence of surface states altering the 'bulk' properties.

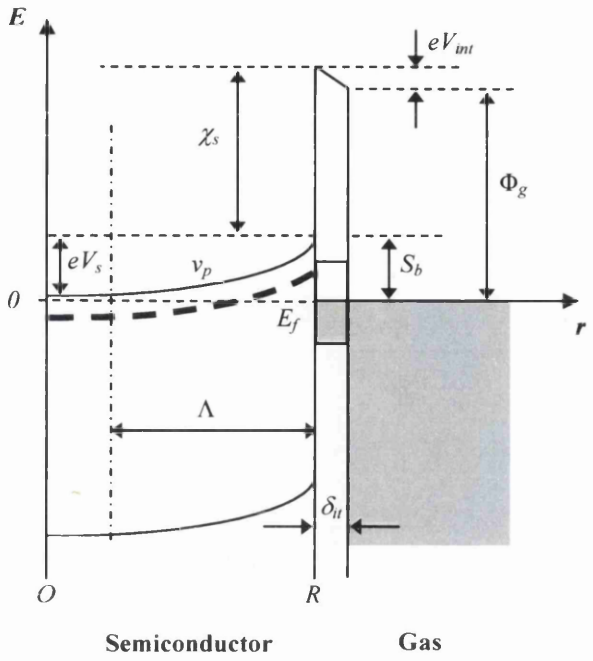


Figure 2-5 Schematic of SnO_2 gas-semiconductor interface

The boundary conditions of the differential equation require that at the surface of the grain the potential is equal to the surface, Schottky-like, barrier height S_b, and that being spherically symmetric, the gradient of v_p at the origin must be zero i.e.

$$v_p(R) = S_b$$
$$\left.\frac{dv_p}{dr}\right|_{r=0} = 0 \tag{2.11}$$

The gradient of the conduction band at the grain boundary forms a measure of the electric field across that boundary, and, as such, can be used to determine the charge in any surface states (see Section 5.2); defining the occupied surface state density N_s as the uniform charge per unit area on the grain surface (hence avoiding the complicated issue of ascertaining the neutral level of the surface states (see Tung [39]) as would be required if N_{ss}, the charge per unit area per electron volt, were used) then

$$N_s = -\frac{\varepsilon_0 \varepsilon_r}{e^2}\left.\frac{dv_p}{dr}\right|_{r=R} \tag{2.12}$$

N_s also includes the 'intrinsic' surface states, neglected in N_{ss}.

Therefore, solving the non-linear Poisson equation Eq. (2.10) in conjunction with the boundary

values Eq. (2.11), given an experimentally measured S_b and n_d for the required operating temperature, then the band bending and surface state density N_s can be assessed.

Importantly, this band bending will remain in evidence whilst the oxygen species remain chemisorbed; once the surface states form and the band bending develops the removal of the surrounding gas e.g. placing of the grains in a vacuum, will not affect the band structure. In this way the nature of these surface states will play a pivotal role for any grains prepared in an oxygen rich environment, even if they later go on to be charge injected in a vacuum. Hence, this modelling of the gas-semiconductor interface and assessment of the formed surface states is not limited in relevance purely to gas sensing issues, but key to the simulation of the charge writing process.

2.3 Numerical Methods for the Solution of the Non-Linear Poisson Equation

The Poisson equation, as given by Eq. (2.9), is non-linear; that is to say that the dependence of the right hand side of the equation on φ means that it is not a linear differential equation in φ and its derivatives. This equation does not have a simple analytical solution and numerical techniques must be employed to determine φ subject to the boundary conditions given by Eq. (2.11).

2.3.1 The Taylor Expansion Method

The first stage in this process is to develop a method to solve Eq. (2.9) assuming that both $\varphi(0)$ and $d\varphi/dr\big|_{R=0}$ are known, $d\varphi/dr\big|_{R=0}$ taking its usual value from Eq. (2.11). As such, let the spatial region of interest be divided into a 1-dimensional mesh of N evenly spaced points, separated by a distance Δ and beginning at zero. It is then possible to evaluate a continuous function f in this discrete space about each lattice point in terms of its previous values; f can be evaluated at site r_{i+1}, abbreviated here as f_{i+1}, by expanding it as a Taylor series about the site r_i

$$f_{i+1} = f_i + \Delta f_i' + \frac{\Delta^2}{2!} f_i'' + \frac{\Delta^3}{3!} f_i''' + \frac{\Delta^4}{4!} f_i^{iv} + \frac{\Delta^5}{5!} f_i^{v} + \cdots \tag{2.13}$$

f' denoting the first derivative of f with respect to r and so forth. By similar means it is also possible to evaluate f at r_{i-1}

$$f_{i-1} = f_i - \Delta f_i' + \frac{\Delta^2}{2!} f_i'' - \frac{\Delta^3}{3!} f_i''' + \frac{\Delta^4}{4!} f_i^{iv} - \frac{\Delta^5}{5!} f_i^{v} + \cdots \tag{2.14}$$

In the determination of v_p, once v_p and v_p' are known on a particular mesh point, v_p'' can be found from the Poisson equation

$$v_p''(r) = -\frac{2}{r} v_p'(r) + \frac{e\rho(v_p)}{\varepsilon} \tag{2.15}$$

For the remainder of this section, v_p will be represented as v, the position of the p subscript sequestered to represent the mesh site at which the potential is being evaluated on.

On the first mesh point, $i = 0$, the radius is zero and the values of v_0 and v_0', and so v_0'', are known from the boundary values:

$$\begin{aligned} v_0 &= \alpha \\ v_0' &= 0 \end{aligned} \tag{2.16}$$

and

$$v_0'' = \frac{e\rho_0(\alpha)}{\varepsilon} \tag{2.17}$$

The values of v_1 , v_1' , and v_1'' are a little more difficult to obtain. For v on r_1 adding the two Taylor expansions Eq.'s (2.13) and (2.14) yields

$$f_{i+1} + f_{i-1} = 2f_i + 2\frac{\Delta^2}{2!} f_i'' + 2\frac{\Delta^4}{4!} f_i^{iv} + \cdots \tag{2.18}$$

Realising that $f_1 = f_{-1}$ in this symmetry and letting $f = v$ then

$$v_1 = v_0 + \frac{\Delta^2}{2!} v_0'' + E \tag{2.19}$$

where v_0 and v_0'' are both known from Eq. (2.16). The error in the expression, E, is fourth order in Δ and can be written

$$E = \frac{\Delta^4}{4!} v_0^{iv} + O\left(\Delta^6\right) \tag{2.20}$$

Subtracting Eq. (2.14) from Eq. (2.13) then

$$f_{i+1} - f_{i-1} = 2\Delta f_i' + 2\frac{\Delta^3}{3!} f_i''' + 2\frac{\Delta^5}{5!} f_i^{v} + \cdots \tag{2.21}$$

Employing the symmetry about the origin, letting $f' = v^{iii}$ in Eq. (2.21) and substituting into Eq. (2.13) with $f = v'$, then

$$v_1' = v_0' + \Delta v_0'' + E \tag{2.22}$$

where

$$E = \frac{\Delta^3}{6} v_0^{iv} + O\left(\Delta^4\right) \tag{2.23}$$

and the error is of third order. v_1'' follows from v and v' in Eq. (2.15) and with the error in v_1' dominant, will also be in error to the third order in Δ (the effect of the $1/r$ term will increase the error in v'' quite considerably near the origin, but less so at the further reaches of the mesh).

For the value of v and its derivates on the third lattice site, $i = 2$, expand about the second lattice site, $i = 1$. Rearranging Eq. (2.14) so that

$$f' = \frac{f_i - f_{i-1}}{\Delta} + \frac{\Delta}{2!} f_i'' - \frac{\Delta^2}{3!} f_i''' + \cdots \tag{2.24}$$

then letting $f' = v^{iii}$ and substituting into Eq. (2.13) yields

$$v_2 = v_1 + \Delta v_1' + \Delta^2\left(\frac{2}{3}v_1'' - \frac{1}{6}v_0''\right) + E \tag{2.25}$$

on letting $f = v$. E is of order Δ^4 and given by

$$E = \frac{\Delta^4}{12}v_0^{iv} + O\left(\Delta^5\right) \tag{2.26}$$

To maintain the error in the potential at fourth order, v_2' is required with its error proportional to $O(\Delta^3)$. Setting $f = v'$ in Eq. (2.13) and $f' = v^{iii}$ in Eq. (2.24) implies that

$$\begin{aligned} v_2' &= v_1' + \Delta\left(\frac{3}{2}v_1'' - \frac{1}{2}v_0''\right) + E \\ E &= \frac{5}{12}\Delta^3 v_1^{iv} + O\left(\Delta^4\right) \end{aligned} \tag{2.27}$$

with E to third order as needed. The second derivative on $i = 2$, v_2'', again follows from Eq. (2.15) and with a $O(\Delta^3)$ error.

For the values of v and its derivates on the lattice sites $i = 3$ to N similar approaches to the above can be applied. Again setting $f' = v^{iii}$ in Eq. (2.24) and substituting into Eq. (2.13) then

$$v_{i+1} = v_i + \Delta v_i' + \Delta^2\left(\frac{2}{3}v_i'' - \frac{1}{6}v_{i-1}''\right) + \frac{3}{24}\Delta^4 v_i^{iv} + \ldots \tag{2.28}$$

Now let $f_i = v_{i-1}^{iv}$ in Eq. (2.13) so that

$$v_i^{iv} = v_{i-1}^{iv} + \Delta v_{i-1}^{v} + \ldots \tag{2.29}$$

and with Eq. (2.18) arranged so that

$$f_i'' = \frac{f_{i+1} - 2f_i + f_{i-1}}{\Delta^2} + \frac{\Delta^4}{12}f_i^{iv} + \cdots \tag{2.30}$$

then setting $f'' = v^{iv}$ in Eq. (2.30) and using Eq.(2.29), Eq.(2.28) can now be written

$$v_{i+1} = v_i + \Delta v_i' + \Delta^2\left(\frac{19}{24}v_i'' - \frac{5}{12}v_{i-1}'' + \frac{1}{8}v_{i-2}''\right) + E \tag{2.31}$$

where the fifth order error is

$$E = \frac{19}{180}\Delta^5 v_i^{v} + O\left(\Delta^6\right) \tag{2.32}$$

Letting $f = v'$ in Eq. (2.13) and substituting in Eq. (2.30) with f'' once again set at v^{iv}, then

$$v_{i+1}' = v_i' + \Delta\left(\frac{23}{12}v_i'' - \frac{4}{3}v_{i-1}'' + \frac{5}{12}v_{i-2}''\right) + E \tag{2.33}$$

where the error E is given to fourth order by

$$E = \frac{9}{24}\Delta^4 v_i^v + O(\Delta^5) \tag{2.34}$$

As usual v_i'' will follow from Eq. (2.15) with E proportional to $O(\Delta^4)$.

In actual practice the formulas for v and its derivative on $i = 2$, Eq.'s (2.25) and (2.27), are replaced with Eq. (2.31) and Eq. (2.33), exploiting the symmetry of the potential once more:

$$v_{i+1} = v_i + \Delta v_i' + \Delta^2\left(\frac{11}{12}v_i'' - \frac{5}{12}v_{i-1}''\right) + E \tag{2.35}$$

with E given to fifth order by Eq. (2.32) , and

$$v_{i+1}' = v_i' + \Delta\left(\frac{7}{3}v_i'' - \frac{4}{3}v_{i-1}''\right) + E \tag{2.36}$$

with E given to fourth order by Eq. (2.34). In this way, the truncation error of the computed solution v is of fifth order in Δ, although the first 'starting step' is of $O(\Delta^4)$. Due to the very small gradient towards the origin, the effect of this lower order mesh point is minimal. Indeed, this can be seen on the density N_t, calculated from the gradient of v at R through Eq. (2.12); on replacing Eq.'s (2.25) and (2.27) with Eq.'s (2.31) and (2.33) and so exchanging an error of $O(\Delta^4)$ with one of $O(\Delta^5)$, the surface state density alters by a negligible 6×10^{-4} %.

In general then, the truncation error of this method is comparable with the common Runge-Kutta methods (see for instance Kreyszig [50], although it should be noted that there are higher order Runge-Kutta methods available e.g. the Runge-Kutta-Felhberg method). The advantage the Runge-Kutta methods have over the Taylor expansion approach is both that they are self-starting, and that propagation errors (instability) are lower as each step is begun anew, all old information discarded. However, their disadvantages are that they are very computationally expensive (four derivative evaluations on each step compared with the Taylor method's two – extremely costly when this method is iterated within self-consistency cycles – see Chapters 2 and 4), and relatively inefficient (discarding all old data on each step). In addition, while they offer exact solutions for polynomials of degree less than or equal to four, about the origin they can be quite inaccurate, much less so than a power series method, see Hamming [51]. Consequentially, although one such Runge-Kutta method is employed by Ref. [15] to solve Eq. (2.10), they shall not be used here. Propagation errors arise through the computed solution differing from the exact solution and this difference being fed into the following computational steps, the total 'propagative' error gradually accumulating on each mesh point until site $i = N$ is reached. If unaware of the true, exact,

solution, this kind of error if difficult to quantitatively assess. However, the manipulation of the forms of Eq.s (2.19), (2.22), (2.31), (2.33), (2.35) and (2.36) to include significant quantities of data from previous mesh points to maximise efficiency has been deliberate. In particular, the decision to input this data in the form of the second derivatives i.e. $A_0 v_i''$, $A_1 v_{i-1}''$ and $A_2 v_{i-2}''$, rather than through preceding values of the function v itself i.e. $B_0 v_i$, $B_1 v_{i-1}$ and $B_2 v_{i-2}$, has been intentional. While the derivation of a specific relationship between truncation error (previously E, now designated the symbol E_T) and the propagation error E_P is unobtainable, it is certain that E_T influences E_P. If multiple v values are used per step, say n in number, then the accumulated error at each step will be broadly proportional to nE_T^v i.e. $O(n\Delta^5)$. If however, multiple v'' values are used per step, say n in number, then the propagative E_P would be expected to be proportional to $n\Delta^2 E_T^{v''}$ i.e. $O(n\Delta^6)$, the $O(\Delta^4)$ error of v' dominating over the $exp(\Delta^5)$ error through v.

A third kind of numerical error arises through round-off errors during computation; for example the fraction 1/3 becoming the computer's truncated decimal 0.333.....333. These are likely to be negligible in comparison to the truncation error.

The number of mesh points in a given interval is proportional to $1/\Delta$, and so in general, a computational method, such as the above, with a truncation error of $O(\Delta^5)$ is said to have a total, or global, error of $O(\Delta^4)$.

While this Taylor expansion method has been specifically created for this problem and optimised by the author to deliver high accuracy and stability over quite large meshes against computational speed, there is a family of Taylor expansion based methods in the literature (see for example Hamming [51]), known as the Predictor-Corrector methods, already 'tuned' for a balance between truncation accuracy and stability. These equations use one expression to estimate, or 'predict', the required value of the function at the $i+1^{th}$ mesh point, typically using several values from the prior i, i-1, i-2... mesh points, usually those of f and f'. The next action is to use a second series expansion expression, which like first, uses preceding data but crucially uses the first estimate of the function value on the $i+1^{th}$ mesh point to create an improved, or 'corrected', evaluation of that same point.

These methods are also not self-starting but require the initial mesh points to be assessed via alternative means (Taylor expansions or Runge-Kutta methods). Applying the Adams-Moulton Predictor-Corrector, global error of $O(\Delta^4)$ as described in Kreyszig [50],

$$\begin{aligned} p_{i+1} &= f_i + \frac{\Delta}{24}\left(55f_i' - 59f_{i-1}' + 37f_{i-2}' - 9f_{i-3}'\right) \\ c_{i+1} &= f_i + \frac{\Delta}{24}\left(9p_{i+1}' + 19f_i' - 5f_{i-1}' + f_{i-2}'\right) \\ f_{i+1} &= c_{i+1} \end{aligned} \tag{2.37}$$

to Eq. (2.10), with the derivative of the predictor given by

$$p_{i+1}' = \frac{p_{i+1} - f_n}{\Delta} + \Delta\left(\frac{9}{8}f_n'' - \frac{11}{12}f_{i-1}'' + \frac{7}{24}f_{i-2}''\right) + O\left(\Delta^4\right) \tag{2.38}$$

calculated via the usual Taylor expansion approach as shown below in footnote [i] of this section, then below ~60 nm on a 250 point mesh for S_b = 1.35 eV and n_d = 1.5×10^{24}m^{-3}, there is no perceptible difference between N_t ascertained with this or with the Taylor expansion method. After this point, errors do begin to appear, growing to ~0.15 % by 70nm and ~0.6 % by 144nm – the radius at which the Taylor expansion method fails to converge. Convergence is possible for another 4nm using the Adams-Moulton approach.

However, the predictor-corrector method is slightly computationally more expensive than the Taylor expansions, involving twice as many evaluations of the potential and one extra evaluation of the first derivative for each mesh point. When used in conjunction with the charge balance equation (see Chapters 5 and 6) to recreate the results of Figure's 6-1 and 6-2, it is found that it does not perform as well as the Taylor method. At 100nm its N_t lies 0.2 % above the Taylor

---

[i] From Eq. (2.13)

$$-\Delta f_i' = f_i - p_{i+1} + \frac{\Delta^2}{2!}f_i'' + \frac{\Delta^3}{3!}f_i''' + \frac{\Delta^4}{4!}f_i^{iv} + \frac{\Delta^5}{5!}f_i^{v} + \cdots$$

which, with the use of Eq.(2.24) and Eq. (2.30) to estimate the third and fourth derivatives of f respectively, can be written as

$$-\Delta f_i' = f_i - p_{i+1} + \Delta^2\left(\frac{19}{24}f_i'' - \frac{5}{12}f_{i-1}'' + \frac{1}{8}f_{i-2}''\right) + \frac{19}{180}\Delta^5 f_i^{v} + \cdots$$

Also from Eq. (2.13) letting $f = f'$ and again employing Eq.s (2.24) and (2.30) it is possible to derive

$$p_{i+1}' = f_i' + \Delta\left(\frac{23}{12}f_i'' - \frac{4}{3}f_{i-1}'' + \frac{5}{12}f_{i-2}''\right) + \frac{13}{24}\Delta^4 f_i^{v} + \cdots$$

Ergo,

$$p_{i+1}' = \frac{p_{i+1} - f_n}{\Delta} + \Delta\left(\frac{9}{8}f_n'' - \frac{11}{12}f_{i-1}'' + \frac{7}{24}f_{i-2}''\right) + O\left(\Delta^4\right)$$

expansions rising to 3 % above by 600nm, but this is the last point of convergence possible with the Adams-Moulton method, and generally, its performance within this range is more temperamental than the alternative. This could possibly be due to the increased truncation error for larger radii (following from the increase in mesh spacing) contributing to greater instability owing to the extra computational steps. Consequentially, it is the Taylor expansion method that has been used throughout this work as being the best all-round method to solve the Poisson non-linear differential equation under these circumstances and operating parameters.

2.3.2 The Bisection Method

Having decided on a means to solve the differential equation Eq. (2.9) the problem now becomes solving it with the relevant boundary conditions of Eq. (2.11), as opposed to the pseudo-boundary values assumed through Eq. (2.16), the value of the potential at the origin, v_0, of course not known.

Aquainted with the required value of v on the radius of the grain (S_b), if its initial value is estimated, e.g. as α, and using the known initial value of the first derivative of v, then running the chosen Taylor series method over the mesh, the computed value of $v(R)$ compared to S_b will give an indication of the accuracy of α. This is the essence of a 'shooting' method.

To solve Eq. (2.9) with Eq. (2.11) then three widely spaced first estimates of v's initial value are chosen: α_1, α_2, and α_3. A 'shoot' with each is then made, running the Taylor expansion method over the mesh from 0 to R; the two 'shoots' which lie either side of, or 'bracket', the known value of the potential at R are then selected and the mid-point of their starting values calculated. This is then used as the starting point of another shoot, discarding the redundant data of the non-bracketing shoot from the first three shoots. From here the procedure is repeated; the mid-point of the two initial conditions whose shoots bracket the known end point are used to produce another shoot, and so on until the initial value paired to the required end value is found to the required precision. This is demonstrated schematically in Figure 2-6.

This method of halving the interval known to contain the initial value generating the required end value is essentially an application of the bisection method, the zero point of the function

$$F(\alpha) = v(R,\alpha) - S_b \tag{2.39}$$

being sought, the potential v's initial value α the variable. Some authors prefer the Newton-Raphson approach[ii] [52], perhaps feeling, as does Ref. [50], that the bisection method is slow. However, as detailed in Hamming [51], although the final rate of convergence is faster with the Newton-Raphson method, it is often the initial rate of convergence that is most important and determines the overall speed at which a solution is achieved. In this, the bisection method leads.

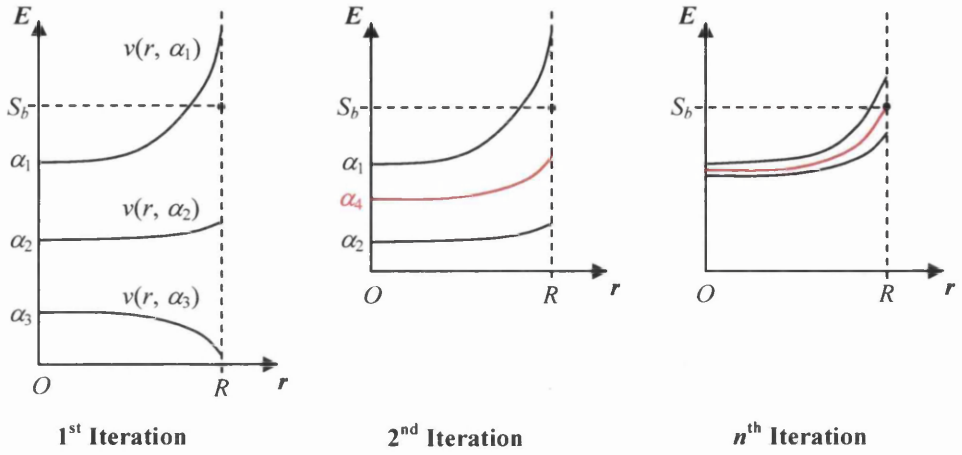


Figure 2-6 Representation of the Shooting-Bisection method. 1st iteration: three shoots from the three initial values α_1, α_2 and α_3. 2nd iteration: mid-point of two bracketing shoots (from α_1 and α_2) used for new shoot (highlighted in red), α_3 shoot discarded. n^{th} iteration: process repeated until $v(R, \alpha_n)$ equals S_b to required precision.

The Newton-Raphson method would also require the computationally expensive evaluation of $\partial F/\partial \alpha$, and is likely to be slow in converging if $|\partial F/\partial \alpha|$ is very small near $F = 0$. If $F = 0$ is an inflection point in F, as is quite likely, then the Newton-Raphson method can run into convergence problems [51], while the bisection method is very robust. This bisection method is also used in Chapters 5 and 6 for an additional purpose: finding the point of charge equality varying the ionised donor density or Fermi level. Here, the sturdy nature of the method is even more significant, discontinuities frequently present in what is often a quite ill conditioned function. This is discussed in more detail in the relevant Chapters.

Employed together to solve Eq. (2.9) with Eq. (2.11), the Taylor-bisection methodology has a global error of $O(\Delta^4)$ and usually takes between 20 and 25 iterations to achieve a computed value of v at R accurate to within 5×10^{-6} % of its specified value. Detailed discussions of the results of the solved Poisson equation are to be found in the next section, Section 2.4.

[ii] Often simply called Newton's Method.

2.4 The Complete Charge Density Model (CCDM) as Applied to Spherical SnO_2 Nanocrystals

This section assesses the developed numerical techniques and appraises the merit of the complete charge density model against the available SnO_2 gas sensing data.

On solving the non-linear Poisson equation Eq. (2.8) via the methods of Section 2.3 for the conduction band bottom, v_p, setting S_b to 1eV and n_d to$1\times10^{25}m^{-3}$, then, at room temperature and with a grain radius R of 50nm, Figure 2-7 results.

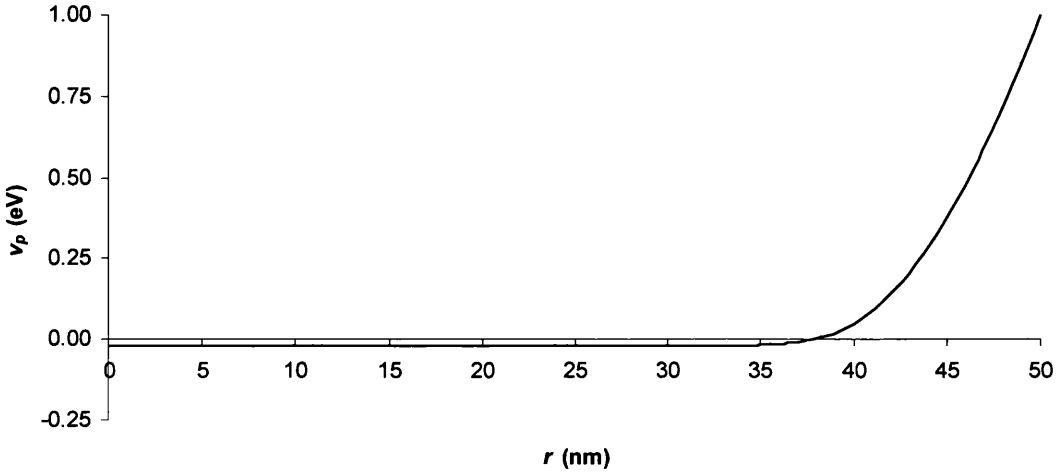


Figure 2-7 Plot of conduction band bottom, $v_p(r)$, from the Taylor-Bisection method solution of the non-linear Poisson equation for a R = 50nm SnO_2 nanocrystal. S_b = 1eV, n_d =$1\times10^{25}m^{-3}$, room temperature.

The density of states permittivity $\varepsilon_r\varepsilon_0$ has been set at ~$10^{-10}Fm^{-1}$ [43] from $\varepsilon_\perp = 14\varepsilon_0$ and $\varepsilon_\parallel = 9\varepsilon_0$ [53], and likewise for the density of states effective mass, $m^* = 0.275m_0$ from $m_\perp^* = 0.299m_0$ and $m_\parallel^* = 0.234m_0$ [54]. The energetic zero of the system is the position of the Fermi level.

As would be expected in a grain of this size there is a considerable flat band region where

$$n_d = N_c\mathscr{F}_{1/2}\left(\frac{E_f - v_p(r)}{k_BT}\right) \tag{2.40}$$

extending from the origin to R_0, where $R_0 \approx$ 36nm, and a space charge region, a region depleted of charge carriers the electrons having moved into the surface states, extending from R_0 to R.

Conventionally [15] this depleted region, Λ, or potential extinction length, is considered to comprise of two distinct parts, a region which is almost depleted Λ' and the Debye length[i] λ_D. The total depletion width is the distance from the grain edge to the point at which $\rho = 0$; at this point the electron density is considerable and such that it equals the density of donors, however, the region of space up to this point is also considered to be depleted of electrons. To resolve this, the depletion width Λ' is defined as the region of space over which the electron presence can be taken as negligible and the Debye length, given by

$$\lambda_D = \sqrt{\frac{\varepsilon_0 \varepsilon_r T}{e^2 n}} \tag{2.41}$$

with the electron density n set at n_d according to Eq. (2.40), describes the spatial region of intermediate charge between the Λ and Λ'. At room temperature, for SnO_2 with an ionised donor density of 1×10^{23} m^{-3} to 1.5×10^{25} m^{-3}, then the Debye length will lie between 12.1nm and 0.9nm. From this definition of the depletion width, if $\Lambda \gg \lambda_D$ it is implied that a reasonable approximation of the band bending can be made by totally neglecting the mobile charge carriers between R_0 and R, Eq. (2.8) becoming simply the linear Poisson equation

$$\nabla^2 v_p(r) = \frac{e^2 n_d}{\varepsilon_0 \varepsilon_r} \tag{2.42}$$

which can be analytically solved. This approximation is known as the depletion approximation (DA) and is put into use in Chapter 6 and briefly in Chapter 5.

Returning to the complete charge density model, if the grain radius R is now reduced below the R_0 of the R = 50nm case, say R = 5nm, then maintaining S_b and n_d, Figure 2-8 results. The conduction

[i] The Debye length is a quantity from plasma physics. In order to understand what the Debye length is, the reader must know a little about plasma. A plasma is defined loosely as an electrically conducting medium in which there are approximately equal numbers of positively and negatively charged particles. Each particle assumes a position such that the total force resulting from all the particles is zero, thus producing a uniform neutrally charge state. If a negative particle, say an electron, is displaced from its equilibrium position, the equilibrium position itself takes on a positive charge and exerts an electrostatic attraction on the electron, causing the electron to oscillate about this equilibrium position. As the interaction between electrons is strong in this phase of matter, they will oscillate collectively at a characteristic frequency depending upon the nature of the particular plasma; these are known as plasma oscillations. A time τ is required for such an oscillation and the Debye length, λ_D, is the distance travelled by the average thermal electron in time $\tau/2\pi$. A plasma can be more precisely defined in terms of this parameter as a partially or fully ionized gas in which constituent electrons may complete many plasma oscillations before they collide with an ion, and that inside each sphere with a radius equal to the Debye length there are many particles, and finally, that the plasma itself is much larger than the Debye length in every dimension.

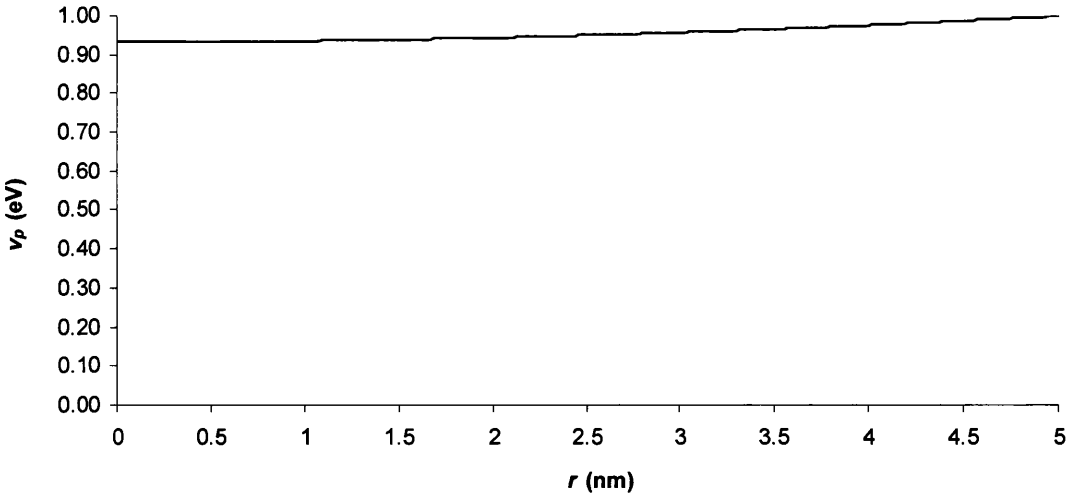


Figure 2-8 Plot of conduction band bottom, $v_p(r)$, from the Taylor-Bisection method solution of the non-linear Poisson equation for a R = 5nm SnO_2 nanocrystal. S_b = 1eV, $n_d = 1\times10^{25}m^{-3}$, room temperature.

band bottom rises up in energy, the curvature of v_p flattening. This has been reported and discussed in References [15,16,17] for example.

Consider now a gas sensing film. Taking the identical spherically symmetric grains to be in Schottky contact with each other, then for $R > \Lambda$, as schematically illustrated in Figure 2-9 for a chain of grains, the band bending in each is fully evolved and the barrier which the electrons have to overcome to conduct is $eV_s = S_b - v_p(0)$, often called the built-in potential V_b.

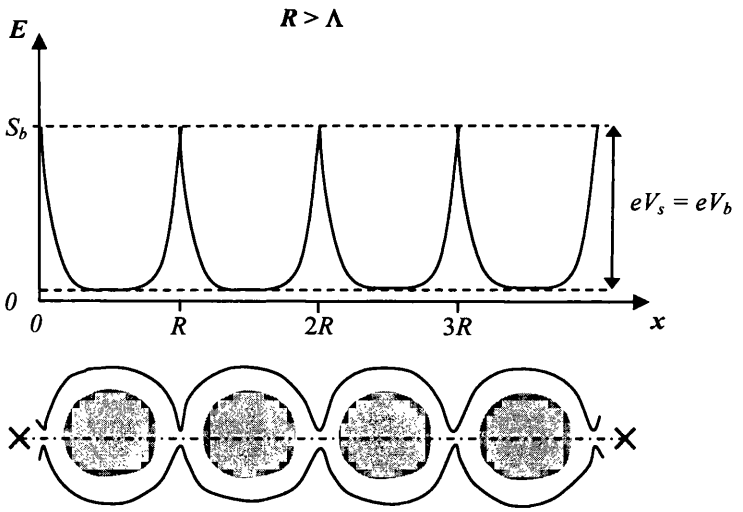


Figure 2-9 Energy diagram representation of a chain of $R>\Lambda$ sintered grains.

The conductivity across a Schottky, or Schottky-like, barrier is given by

$$G = G_0 e^{-eV_s / k_B T} \tag{2.43}$$

G_0 is generally treated as a constant, and is taken to be dependent on several factors including the electron mobility [55] and the electron density [55,56].

On reducing the grain radius such that the whole grain is depleted of charge carriers, then flattening of the band bending occurs, see Figure 2-10, and $eV_s = S_b - v_p(0)$ is no longer the built-in potential V_b, but much lower in energy.

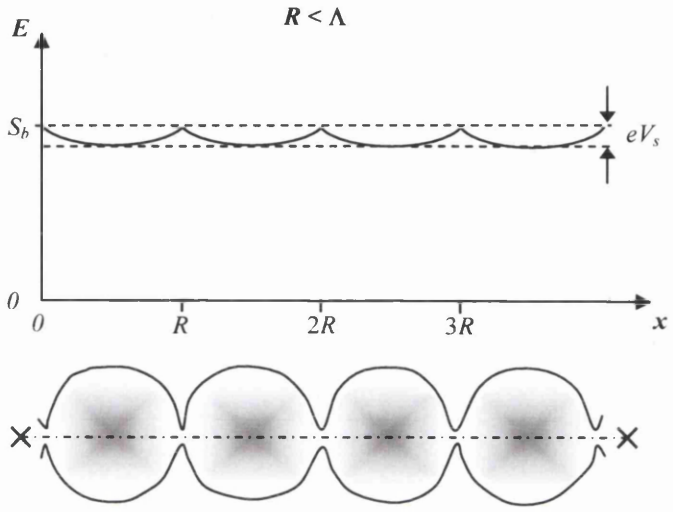


Figure 2-10 Energy diagram representation of a chain of $R<\Lambda$ sintered grains.

This implies, from Eq. (2.43), that the conductance of the film should increase, the barrier that must be 'overcome' for conduction to take place now lowered. Defining Λ_D as the depletion width of the smallest grain to be considered to possess both a depletion width and a 'bulk' region (see Section 5.6 for more detailed discussion about depletion width behaviour and the variation of Λ with R), then the conductance of grains $R < \Lambda_D$ should be markedly better than those of $R > \Lambda_D$. Provided, of course, that this ease of conduction is not completely countered by the diminution in the number of available charge carriers, grains $R < \Lambda_D$ not possessing an undepleted region of semiconductor.

Band bending implies surface states within the confines of this model, and the alteration in band bending behaviour will have a corresponding manifestation in the surface state density. In order to combine an investigation of this and to simultaneously provide a validation of the methods of Section 2.4 (which is, after all, the main purpose of this whole Chapter), this surface state behaviour will be compared directly with work published in the literature, namely Malagú *et al.*

[15]. To do this a few extra conventions and approximations must be adopted to align the model developed here with that of the stated reference. The energetic zero of the system is now set at the bottom of the conduction band of a large, $R \gg \Lambda_D$, grain. The position of the Fermi level relative to this minimum is such that the conductive electron gas can be considered non-degenerate and so the approximation Eq. (1.41) can be applied to reduce the evaluation of the Fermi-Dirac integral to the simple calculation of an exponential:

$$n = N_c e^{(E_f - v_p(r))/k_B T} \tag{2.44}$$

The chosen work of literature uses the more conventional Poisson potential φ i.e. $\varphi(r) = -v_p(r)/e$ and works with N_t, the surface acceptor density, instead of the occupied surface state density N_s, its negative.

Finally, it is assumed that the donor states are completely ionised and have a density N_d, which is equal to $N_d = N_c e^{E_f/k_B T}$. Unfortunately, for Malagú *et al.*'s $N_d = 5\times10^{24}$ m^{-3} (see also footnote [ii] of Section 6.1), this implies that the Fermi level lies only 0.0516 eV (<1 $k_B T$) below the bulk conduction band bottom, implying that the system is not non-degenerate, and as such, that the use of Eq. (2.44) is inappropriate and could lie in error in excess of 5% [35]. Inverting the Fermi-Dirac integral directly using Nilsson's equation [57], accurate to within 0.5%, supports this assessment, placing E_f at -0.0434 eV. Nonetheless, to test the accuracy of the methods of Section 2.3, the non-degenerate approximation is still adopted.

At an operating temperature of 673K the value of potential at the grain boundary is set at $V = -0.68$V, and along with the donor density, is considered constant for all radii. In the new format Eq.'s (2.8) to (2.10) become

$$\frac{d^2\varphi}{dr^2} + \frac{2}{r}\frac{d\varphi}{dr} = -\frac{e}{\varepsilon_0\varepsilon_r}\left(N_d - N_d e^{e\varphi/k_B T}\right) \tag{2.45}$$

$$\varphi(R) = V \tag{2.46}$$

$$-\left.\frac{d\varphi}{dr}\right|_{r=0} = 0 \tag{2.47}$$

$$N_t = -\frac{\varepsilon_r\varepsilon_0}{e}\left.\frac{d\varphi}{dr}\right|_{r=R} \tag{2.48}$$

Solving these equations using the techniques outlined in Section 2.3, then for the radii R = 100nm, 30nm, 10m, and 5nm, the band bending of Ref. [15] is recreated, as seen in Figure 2-11.

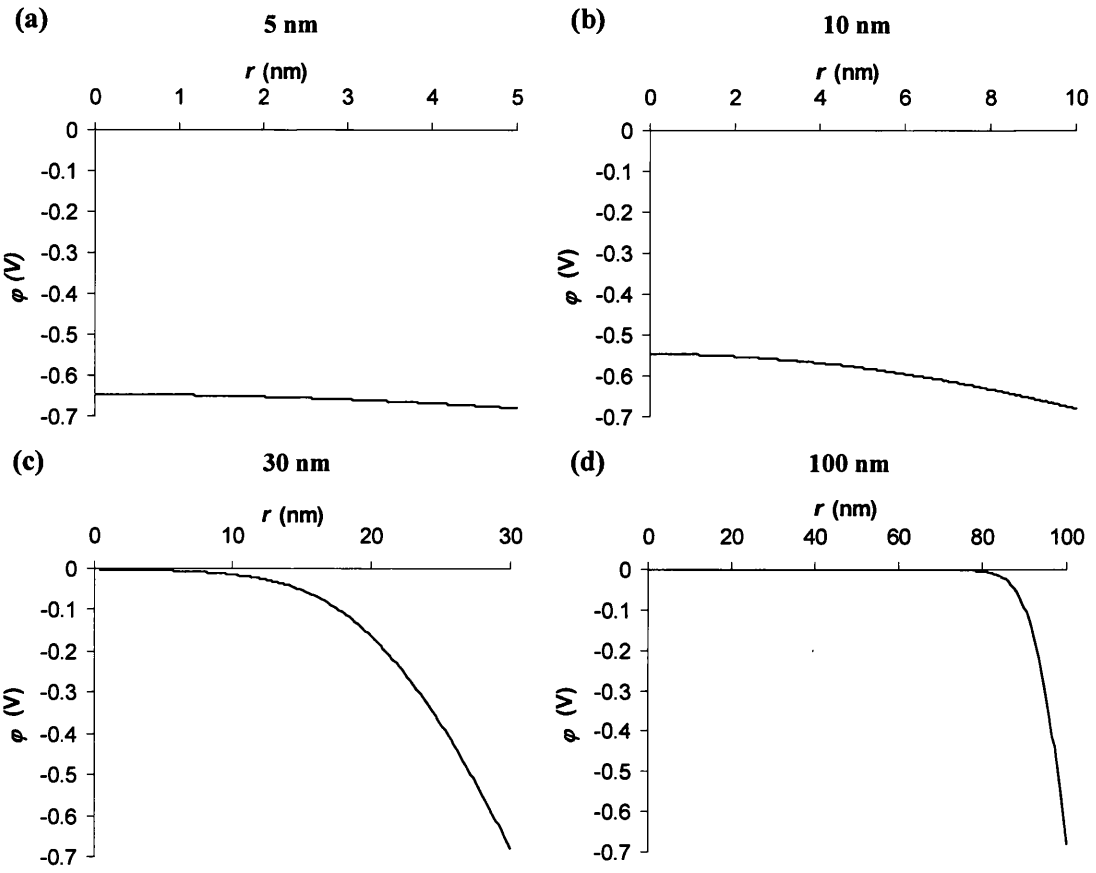


Figure 2-11 Plots of the potential φ against the radial coordinate r for (a) R = 5nm (b) R = 10nm (c) R = 30nm (d) R = 100nm SnO_2 grains at V_b = -0.68 V, N_d = $5\times10^{24}m^{-3}$, 673 K.

The $\varphi(R)$ - $\varphi(0)$ difference of the 10 and 5nm grains identical to the two decimal places of Malagú *et al*'s -0.13 V and -0.03 V respectively, and Λ_D lies at the reported ~22nm. The surface acceptor density trend in the range 0.1nm to 100nm is plotted in Figure 2-12, and appears identical to Ref. [15]'s Figure 6.

It can therefore be assumed that the methods of Section 2.3 are perfectly adequate to solve the non-linear Poisson equation for these kinds of dimensions and conditions. The use of the non-degenerate approximation to the electron density does not appear to have an undue effect, the discrepancy between the usual Fermi-Dirac integral version of Eq. (2.44) most apparent at larger radii i.e. when $R > \Lambda_D$ and the potential reaches its 'bulk' position (and therefore lying closest to the Fermi level). By R = 100nm the Fermi-Dirac integral trend lies 3.5 % above the non-degenerate approximation.

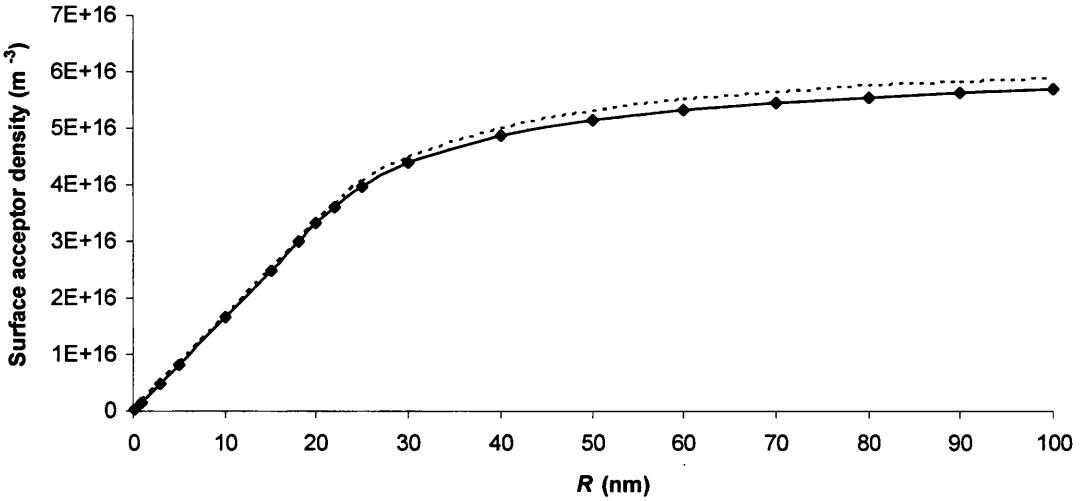


Figure 2-12 Plot of surface acceptor density against SnO_2 grain radius for the Malagú *et al.* [15] system: V_b = -0.68 V, $N_d = 5\times10^{24}m^{-3}$, 673 K. The dotted line indicates the replacement of the non-degenerate approximation of the electron density with the Fermi-Dirac integral based density. By R = 100nm this lies 3.5 % above the non-degenerate approximation

What effect does the behaviour of Figure 2-12 have on the sensing abilities of gas sensors?

Whilst N_s ($-N_t$) is not equal to N_{ss}, intrinsic states aside and assuming that S_b - ϕ_0 is constant with respect to grain radius, then a decrease in N_s would certainly be reflected in N_{ss}. As $R \rightarrow 0$ so does N_s, implying that $\gamma_{gs} \rightarrow 1$, and therefore $\partial S_b / \partial \Phi_g \rightarrow 1$. This means that changes in the work function of the target gas (i.e. different gasses) will cause equivalent changes in the surface barrier height S_b. This will follow through to larger variations in conductance via Eq. (2.43), implying that the sensitivity of the sensor (R_{air}/ R_{gas}) will be high.

As $R \rightarrow \infty$ and the grains have the properties of the bulk semiconductor, then N_s is large and $\gamma_{gs} \rightarrow$ 0, implying that $\partial S_b / \partial \Phi_g \rightarrow 0$ and changes in the gas work function will be poorly reflected in corresponding changes in S_b. As such, the conductivity between samples in air and samples in the test gas will be similar and the sensitivity of the sensor low in comparison with the small R regime.

Of course, based on this, when the grains are exposed to air (forming surface states) then the surface state densities will be lower the smaller the grain, and so, importantly, with fixed Φ_g and

fixed χ_s[ii], the change in N_s with grain size implies a change in S_b given Eq. (2.1), now recast as

$$S_b = \Phi_g - \chi_s - e^2 \frac{N_s \delta_{it}}{\varepsilon_{it}} \tag{2.49}$$

Over Figure 2-12 then, with a change in N_t of ~ 5.7×10^{16} m^{-2} and a δ_{it} = 0.5 to 2nm (from the metal-semiconductor interface of Tung [39]), then between R = 0 and 110nm, a ΔV of 0.046 to 0.183V would be expected from Eq. (2.49) i.e. between 6.7% and 26.9% of V's (S_b for the more usual terminology of this work) given value. A change in V (S_b) would generally be expected to affect grains more when $R > \Lambda$ (and perhaps influence the value of Λ_D itself) than when $R < \Lambda$[iii]. If V increases in a region where it does significantly affect N_t, then N_t would also increase, ergo, it would be expected that the trend of Figure 2-12 below 22nm would not change, or rise only slightly, while that above would become noticeably steeper.

If Fermi level unpinning were in evidence, then it would be seen as a diminution of S_b from large to small grains for constant Φ_g. Indeed, Maffeis *et al.* [31] does observe such a decrease: 1.4(±0.1)eV at R = 15nm and 1.3(±0.05)eV at R = 4nm, although with the experimental limits of error, this is not conclusive evidence. Based on Eq. (2.49), δ_{it} = 0.5 to 2nm [39] and the work of Chapter 6 in assessing N_s, then ΔS_b between the two grains would be expected to lie between 0.006eV and 0.024eV.

Of course, Eq. (2.49) is not itself exact, as the quantity of charge $(S_b - \phi_0)N_{ss}$ excluded the intrinsic space charge, and the total voltage drop caused by N_s, although a representation of the total

---

[ii] The determination of the electron affinity of Tin dioxide is not completely straightforward. Values of SnO_2's work function range from 4.3eV [58] to 4.7±0.2eV [59]. It is a quantity very dependent on the preparation of the sample. It can be inferred from the work of Kulger *et al.* [60] on Indium-SnO_2 (ITO) that there will be a similar dependence of SnO_2 on the cleaning method of the sample: ITO cleaned with Ne^+ spluttering has Φ = 4.0eV while cleaning with H_2O_2 gives Φ = 4.8eV. Batzill *et al.* [61] suggest a value of Φ_{SnO_2} of 4.55eV for a sample cleaned with Ar^+ and annealed at 900K, while Shen *et al.* [62] concludes that the annealing process over the temperature range 300K to 1100K introduces a work function shift in Tin dioxide of between approximately +0.2eV and -0.7eV (using error bar extremes).

For this work, an estimation of the electron affinity is calculated from Eq. (2.49). Using the Schottky barrier heights as measured by Maffeis *et al.* [31] and an approximated value of the work function of a ~0.9nm spherical Tungsten tip (Φ_W = 4.53eV for a Tungsten sphere of radius 10nm [63]), a value for the electron affinity is calculated at each iteration of the P-KS-CB method (see Chapter 6) from Eq. (2.49). Generally, χ_{SnO_2} can be taken to be ~3.2eV.

[iii] For larger grains where $R > \Lambda$, a change in $V(S_b)$ will have a profound affect on the gradient of the potential at R, while for grains where $R < \Lambda$, the significant 'damping' of the exponential term in Eq. (2.45) over the whole of the potential will reduce the impact of any changes in $V(S_b)$.

surface state density (inclusive of the influence of the intrinsic space charge), may not accurately allow for all the effects of the natural surface states.

Regardless of S_b's movement, any decrease of N_s will be seen as an increase in sensor sensitivity, and when the grains become completely depleted there would be expected to be considerable sensitivity enhancement. However, while the sensitivity of the sensing film will indeed increase with the decrease of grain size, if $\Lambda \approx 22$nm for $V = -0.68$ and $N_d = 5\times10^{24}$ m^{-3} is representative for gas sensing films as a whole, then full depletion ($R < 22$nm) would not seem to explain the notable increase in sensitivity below a radius of 10nm (the 20nm diameter of [41,42]), nor the results of Ref.'s [45,46,47].

Notwithstanding the discussion of the complexities of conductance in a gas sensing films in Section 2.1, a simple qualitative evaluation of the sensitivity of the Malagú system can be offered. Define the sensitivity S of a sensing film as the ratio of its resistance in air and its resistance after exposure to the target gas i.e.

$$S = \frac{R_{air}}{R_{gas}} \tag{2.50}$$

or in terms of conductance

$$S = \frac{G_{gas}}{G_{air}} \tag{2.51}$$

Assuming narrow sinter necks such that the dominant conduction mechanism is over the surface barriers, then removing the electron density from G_o of Eq. (2.43) the conductance can be expressed as

$$G = G_0' n_{film} e^{-eV_s/k_BT} \tag{2.52}$$

where n_{film} is the electron density of the gas sensing film. This quantity will depend on the density of the grains in the film and the amount of inter-granular space. While this of course could be estimated if the packing configuration was known, since the film structure can be taken as constant across gas exposure, then assuming uniform grain size across a film, the relative electron densities can be expressed as the ratio of the relevant grain electron densities. Therefore, Eq. (2.51) can be written

$$S = \frac{n_{gas} e^{eV_s^{air}/k_BT}}{n_{air} e^{eV_s^{gas}/k_BT}} \tag{2.53}$$

The action of the reducing target gas is to return electrons to the conduction band. Ergo, whilst this is clearly dependent on the type of gas, its concentration in the atmosphere, length of exposure, speed of reaction, permeability of the film to the gas and so on, the maximum possible sensitivity, independent of the exact reducing gas, is give by a total return of all the charge in the surface states to the conduction band. This implies the disappearance of V_s^{gas}, and as the donor density has been assumed to be constant across all sizes of grain and all donors fully ionised, then

$$S_{\max} = \frac{N_d e^{eV_s^{air}/k_B T}}{n_{air}} \tag{2.54}$$

n_{air} and V_s^{air} are simply the electron density and depth of well of the Malagú system. Figure 2-13 plots this maximum sensitivity against grain radius, clearly implying that although flattening of the conduction band does improve the sensitivity, it does so for $R < 20$nm, and contrary to a particular increase below a radius of 10nm, the sensitivity actually levels off. Naturally, the actual sensitivity would be vastly lower than this and gas dependent, but the general trend should still be relevant, neglecting that is any dependency of permeability, porosity and the like, in the film on grain radius.

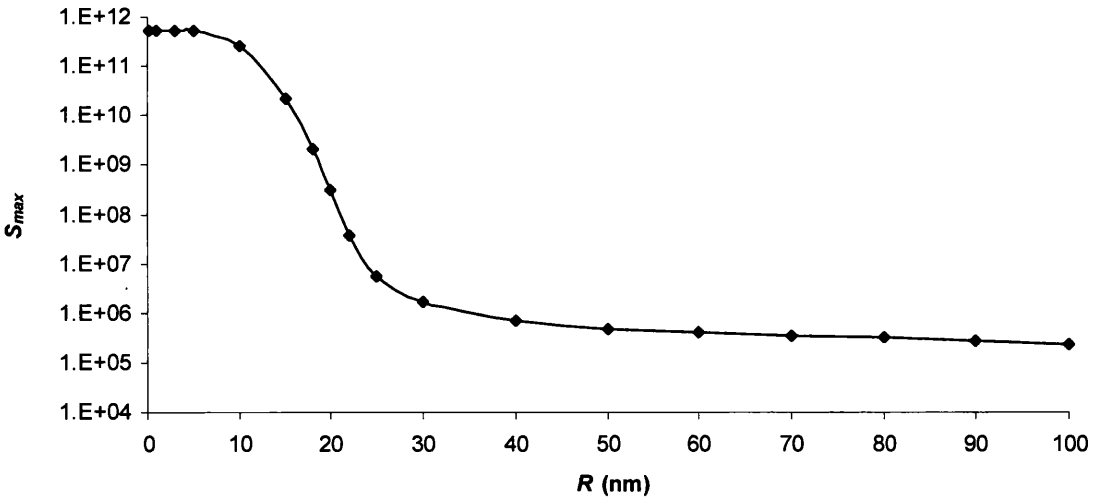


Figure 2-13 Plot of the theoretical maximum sensitivity, S_{max}, against SnO_2 grain radius for the Malagú *et al* system.

While not the main focus of this work, this sensitivity issue will be briefly revisited in Chapter 6 and an alternative model presented rectifying this discrepancy between theory and experimental observation.

Returning to Maffeis *et al.* [31], in addition to the surface barrier heights, the authors also measured the ratio of surface state densities, and place that of the R =15nm nanocrystals 5.90 times greater than that of the R = 4nm nanocrystals. Applying the complete charge density method to their system, then the ratio predicted is 3.75[iv] indicating that some problem exists with this theoretical model.

A possible cause of this discrepancy is the neglect of quantum effects. As was seen in Section 1.4, spatial confinement causes the formation of discrete energy levels, and so a small enough grain is, in effect, a quantum dot. As a result, its density of states will mirror the delta functions of Eq. (1.58) and Figure 1-20, and not the smooth continuous profile of Figure 1-17 leading to the bulk expression for the electron density, Eq. (1.34), used here. The effect of the discrete energy levels, and the point at which this quantised behaviour diverges from its bulk counterpart, is the subject of the following Chapters, the charge written grains of Figure 1.1 having a radius of 4nm, and so likely to lie in the realm of the quantum dot rather that in the bulk, if the work of this section can be taken to provide even a slight guide.

In summary then, whilst the numerical techniques of Section 2.3 applied to the non-linear Poisson equation of the CCDM appear to exactly recreate the results published in literature, accurately reproducing the theorised band bending flatting of $R < \Lambda_D$ grains, it appears that neither this decrease in well depth nor the proposed un-pinning of the Fermi level through the diminution of the surface state density quite explains all the characteristics of the experimentally observed sensitivity increase of gas sensing films as the radius of the grains diminishes, particularly below a radius of 10nm. Neither does the complete charge density model completely account for the measured surface state density ratio between 4nm and 15nm radius grains, predicting a ratio of 3.75 compared to the observed 5.90. Ergo, it would appear that the CCDM as it stands is not sufficient to simulate the electronic characteristics of nanometric SnO_2 grains, and it is proposed that the effects of quantisation must be incorporated into the model.

The pinning or unpinning of the Fermi level and its effects on S_b are not precisely treated in this

[iv] The quoted ratio of 3 for the theoretical value follows from the ratio between R = 5nm and R = 15nm grains, or, unfortunately as it appears from the following paper, Malagú *et al.* [17], between R = 10nm and R = 30nm grains, some confusion having arisen between the experimentalists' diameters and the theorist's radii. This radius/diameter error also appears to be evident in Malagú *et al.* [15] and after discussion with the experimental contributors, Malagú *et al.* [16].

work. For the main results, see Chapter 6, the error bars reflect the known uncertainty in the Schottky-like barrier heights from Maffeis *et al.* [31]'s work. For radii in excess of R =15nm this is likely to be sufficient – for a bulk grain, based on Eq. (2.49) there is likely to be a maximum shift in S_b of ~0.13eV from that of the smallest sample, only 0.03eV greater than that allowed for.

2.5 Conclusions

This Chapter saw the first steps along the path to developing the models necessary to simulate the electronic structure of a nanometric grain during charge injection.

The surface states of semiconductor were treated. These states form from the free 'dangling' bonds of the semiconductor lattice ions at the edge of the lattice, these ions no longer surrounded on all sides like their fellows in the bulk of the semiconductor. The drain of electrons from the conduction band into these states causes the conduction and valence bands to bend and results in the formation of a space charge, or depletion, region and a surface barrier. A qualitative model of this is developed using a metal-semiconductor interface and the relationship between the surface barrier and the surface states is investigated. This includes mention of 'Fermi level pinning' where a large quantity of charge trapped in the surface states renders the interface independent of the type of metal (or gas), the surface conduction band fixed relative to the Fermi level by the surface charge. The gas-semiconductor interface is then discussed, along with the formation of surface states by the chemisorption (strong chemical bonding) of oxygen species on the SnO_2 surface. Once formed, unless the oxygen species are removed by a reducing gas and the electrons returned to the conduction band, these surface states and surface state induced band bending will remain, even if the grains are placed in a vacuum (as in the charge injection process for example), and so their detailed treatment is essential for the objectives of this work.

To model the bending of the conduction band due to these surface states, the non-linear Poisson equation complete charge density model of the literature is introduced, and a suite of numerical methods created to solve it. The ODE is discretised over a discrete spatial mesh and a Taylor series expansion method is used, within a shooting method framework employing a bisection methodology, to solve for the conduction band bottom satisfying the relevant boundary conditions of a spherical nanoparticle. In this way the conduction band bottom is determined which satisfies the known value of the surface barrier height to 5×10^{-6} % within 20 to 25 iterations, the global error of the solution of fourth order in the inter mesh point spacing. The composition of the Taylor series formulas has been carefully orchestrated to maximise stability and speed, and the chosen methods compare favourably with the alternative numerical techniques for these criteria.

To evaluate the performance of the numerical suite as a whole, these techniques are compared favourably to work from the literature, where the surface state and band bending properties of spherical SnO_2 nanocrystals have been investigated in a gas-sensing context. It is found that when the grains become too small to possess a distinct region un-depleted of charge carriers, then the curvature of the conduction band flattens and the depth of the potential well, formed between the conduction band at the surface and at the centre of the grain, diminishes. For the system described by Malagú *et al* [15], the depths of R = 10nm and R = 5nm grains are correctly calculated to be 0.13V and 0.03V respectively, and the behaviour of the surface state acceptor density with grain radius appears to be identical to that published.

Remaining within the gas-sensing milieu, conduction between the grains of a gas-sensing film is discussed, and is considered to be dominated by the formation of inter-granular connections, or sinter necks, during annealing and operates through three distinct processes, depending on both the size of, and the concentration of, charge carriers in the sinter neck. The publications of literature considered within this project mainly treat the transport mechanism between grains as solely that of conduction over a Schottky-like barrier, formed at the sinter neck. As such, two alternative reasons to explain the experimentally observed increase in sensitivity of the sensing films on the diminution of particle size, particularly beneath R = 10nm, are considered with simple models. The first proposed explanation attributes these effects to the unpinning of the Fermi level, and consequent movement of the barrier height. Based on the data from the complete charge density model this is thought in general to be insufficient to the task. The alternative hypothesis, the decrease in well depth as a consequence of band bending flattening, is again evaluated from the complete charge density model and is also considered to be wanting, appearing to place the marked upturn in sensitivity at R = 20nm instead of the observed R = 10 nm.

Furthermore, in Maffeis *et al.* [31], STS measurements place the ratio of surface state densities between 4nm and 15nm radii grain to be 5.90 while the complete charge density model calculates the ratio to lie at 3.75.

These discrepancies would seem to indicate that some problem exists with the current model of the surface states and conduction band. It is proposed that at small dimensions, the neglect of the quantisation of the energy spectrum in the CCDM could contribute to these incongruities.

The next Chapter extends the complete charge density model and begins to address the issue of the electron population of the grain being confined to the discrete energy levels.

Chapter 3 Solving the Schrödinger Equation

Having generated a conduction band profile from the solution of Poisson's equation, the next stage in improving the sophistication, and hopefully the accuracy, of the computational model of the nanocrystals is to ascertain their eigenstates; their permissible energy levels and corresponding wave functions.

Section 3.1 deals with separating the electron wave function into two parts – the radial and the angular, and solving the angular equation. The angular eigenstates can be determined exactly, with the eigenvalues forming an implicit part of the radial equation. The quantum mechanics necessary to interpret the angular component is developed along a path broadly similar to Bohm [64].

Section 3.2 develops a method of approximating the eigenstates of the radial equation for an arbitrary radially dependant potential: the ODE is approximated on a discrete grid of spatial points, and a tri-diagonal matrix formed modelling the space. From this matrix, with some manipulation, an orthonormal set of eigenvectors can be formed, which with their corresponding eigenvalues, constitute the eigenstates of the radial Schrödinger equation.

In Section 3.3, the analytical solutions of the radial equation for a square potential well are derived and compared, favourably, with the approximate solutions from the finite difference method developed in the preceding section. Section 3.4 compares the finite difference results for this nearly free electron model with those from an alternative, tight-binding approach, again favourably.

Section 3.5 discusses the issue of self-consistency, and its importance especially when modelling higher electron densities, and proposes a straightforward and readily workable iterative scheme to address the many problems in achieving this.

In the conclusion to this chapter, Section 3.6, the procedures and methods that are herein developed are extended to allow for a variable spatial mesh, allowing larger systems to be considered whilst still retaining the necessary resolution through a small mesh spacing in the regions of particular importance. In addition, the prospect of non-homogeneous media is discussed and the required modifications made to ensure the accuracy of the eigenstates in this more complex situation.

3.1 Separation of Variables and Spherical Harmonics

In a three-dimensional space, the time independent wave equation assumes the form

$$\nabla^2\psi_n + \frac{2m}{\hbar^2}(E_n - V)\psi_n = 0 \tag{3.1}$$

Should the symmetry of the space be such that spherical polar coordinates make a convenient coordinate system, see Figure 3-1,

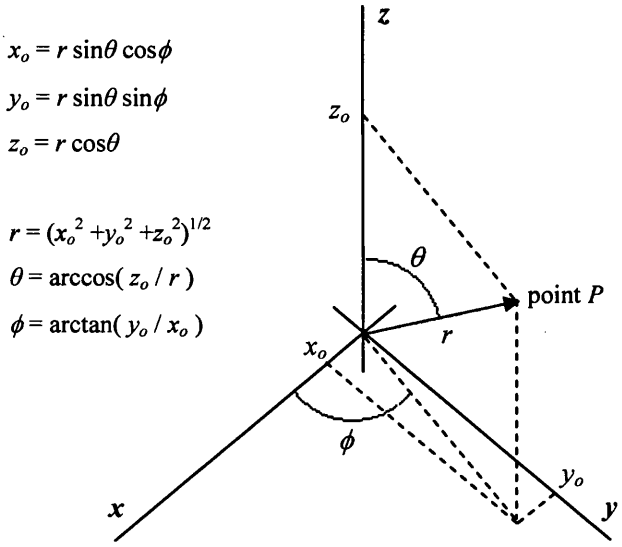


Figure 3-1 Three dimensional coordinate systems: the relationship between Cartesian (x, y, z) and spherical polar (r, θ, ϕ)

then the Laplacian acting on ψ becomes

$$\begin{aligned}\nabla^2\psi_n &= \frac{1}{r}\frac{\partial^2}{\partial r^2}(r\psi_n) + \frac{1}{r^2}\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta} + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi}\right]\psi_n \\ &= \frac{1}{r}\frac{\partial^2}{\partial r^2}(r\psi_n) + \frac{1}{r^2}\Omega\psi_n\end{aligned} \tag{3.2}$$

If the potential V is dependant on only one variable r, then the solution of Eq. (3.1) can be greatly simplified as its eigenvectors can be separated into two functions, one solely dependant on r and the other a function of θ and ϕ. Letting

$$\psi_n = f_n(r)Y(\theta,\phi) \tag{3.3}$$

then Eq. (3.1) becomes

$$\left[\frac{1}{r}\frac{\partial^2}{\partial r^2}(rf_n(r)) + \frac{2m}{\hbar^2}(E_n - V(r))f_n(r)\right]Y(\theta,\phi) + \frac{f_n(r)}{r^2}\Omega Y(\theta,\phi) = 0$$

i.e.

$$r^2\left[\frac{1}{f_n(r)r}\frac{\partial^2}{\partial r^2}(rf_n(r))+\frac{2m}{\hbar^2}\left(E_n-V(r)\right)\right]=-\frac{\Omega Y(\theta,\phi)}{Y(\theta,\phi)} \tag{3.4}$$

and therefore, a function in r must equal, for all values of r, θ and ϕ, a function in θ and ϕ. Ergo, each function must be constant:

$$\Omega Y(\theta,\phi) = cY(\theta,\phi) \tag{3.5}$$

$$\frac{1}{r}\frac{d^2}{dr^2}(rf_n)+\frac{2m}{\hbar^2}\left(E_n-V(r)\right)=-\frac{cf_n}{r^2} \tag{3.6}$$

The task of ascertaining the eigenstates of Eq. (3.1) has now been reduced to solving Eq. (3.5) and then Eq. (3.6). Eq. (3.6) is dependant on the form of V and as such cannot always be simply solved analytically and its treatment is discussed further in Section 3.2. The remainder of this section is dedicated to determining the physically admissible values of c and the corresponding eigenfunctions $Y(\theta,\phi)$, generally referred to as spherical harmonics. Intrinsic to this is the quantum mechanical formulation of angular momentum, the operator Ω closely related to the total angular momentum operator L^2. Beginning then with the concept of angular momentum, define the angular momentum operator such that

$$\boldsymbol{L}=\boldsymbol{r}\times\boldsymbol{p} \tag{3.7}$$

then its components are

$$\begin{aligned} L_x &= yp_z - zp_y \\ L_y &= zp_x - xp_z \\ L_z &= xp_y - yp_x \end{aligned} \tag{3.8}$$

which in QM terms can be written

$$\begin{aligned} L_x &= \frac{\hbar}{i}\left[y\frac{\partial}{\partial z}-z\frac{\partial}{\partial y}\right] \\ L_y &= \frac{\hbar}{i}\left[z\frac{\partial}{\partial x}-x\frac{\partial}{\partial z}\right] \\ L_z &= \frac{\hbar}{i}\left[x\frac{\partial}{\partial y}-y\frac{\partial}{\partial x}\right] \end{aligned} \tag{3.9}$$

on defining the momentum components $p_x=\frac{\hbar}{i}\frac{\partial}{\partial x}$ etc.

In the spherically symmetric space of Eq. (3.5)

$$\begin{aligned} x &= r\sin\theta\cos\phi \\ y &= r\sin\theta\sin\phi \\ z &= r\cos\theta \end{aligned}$$

recalling Figure 3-1, and therefore

$$dx = \sin\theta\cos\phi\, dr + r\cos\theta\cos\phi\, d\theta - r\sin\theta\sin\phi\, d\phi$$
$$dy = \sin\theta\sin\phi\, dr + r\cos\theta\sin\phi\, d\theta + r\sin\theta\cos\phi\, d\phi$$
$$dz = \cos\theta\, dr + r\sin\theta\, d\theta$$

then on separating the Cartesian derivatives into their polar components

$$\frac{\partial}{\partial x} = \frac{\partial r}{\partial x}\frac{\partial}{\partial r} + \frac{\partial\theta}{\partial y}\frac{\partial}{\partial\theta} + \frac{\partial\phi}{\partial z}\frac{\partial}{\partial\phi}$$
$$= \sin\theta\cos\phi\frac{\partial}{\partial r} + \frac{\cos\theta\cos\phi}{r}\frac{\partial}{\partial\theta} - \frac{\sin\phi}{r\sin\theta}\frac{\partial}{\partial\phi}$$
$$\frac{\partial}{\partial y} = \sin\theta\sin\phi\frac{\partial}{\partial r} + \frac{\cos\theta\sin\phi}{r}\frac{\partial}{\partial\theta} + \frac{\cos\phi}{r\sin\theta}\frac{\partial}{\partial\phi}$$
$$\frac{\partial}{\partial z} = \cos\theta\frac{\partial}{\partial r} - \frac{\sin\theta}{r}\frac{\partial}{\partial\theta}$$

and so finally

$$L_x = i\hbar\left[\sin\phi\frac{\partial}{\partial\theta} + \cot\theta\cos\phi\frac{\partial}{\partial\phi}\right]$$
$$L_y = \frac{\hbar}{i}\left[\cos\phi\frac{\partial}{\partial\theta} - \cot\theta\sin\phi\frac{\partial}{\partial\phi}\right] \qquad (3.10)$$
$$L_z = \frac{\hbar}{i}\frac{\partial}{\partial\phi}$$

The absolute magnitude of the angular momentum, $|L|$, follows from

$$L^2 = L_x^2 + L_y^2 + L_z^2 \qquad (3.11)$$

so

$$L^2 = -\hbar^2\left[x^2\left(\frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) + y^2\left(\frac{\partial^2}{\partial z^2} + \frac{\partial^2}{\partial x^2}\right) + z^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) - 2xy\frac{\partial^2}{\partial x\partial y}\right.$$
$$\left. - 2yz\frac{\partial^2}{\partial y\partial z} - 2zx\frac{\partial^2}{\partial z\partial x} - 2x\frac{\partial}{\partial x} - 2y\frac{\partial}{\partial y} - 2z\frac{\partial}{\partial z}\right]$$

Adding this to the square of the scalar product of $\boldsymbol{r}$ and $\boldsymbol{p}$

$$(\boldsymbol{r}\cdot\boldsymbol{p})^2 = -\hbar^2\left[x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y} + z\frac{\partial}{\partial z}\right]^2$$
$$= -\hbar^2\left[x^2\frac{\partial^2}{\partial x^2} + y^2\frac{\partial^2}{\partial y^2} + z^2\frac{\partial^2}{\partial z^2} + 2xy\frac{\partial^2}{\partial x\partial y} + 2yz\frac{\partial^2}{\partial y\partial z} + 2zx\frac{\partial^2}{\partial z\partial x} + x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y} + z\frac{\partial}{\partial z}\right]$$

then

$$L^2 + (\boldsymbol{r}\cdot\boldsymbol{p})^2 = -\hbar^2(x^2 + y^2 + z^2)\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) + \hbar^2\left(x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y} + z\frac{\partial}{\partial z}\right) \qquad (3.12)$$
$$= r^2p^2 + i\hbar^2\boldsymbol{r}\cdot\boldsymbol{p}$$

which can be expressed as

$$p^2 = \frac{L^2}{r^2} - \frac{\hbar^2}{r^2}\left(r\frac{\partial}{\partial r}\right)^2 - \frac{\hbar^2}{r}\frac{\partial}{\partial r} \tag{3.13}$$

Given that the Hamiltonian, H, of $H\psi_n = E_n\psi_n$ can be written

$$H = \frac{p^2}{2m} + V \tag{3.14}$$

then on comparing Eq. (3.4) with Eq.'s (3.13) and (3.14) it is clear that

$$L^2 = -\hbar^2\Omega \tag{3.15}$$

implying that the problem of determining the physically permissible values of c is equivalent to ascertaining the eigenvalues of the total angular momentum operator L^2.

Interestingly, from the commutation relations of the angular momentum operators

$$\begin{aligned}
\left[L_x, L_y\right] &= L_xL_y - L_yL_x \\
&= -\hbar^2\left[\left(y\frac{\partial}{\partial z} - z\frac{\partial}{\partial y}\right)\left(z\frac{\partial}{\partial x} - x\frac{\partial}{\partial z}\right) - \left(z\frac{\partial}{\partial x} - x\frac{\partial}{\partial z}\right)\left(y\frac{\partial}{\partial z} - z\frac{\partial}{\partial y}\right)\right] \\
&= -\hbar^2\left(y\frac{\partial}{\partial x} - x\frac{\partial}{\partial y}\right) \\
&= i\hbar L_z \\
\left[L_y, L_z\right] &= i\hbar L_x \\
\left[L_z, L_x\right] &= i\hbar L_y
\end{aligned} \tag{3.16}$$

and

$$\begin{aligned}
\left[L^2, L_z\right] &= L^2L_z - L_zL^2 \\
&= \left(L_x^2 + L_y^2\right)L_z - L_z\left(L_x^2 + L_y^2\right) \\
&= L_x\left(L_xL_z - L_zL_x\right) + \left(L_xL_z - L_zL_x\right)L_x + L_y\left(L_yL_z - L_zL_y\right) + \left(L_yL_z - L_zL_y\right)L_y \\
&= -i\hbar\left(L_xL_y + L_yL_x - L_yL_x - L_xL_y\right) \\
&= 0
\end{aligned} \tag{3.17}$$

a number of conclusions can be drawn. From Eq. (3.17), since L^2 commutes with L_z, symmetry implies that L^2 will also commute with L_x and L_y. Then it is possible to measure simultaneously the eigenstates of L^2 and any single component of L. However, as the components L_x, L_y, and L_z do not commute with each other, only one of these can be specified at a time.

The eigenvalues of L_z must satisfy

$$L_zY(\theta,\phi) = \frac{\hbar}{i}\frac{\partial}{\partial\phi}Y(\theta,\phi) = k_1Y(\theta,\phi) \tag{3.18}$$

On splitting $Y(\theta,\phi)$ into two functions, $A(\theta)B(\phi)$, then

$$\frac{\hbar}{i}\frac{\partial}{\partial\phi}B(\phi)=k_1B(\phi)$$

which has the solution $B(\phi)=e^{ik_1\phi/\hbar}$. $B(\phi)$ must be a single valued function of x, y and z, and therefore it must be periodic in ϕ with period 2π, implying that $k_1/\hbar=m$ where m is an integer. Ergo,

$$L_z=m\hbar \tag{3.19}$$

How are L_z's eigenfunctions related to the eigenfunctions of L^2? Substituting $A(\theta)e^{im\phi}$ into Eq. (3.15) then

$$\begin{aligned}L^2A(\theta)e^{im\phi}&=-\hbar^2\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta}+\frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi}\right]A(\theta)e^{im\phi}\\&=-\hbar^2\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right)-\frac{m^2}{\sin^2\theta}\right]A(\theta)e^{im\phi}\\&=cA(\theta)e^{im\phi}\end{aligned}$$

thus

$$-\hbar^2\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right)-\frac{m^2}{\sin^2\theta}\right]A(\theta)=cA(\theta) \tag{3.20}$$

So $A(\theta)$ can be taken as dependant on both m and its own eigenvalue c, relabelled k_2 i.e. $A^m_{k_2}(\theta)$, and evidentially $e^{im\phi}A^m_{k_2}(\theta)$ serves as an eigenfunction for both L_z and L^2.

Determining $e^{im\phi}A^m_{k_2}(\theta)$'s corresponding eigenvalue for L^2, c, is a little more involved.

If

$$L^2Y(\theta,\phi)=cY(\theta,\phi) \tag{3.21}$$

then

$$L^2\left(L_zY\right)=L_zL^2Y=L_zcY=c\left(L_zY\right) \tag{3.22}$$

therefore if Y is an eigenfunction of L^2, then (L_zY) is also an eigenfunction belonging to the same eigenvalue c.

Now, for

$$L_z Y = m\hbar Y$$

multiplication by the operator $L_+ = L_x + iL_y$ results in

$$L_+ L_z Y = m\hbar L_+ Y \tag{3.23}$$

but from the previous commutation relations, Eq. (3.16)

$$(L_x + iL_y)L_z - L_z(L_x + iL_y) = -\hbar(L_x + iL_y) \tag{3.24}$$

implying that

$$L_z L_+ Y = (m+1)\hbar L_+ Y \tag{3.25}$$

and similarly, from the operator $L_- = L_x - iL_y$

$$L_z L_- Y = (m-1)\hbar L_- Y \tag{3.26}$$

Therefore, if Y is an eigenfunction of L_z where $L_z = m\hbar$ then $L_\pm Y$ is also an eigenfunction of L_z but belonging to $L_z = (m \pm 1)\hbar$. Significantly, both Y and $L_\pm Y$ are eigenfunctions of L^2 corresponding to the same eigenvalue. This implies that starting with a given eigenfunction of L_z, then it is possible to generate all the eigenfunctions of L_z belonging to the same eigenvalue of L^2.

$L_\pm$ acts to raise/lower m by unity leaving c unchanged, however as

$$\begin{aligned} L^2 &= L_x^2 + L_y^2 + L_z^2 \\ &= L_x^2 + L_y^2 + m^2\hbar^2 \end{aligned} \tag{3.27}$$

and the mean values of L_x^2 and L_y^2 must always be positive[i], then $L^2 \geq \hbar^2 m^2$ and thus there must be an upper limit on $|\hbar m|$ of $\sqrt{L^2}$. Therefore, there must exist some value of m for which $L_\pm Y_{k_2}^m$ vanishes, else repeat application of the $L_\pm$ operator will lead to an infinitely large set of eigenvalues of L_z for given L^2.

In the state where $|m|$ is at its maximum for given L^2, then

$$L_+ Y^{m_1} = 0 \quad \text{or} \quad L_- Y^{m_2} = 0 \tag{3.28}$$

where m_1 is the maximum positive value of $L_z/\hbar$ and m_2 is the maximum negative value (for given L^2), then

---

[i] Consider a rotation of the axes so that the new z axis is parallel to the old x axis, then $L_x^2 = -\hbar^2 \frac{\partial^2}{\partial \phi^2}$ with positive eigenvalues, and similarly for L_y^2 [64].

$$\begin{aligned} L^2 Y^{m_1} &= (L_x^2 + L_y^2 + L_z^2) Y^{m_1} \\ &= (L_- L_+ + L_z^2 + \hbar L_z) Y^{m_1} \\ &= (L_- L_+ + \hbar^2 (m_1^2 + m_1)) Y^{m_1} \end{aligned} \tag{3.29}$$

since $L_+ Y^{m_1} = 0$ from Eq. (3.28) then

$$L^2 Y^{m_1} = \hbar^2 (m_1^2 + m_1) Y^{m_1} \tag{3.30}$$

Repeating for L_-, it is found that

$$L^2 Y^{m_2} = \hbar^2 (m_2^2 - m_2) Y^{m_2} \tag{3.31}$$

For Eq. (3.30) and Eq. (3.31) to be simultaneously true then

$$m_2 (m_2 - 1) = m_1 (m_1 + 1) \tag{3.32}$$

implying that

$$m_2 = -m_1 \quad \text{or} \quad m_2 = m_1 + 1$$

The second solution is inadmissible as m_1 is the largest positive value of $L_z/\hbar$, and so the maximum negative value of m must be the negative of the maximum positive value. Convention represents the term m_1 by the integer l, and so

$$L^2 Y_l^m = \hbar^2 l(l+1) Y_l^m \tag{3.33}$$

It can be useful to visualise $\sqrt{l(l+1)}$ as the magnitude of the vector $\boldsymbol{L}/\hbar$ and m as the projection of $\boldsymbol{L}$ onto the z-axis, as illustrated in Figure 3-2.

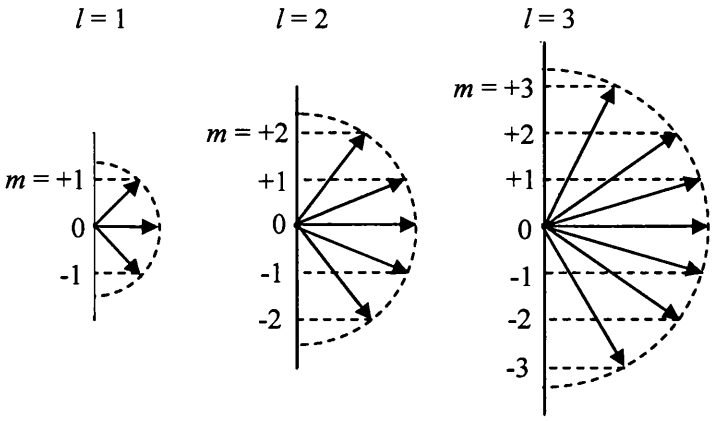


Figure 3-2 Relation between the l and m quantum numbers

Although, importantly, even when $|m| = l$ the angular momentum does not entirely orientate along the z-axis, having residual x- and y- components. This is a consequence of the non-commutation of L_x and L_y with L_z, and implies that they cannot be fixed at any particular value in a state where

L_z is definite. However, it must be made clear that it is not the case that the angular momentum has some definite direction, only with the caveat that that direction cannot be measured with total precision. Rather, for definite L^2 and L_z, the cone of directions corresponding to L_x and L_y consistent with given L^2 and L_z, are distributed through simultaneously.

Having determined L^2's eigenvalues, it now remains to ascertain the exact form of its eigenfunctions. As discussed previously, given an eigenfunction of L_z, it is possible to generate new eigenfunctions of L_z belonging to the same L^2 via repeated operation of the ladder operators, $L_\pm$.

For $m = l$, $L_+ Y_l^l = 0$, and so given that $Y_l^l(\theta,\phi) = A_l^l(\theta)e^{il\phi}$, then

$$i\frac{\partial Y_l^l}{\partial \phi} = -lY_l^l \tag{3.34}$$

which from the representation of L_+ in spherical polar coordinates,

$$L_+ = L_x + iL_y = \hbar e^{i\phi}\left[\frac{\partial}{\partial\theta} + i\cot\theta\frac{\partial}{\partial\phi}\right] \tag{3.35}$$

implies that

$$\frac{\partial Y_l^l}{\partial\theta} = l\cot\theta Y_l^l \tag{3.36}$$

On integrating with respect to θ,

$$\begin{aligned} &\ln Y_l^l = l\ln(\sin\theta) + K(\phi) \\ &\Rightarrow Y_l^l = g(\phi)\sin^l\theta \end{aligned} \tag{3.37}$$

$g(\phi)$ must make Y_l^l an eigenfunction of L_z with a corresponding eigenvalue of $\hbar l$, so $g(\phi) = e^{il\phi}$, correct to an arbitrary multiplicative constant, which must be determined from normalisation. Therefore

$$Y_l^l = Ce^{il\phi}\sin^l\theta \tag{3.38}$$

Now, applying the lowering operator L_-, any arbitrary state can be achieved. As such

$$\begin{aligned} L_-Y_l^l &= \hbar e^{i\phi}\left[-\frac{\partial}{\partial\theta} + i\cot\theta\frac{\partial}{\partial\phi}\right]C\sin^l\theta\, e^{il\phi} \\ &= \hbar Ce^{i(l-1)\phi}\left[-\frac{\partial}{\partial\theta} + l\cot\theta\right]\sin^l\theta \end{aligned} \tag{3.39}$$

Since for an arbitrary function $f(x)$

$$\left[\frac{d}{dx}+l\cot x\right]f(x)=\frac{1}{\sin^{l}x}\frac{d}{dx}\left[\sin^{l}x\,f(x)\right] \tag{3.40}$$

it follows that

$$Y_l^{l-1}=-\hbar C'\frac{e^{i(l-1)\phi}}{\sin^{l}\theta}\frac{\partial}{\partial\theta}\sin^{2l}\theta \tag{3.41}$$

Repeating,

$$Y_l^{l-2}=(-1)^2\hbar^2C''\frac{e^{i(l-2)\phi}}{\sin^{l-1}\theta}\frac{\partial}{\partial\theta}\left(\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\sin^{2l}\theta\right) \tag{3.42}$$

and in general,

$$Y_l^m=(-1)^{l-m}\hbar^{l-m}C^{l-m}\frac{e^{im\phi}}{\sin^{m}\theta}\left(\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\right)^{l-m}\sin^{2l}\theta \tag{3.43}$$

To determine Y_l^m exactly, the normalisation parameters $C^{l\text{-}m}$ have to be set to ensure

$$\int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi}Y_{l_1}^{m_1}{}^{*}(\theta,\phi)Y_{l_2}^{m_2}(\theta,\phi)\sin\theta\,d\theta\,d\phi=\delta_{l_1,l_2}\delta_{m_1,m_2} \tag{3.44}$$

In order to do this, again look at the action of $L_{\pm}$. It can be expressed as

$$L_{\pm}Y_l^m=C_{\pm}^{lm}Y_l^{m\pm1} \tag{3.45}$$

then

$$\begin{aligned}\left|C_{\pm}^{lm}\right|^2\int\left(Y_l^{m\pm1}\right)^{*}Y_l^{m\pm1}d\Omega&=\int\left(L_{\pm}Y_l^m\right)^{*}L_{\pm}Y_l^m d\Omega\\&=\int Y_l^{m^{*}}L_{\mp}L_{\pm}Y_l^m d\Omega\\&=\int Y_l^{m^{*}}(L^2-L_z^2\mp\hbar L_z)Y_l^m d\Omega\\&=\hbar^2\left(l(l+1)-m(m\pm1)\right)\end{aligned} \tag{3.46}$$

where

$$d\Omega=\sin\theta d\theta d\phi \tag{3.47}$$

and so with a convenient choice of phase

$$\begin{aligned}C_{\pm}^{lm}&=\hbar\sqrt{l(l+1)-m(m\pm1)}\\&=\hbar\sqrt{(l\mp m)(l\pm m+1)}\end{aligned} \tag{3.48}$$

Therefore

$$\begin{aligned}\hbar^{l-m}C^{l-m}&=\hbar^{l-m}C\left[C_{-}^{ll}C_{-}^{ll-1}C_{-}^{ll-2}\cdots\times C_{-}^{lm+1}\right]^{-1}\\&=\frac{C}{\sqrt{(l+l)(l-l+1)\times(l+l-1)(l-l+1+1)\times\cdots\times(l+m+1)(l-m)}}\end{aligned}$$

$$= \frac{C}{\sqrt{2l \times 2(2l-1) \times 3(2l-2) \times \cdots \times (l-m)(l+m+1)}}$$
$$= C\sqrt{\frac{(l+m)!}{(2l)!(l-m)!}} \tag{3.49}$$

For $m = l$, if the relevant spherical harmonic were correctly normalised then

$$\int Y_l^{l^*} Y_l^l \, d\Omega = 1 \tag{3.50}$$

so that

$$CC^* \int_0^{2\pi} \int_0^{\pi} \sin^{2l+1} d\theta \, d\phi = 1 \tag{3.51}$$

Evaluating the θ integral as a beta function (Appendix I, Gamma and Beta Functions), then

$$|C| = \sqrt{\frac{(2l+1)!!}{4\pi(2l)!!}} = \frac{\sqrt{(2l)!}}{2^l l!}\sqrt{\frac{2l+1}{4\pi}} \tag{3.52}$$

With the phases chosen to agree with the conventional Condon-Shortley phase, using for instance $C = (-1)^l |C|$ [65], then

$$Y_l^m = (-1)^{l-m}(-1)^l \frac{\sqrt{(2l)!}}{2^l l!}\sqrt{\frac{2l+1}{4\pi}}\sqrt{\frac{(l+m)!}{(2l)!(l-m)!}} \frac{e^{im\phi}}{\sin^m \theta}\left(\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\right)^{l-m} \sin^{2l}\theta$$
$$= (-1)^m \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}} P_l^m(\cos\theta) e^{im\phi} \qquad m \geq 0 \tag{3.53}$$

with

$$Y_l^{-m} = (-1)^m (Y_l^m)^* \tag{3.54}$$

and where the associated Legendre polynomials are defined by

$$P_l^m(\cos\theta) = \frac{1}{2^l l!}\frac{(l+m)!}{(l-m)!}\frac{1}{\sin^m\theta}\left(\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\right)^{l-m} \sin^{2l}\theta \qquad m \geq 0 \tag{3.55}$$

which for negative m are given by

$$P_l^{-m}(\cos\theta) = (-1)^m \frac{(l-m)!}{(l+m)!} P_l^m(\cos\theta) \tag{3.56}$$

The associated Legendre functions obey the orthogonality relation (see Appendix II, Orthogonality of the Associated Legendre Functions for identical m)

$$\int_0^{\pi} P_p^m(\cos\theta) P_q^m(\cos\theta) \sin\theta \, d\theta = \frac{2}{2p+1}\frac{(p+m)!}{(p-m)!}\delta_{p,q} \tag{3.57}$$

and so the function

$$\zeta_l^m(\cos\theta) = \sqrt{\frac{2l+1}{2}\frac{(l-m)!}{(l+m)!}} P_l^m(\cos\theta) \qquad -l \le m \le l \tag{3.58}$$

is therefore orthonormal with respect to the polar angle θ. Similarly for azimuthal angle ϕ,

$$\int_0^{2\pi} e^{-im_1\phi} e^{im_2\phi} d\phi = 2\pi\delta_{m_1,m_2} \tag{3.59}$$

implying that the function

$$\xi_m(\phi) = \frac{e^{im\phi}}{\sqrt{2\pi}} \tag{3.60}$$

is also orthonormal. The product of the two functions, $Y_l^m(\theta,\phi)$, is therefore orthonormal over the spherical surface, as required by Eq. (3.44).

The associated Legendre functions and spherical harmonics for l = 0 to 3 are listed explicitly in Table 3-1.

Table 3-1 Table of associated Legendre functions and spherical harmonics for l = 0 to 3. For m = 0, the associated Legendre functions are equivalent to the Legendre functions for the same value of l.

Associated Legendre Functions, $P_l^m(\theta)$		Spherical Harmonics, $Y_l^m(\theta,\phi)$
$P_0^0(\theta) = 1$		$Y_0^0(\theta,\phi) = \frac{1}{\sqrt{4\pi}}$
$P_1^0(\theta) = \cos\theta$		$Y_1^0(\theta,\phi) = \sqrt{\frac{3}{4\pi}}\cos\theta$
$P_1^1(\theta) = \sin\theta$	$P_1^{-1}(\theta) = -\frac{1}{2}P_1^1(\theta)$	$Y_1^{\pm1}(\theta,\phi) = \mp\sqrt{\frac{3}{8\pi}}\sin\theta\, e^{\pm i\phi}$
$P_2^0(\theta) = \frac{1}{2}(3\cos^2\theta - 1)$		$Y_2^0(\theta,\phi) = \sqrt{\frac{5}{16\pi}}(3\cos^2\theta - 1)$
$P_2^1(\theta) = 3\cos\theta\sin\theta$	$P_2^{-1}(\theta) = -\frac{1}{6}P_2^1(\theta)$	$Y_2^{\pm1}(\theta,\phi) = \mp\sqrt{\frac{5}{24\pi}}3\cos\theta\sin\theta\, e^{\pm i\phi}$
$P_2^2(\theta) = 3\sin^2\theta$	$P_2^{-2}(\theta) = \frac{1}{24}P_2^2(\theta)$	$Y_2^{\pm2}(\theta,\phi) = \sqrt{\frac{15}{32\pi}}\sin^2\theta\, e^{\pm i2\phi}$
$P_3^0(\theta) = \frac{1}{2}(5\cos^2\theta - 3)\cos\theta$		$Y_3^0(\theta,\phi) = \sqrt{\frac{7}{16\pi}}(5\cos^2\theta - 3)\cos\theta$
$P_3^1(\theta) = \frac{3}{2}(5\cos^2\theta - 1)\sin\theta$	$P_3^{-1}(\theta) = -\frac{1}{12}P_3^1(\theta)$	$Y_3^{\pm1}(\theta,\phi) = \mp\sqrt{\frac{21}{64\pi}}(5\cos^2\theta - 1)\sin\theta\, e^{\pm i\phi}$
$P_3^2(\theta) = 15\cos\theta\sin^2\theta$	$P_3^{-2}(\theta) = \frac{1}{120}P_3^2(\theta)$	$Y_3^{\pm2}(\theta,\phi) = \sqrt{\frac{105}{32\pi}}\cos\theta\sin^2\theta\, e^{\pm i2\phi}$
$P_3^3(\theta) = 15\sin^3\theta$	$P_3^{-3}(\theta) = -\frac{1}{720}P_3^3(\theta)$	$Y_3^{\pm3}(\theta,\phi) = \mp\sqrt{\frac{35}{64}}\sin^3\theta\, e^{\pm i3\phi}$

The probability density distribution of an electron at a point (r, θ, ϕ) is dependant on the square of its wave function, and as such, the angular probability density is proportional to $P_l^m(\cos\theta)^2$, examples of which are given in Figure 3-3.

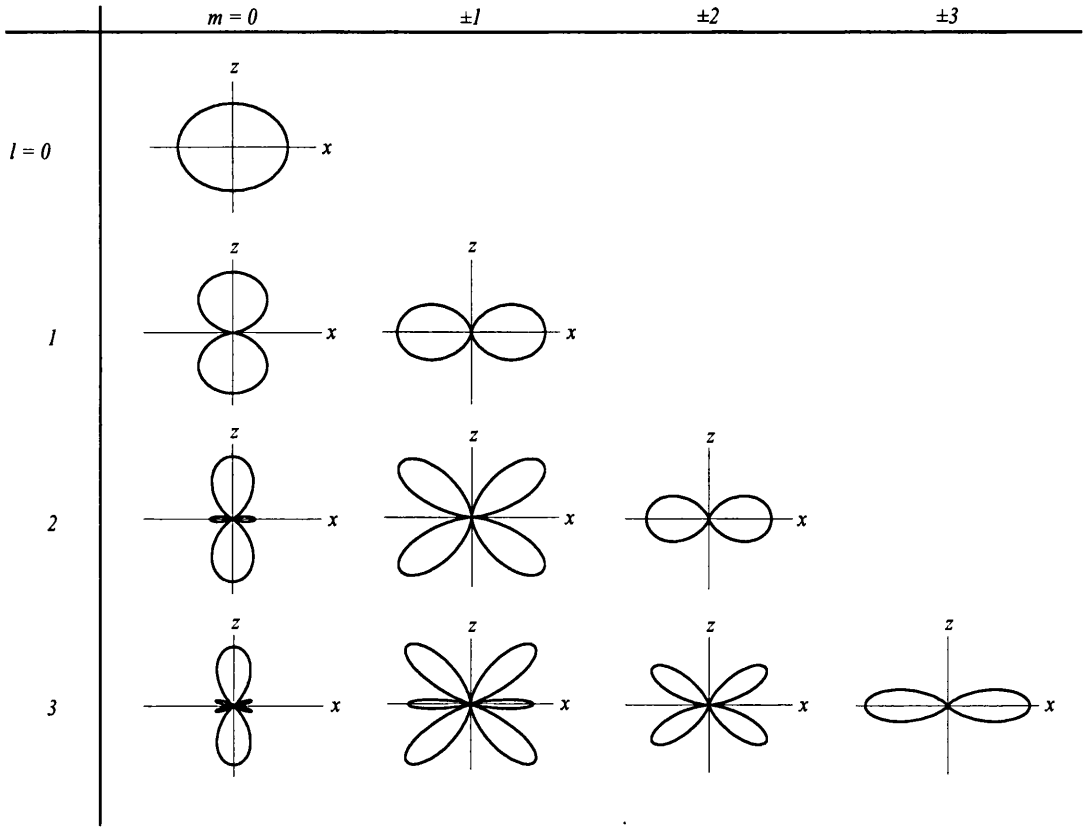


Figure 3-3 Plots of $P_l^m(\cos\theta)^2$, representing the angular probability density, for $l = 0$ to 3

As the value of m increases, the probability density shifts from the z-axis towards the x-axis, referred to as the equatorial plane. When $|m| = l$, then $|P_l^l(\cos\theta)|^2 \propto \sin^{2l}\theta$ and the function has a maximum at $\theta = \pi/2$ (on the equatorial plane); the lower m, the more diffuse this maximum is, meaning that the range of latitude angles, θ, over which the particle can be found increases as m is decreased. Indeed, at $l = 0$, $m = 0$ the particle covers the full range of θ evenly.

For the total angular momentum quantum number, as l increases, the width of the peak decreases, and so the function maximum becomes sharper. Ergo, at large l and large m the classical limit is approached and the particle can be considered to tend to an orbit almost exactly in the equatorial plane. Small fluctuations remain however as although the total and z-component of the angular

momentum are defined exactly, the x- and y- components remain unspecified, and so the angular momentum vector cannot be perfectly aligned along any one axis.

The exact forms of the angular density probability distributions are merely extensions of Figure 3-3 into three dimensions, see for instance Figure 3-4 in which $Y_0^{0\,2}$, $Y_2^{0\,2}$ and $Y_2^{\pm 1\,2}$ have been modelled in full for the purpose of illustration.

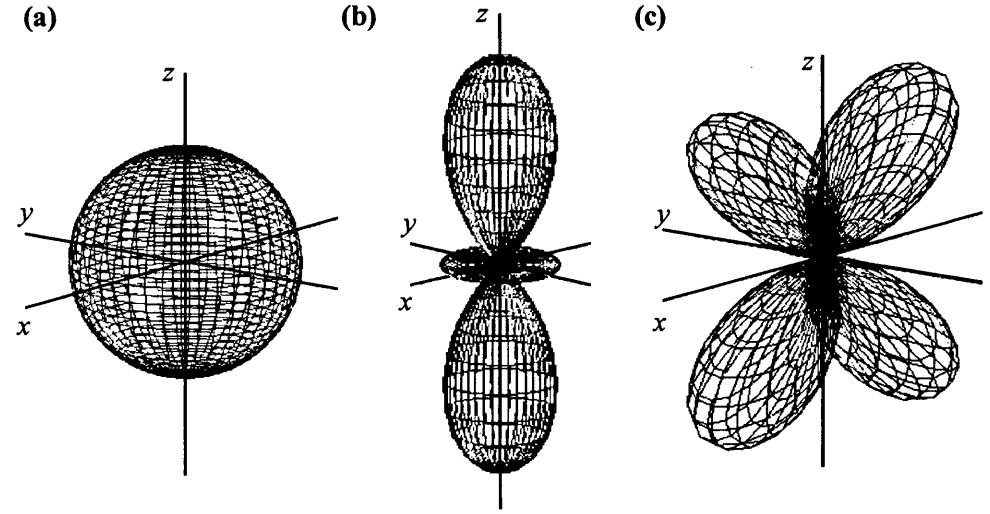


Figure 3-4 Examples of the angular probability density, $Y_l^m(\theta,\phi)^2$, for (a) Y_0^0 (b) Y_2^0 and (c)

An alternative formula for the associated Legendre function, expressed in terms of its equivalent Legendre function is

$$P_l^m(x) = (1-x^2)^{m/2}\frac{d^m}{dx^m}P_l(x) \tag{3.61}$$

where the $P_l(x)$ can be given by Rodrigues' formula

$$P_l(x) = \frac{1}{2^l l!}\frac{d^l}{dx^l}(x^2-1)^l \tag{3.62}$$

and x is of course cos θ, see for instance Arfken and Weber [65]. Then, given that the highest power of x in the expansion of $(x^2\text{-}1)^l$ is $2l$, it follows that $d^mP_l(x)/dx^m$ is a polynomial of degree l - m [64]. This is multiplied by the factor $(1 - x^2)^{m/2} = \sin^m\theta$, so therefore, for x between +1 and -1, $P_l^m(x)$ has l - m zeros. From the real part of $e^{im\phi}$, cos $m\phi$, it is clear that in the same interval the azimuthal dependant function has m zeros, and thus the complete angular wave function, the spherical harmonic, has $(l - m) + m = l$ zeros (or nodes) between $x = 1$ and -1. This is important in the selection of appropriate wave functions discussed in Section 6.2.

In summary then, recalling Eq.'s (3.5) and (3.6), in a spherically symmetric space where V is a function of r only, the wave function solutions to Schrödinger's equation for specified angular momenta quantum numbers, l and m, are given by

$$\psi_{nlm}(r,\theta,\phi) = f_{nl}(r)Y_l^m(\theta,\phi) \tag{3.63}$$

The angular function, $Y_l^m(\theta,\phi)$, is termed the spherical harmonic and is defined by Eq. (3.53)

$$Y_l^m = (-1)^m \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}} P_l^m(\cos\theta)e^{im\phi} \qquad m \geq 0$$

while the radial component, $f_{nl}(r)$, must satisfy the equation

$$\left[\frac{d^2}{dr^2} + \frac{2}{r}\frac{d}{dr} - \frac{l(l+1)}{r^2}\right] f_{nl}(r) + \frac{2m}{\hbar^2}\left[E_n - V(r)\right] f_{nl}(r) = 0 \tag{3.64}$$

the solution of which occupies the next section.

3.2 The Radial Component of the Schrödinger Equation Solution

The form of $f_{nl}(r)$ satisfying Eq. (3.64) is dependant on the potential $V(r)$ and as such does not have standard solution in the way that Eq. (3.5) does. This section outlines a numerical approach to ascertain the approximate eigenstates of Eq. (3.64), where its 1-dimensional continuous space is broken up into $N+1$ discrete, infinitesimal points of predetermined spacing over which its derivates and potential can be estimated. Representing part of this discrete space as a square (N-1)×(N-1) matrix, M, f_{nl} and E_{nl} follow from the solution of the standard eigenvector/value problem $Mv = \lambda v$.

In a similar fashion to Sub-Section 2.3.1, let the spatial region of interest be divided into a mesh of N+1 points, beginning at zero and set an evenly spaced a distance of Δ apart, then a function f at the site r_{i+1}, abbreviated here as f_{i+1}, can be expanded as a Taylor series about the site r_i as

$$f_{i+1} = f_i + \Delta f_i' + \frac{\Delta^2}{2!} f_i'' + \frac{\Delta^3}{3!} f_i''' + \frac{\Delta^4}{4!} f_i^{iv} + \cdots \tag{3.65}$$

f' denoting the first derivative of f with respect to r and so on[i]. Similarly, for f at site r_{i-1}

$$f_{i-1} = f_i - \Delta f_i' + \frac{\Delta^2}{2!} f_i'' - \frac{\Delta^3}{3!} f_i''' + \frac{\Delta^4}{4!} f_i^{iv} - \cdots \tag{3.66}$$

On subtracting Eq. (3.65) from Eq. (3.66) then

$$f_{i+1} - f_{i-1} = 2\left[\Delta f_i' + \frac{\Delta^3}{3!} f_i''' + \frac{\Delta^5}{5!} f_i^{v} + \cdots\right] \tag{3.67}$$

and by rearranging, the first derivative of f can be approximated by

$$f_i' = \frac{f_{i+1} - f_{i-1}}{2\Delta} + O(\Delta^2) \tag{3.68}$$

to second order in Δ. Likewise, on addition of Eq.'s (3.65) and (3.66), the second derivative can be approximated by

$$f_i'' = \frac{f_{i+1} - 2f_i + f_{i-1}}{\Delta^2} + O(\Delta^2) \tag{3.69}$$

Via these two relations, Eq. (3.68) and Eq. (3.69), Eq. (3.64) can be estimated on this mesh to second order as

---

[i] Historically, the prime notation $y' = dy/dx$ was introduced in the late eighteenth by Lagrange as an abbreviation of Leibniz's dy/dx notation for a derivative [65]

$$\left(\frac{2}{\Delta^2}+\frac{2m}{\hbar^2}V_i+\frac{l(l+1)}{r_i^2}\right)f_i^{nl}-\left(1+\frac{\Delta}{r_i}\right)\frac{f_{i+1}^{nl}}{\Delta^2}-\left(1-\frac{\Delta}{r_i}\right)\frac{f_{i-1}^{nl}}{\Delta^2}=\frac{2m}{\hbar^2}E_n f_i^{nl}+O(\Delta^2) \tag{3.70}$$

As with Burden *et al.* [66]'s finite difference method, a (N-1)×(N-1) un-symmetric tri-diagonal matrix, M, is formed

$$M=\begin{bmatrix} b_1 & c_1 & & & & & & \\ a_2 & b_2 & c_2 & & & O & & \\ & a_3 & b_3 & c_3 & & & & \\ & & & \ddots & & & & \\ & & & & a_{N-3} & b_{N-3} & c_{N-3} & \\ & & O & & & a_{N-2} & b_{N-2} & c_{N-2} \\ & & & & & & a_{N-1} & b_{N-1} \end{bmatrix} \tag{3.71}$$

where

$$\begin{aligned} a_i &= -\frac{1}{\Delta^2}\left(1+\frac{\Delta}{r_1}\right) \\ b_i &= \frac{2}{\Delta}+\frac{2m}{\hbar^2}V_i+\frac{l(l+1)}{r_i^2} \\ c_i &= -\frac{1}{\Delta^2}\left(1-\frac{\Delta}{r_1}\right) \end{aligned} \tag{3.72}$$

and letting

$$\begin{aligned} v_i^{nl} &= f_i^{nl} \qquad 1<i<N-1 \\ \lambda_{nl} &= \frac{2m}{\hbar^2}E_{nl} \end{aligned} \tag{3.73}$$

then Eq. (3.70) can be represented by

$$(M-\lambda_{nl}I)v_{nl}+\left(a_1 f_0^{nl}+c_{N-1}f_N^{nl}\right)=0 \tag{3.74}$$

As $r\rightarrow\infty$, $f_{nl}(r)$ is required to tend to zero since the normalisation constraint requires

$$\int_0^\infty f_{nl}^*(r)f_{nl}(r)r^2dr=1 \tag{3.75}$$

if f_{nl} is correctly normalised. From this, if $r_i\rightarrow\infty$ as $i\rightarrow\infty$, then f_i^{nl} must also tend to zero in the same limit. Therefore for sufficiently large N, $f_N^{nl}\approx 0$, and $c_{N-1}f_N^{nl}$ in Eq. (3.74) can be set to zero for all l.

The boundary value at the centre of the potential is a little more difficult. At r = 0, the term $l(l+1)f_{nl}(r)/r^2$ will be infinite unless l or $f_{nl}(0)$ is equal to zero. As consequence of the spherical

symmetry and the physical necessity for a smooth wave function, if $f_{nl}(0)$ is not equal to zero, then its first derivative at this point must i.e. $df_{nl}(0)/dr = 0$. Thus, if $l > 0$ then $a_1 f_0^{nl} = 0$ and Eq. (3.74) reduces to

$$Mv_{nl} = \lambda_{nl} v_{nl} \tag{3.76}$$

For $l = 0$, if $df_{nl}(0)/dr = 0$ then $f_0^{nl} \approx f_1^{nl}$. Defining

$$\beta_1 = b_1 + a_1$$

then substituting β_1 for b_1 in M, now referred to as Q, Eq. (3.74) becomes

$$Qv_{nl} = \lambda_{nl} v_{nl} \tag{3.77}$$

Should f_0^{nl} equal zero and not the derivative, then for a sufficiently small Δ,

$$\left|f_1^{nl}\right| = \left|f_0^{nl}\right| + \delta \tag{3.78}$$

where δ is a small quantity which tends to zero as Δ tends to zero. Thus $f_1^{nl} \approx f_0^{nl} = 0$ and therefore $Q \approx M$ and so Eq. (3.77) alone is adequate to represent Eq. (3.74) in this case.

Equations (3.76) and (3.77) are standard eigenvector/value problems which can be solved via several approaches, such as a QR-factorisation technique (see Appendix II, Introduction to Eigenvalues and Eigenvectors) to determine the eigenvalues, accompanied by a process such as the inverse power method (see also Appendix II, Introduction to Eigenvalues and Eigenvectors), to ascertain the eigenvectors. However, any linear algebra package, for instance LAPACK [67] (or alternatively if M or Q is sparse, ARPACK [68]), should contain suitable subroutines to tackle these two equations.

f_{nl} and E_{nl} can follow immediately from v_{nl} and λ_{nl} via Eq. (3.73) and the appropriate boundary values, to provide tenable approximations for the eigenstates of Eq. (3.64). However, the Hamiltonian H is a Hermitian operator and as such, its eigenvectors, ψ_{nlm}, will span the vector space on which it is defined. Moreover, these eigenvectors can be chosen to be a complete orthonormal set[ii] [69], this orthonormality forming an essential feature of the DFT processes applied in later Chapters. The angular components of these eigenvectors, the spherical harmonics, are already orthonormal with respect to the quantum numbers l and m, therefore it is enough to require f_{nl} to be orthonormal with respect to n for given l. Ergo, the most suitable radial wave

[ii] A set of vectors $\{x_1, x_2, \ldots\}$ is deemed to be orthonormal if $(x_i, x_j) = \delta_{ij}$ for all i and j.

functions are formed from v_{nl} and the relevant boundary values, and then orthoganalised and normalised to satisfy Eq. (3.75). One possible orthogonalisation procedure, described by Weissbluth [70] is particularly suitable due to its simplicity.

A transformation operator, A, is required that when applied to the non-orthonormal set of orbitals $f_{\mu l}$, formed from $v_{\mu l}$ as above, will generate a new set of orthonormal orbitals g_{nl}, such that

$$g_{nl}(r) = \sum_{\mu} f_{\mu l}(r) A^{l}_{\mu n} \tag{3.79}$$

$$\int_0^\infty g^*_{nl}(r) g_{ml}(r) r^2 dr = \delta_{nm} \tag{3.80}$$

In an effort to keep the following arguments as clear as possible, the angular momentum quantum number l will be dropped from the sub- and super- scripts where practicable, although naturally it is implicit in all the operators below, acting as they do on reduced sets of vectors, all of equal l.

Defining the overlap integral $S_{\mu\nu}$ such that

$$S_{\mu\nu} = \int_0^\infty f^*_{\mu l}(r) f_{\nu l}(r) r^2 dr \tag{3.81}$$

then

$$\begin{aligned}
\delta_{nm} &= \int_0^\infty g^*_n(r) g_m(r) r^2 dr \\
&= \sum_{\mu\nu} A^*_{n\mu} \left(\int_0^\infty f^*_\mu(r) f_\nu(r) r^2 dr \right) A_{\nu m} \\
&= \sum_{\mu\nu} A^*_{n\mu} S_{\mu\nu} A_{\nu m}
\end{aligned}$$

i.e.

$$I = A^\dagger S A \tag{3.82}$$

Assuming that the operator A is Hermitian, $A^\dagger = A$, then

$$A = S^{1/2} \tag{3.83}$$

If S could be diagonalised via unitary transforms[iii] (see also Appendix II, Introduction to Eigenvalues and Eigenvectors), as is always possible since S is either Hermitian or real symmetric

[iii] A unitary transformation is a linear transformation $y = Ux$ with a unitary matrix (operator) U. U is deemed unitary if its inverse is equal to its adjoint i.e. $U^{-1} = U^\dagger$. A usefull property of unitary matricies is

as can be seen from Eq. (3.81), then

$$S' = U^{-1} S U \tag{3.84}$$

where S' is a diagonal matrix, the S'_{ii} components equal of course to the eigenvalues of S. The matrix $(S')^{-1/2}$ is naturally also diagonal, its $(S')^{-1/2}_{ii}$ components simply equal to $1/\sqrt{S'_{ii}}$, and then finally $S^{-1/2}$ straight forwardly follows from

$$S^{-1/2} = U (S')^{-1/2} U^{1} \tag{3.85}$$

A is now defined, and the orthonormal eigenvectors g_{nl} can be formed via linear combinations of $f_{\mu l}$ via Eq. (3.79).

Applying the radial operator to the new set of eigenvectors, g_{nl}

$$\begin{aligned} -\left[\frac{d^2}{dr^2} + \frac{2}{r}\frac{d}{dr} - \frac{l(l+1)}{r^2} - \frac{2m}{\hbar^2}V(r)\right] g_{nl}(r) &= -\sum_{\mu}\left[\frac{d^2}{dr^2} + \frac{2}{r}\frac{d}{dr} - \frac{l(l+1)}{r^2} - \frac{2m}{\hbar^2}V(r)\right] f_{\mu l}(r) A^{l}_{\mu n} \\ &= \frac{2m}{\hbar^2}\sum_{\mu} E_{\mu l} f_{\mu l}(r) A^{l}_{\mu n} \\ &= \frac{2m}{\hbar^2}\varepsilon_{nl} g_{nl}(r) \end{aligned}$$

then the corresponding eigenvalues are

$$\varepsilon_{nl} = \frac{\sum_{\mu} E_{\mu l} f_{\mu l}(r) A^{l}_{\mu n}}{\sum_{\mu} f_{\mu l}(r) A^{l}_{\mu n}} \tag{3.86}$$

In all probability, some small variation in ε_{nl} will occur over values of r_i due to the unavoidable effects of cumulative computational rounding errors and so forth, and so the mean should be taken over the mesh for each n and l. Thus, for the discretised space of N points

$$\begin{aligned} g^{nl}_{i} &= \sum_{\mu} f^{\mu l}_{i} A^{l}_{\mu n} \\ \varepsilon_{nl} &= \frac{1}{N}\sum_{i=0}^{N} \frac{\sum_{\mu} E_{\mu l} f^{\mu l}_{i} A^{l}_{\mu n}}{\sum_{\mu} f^{\mu l}_{i} A^{l}_{\mu n}} \end{aligned} \tag{3.87}$$

In practise, it is more efficient to construct the eigenvectors from Eq. (3.87) and then compute the eigenvalues anew from g and M (or Q).

---

that if a matrix A is Hermitian, then there will exist a diagonal matrix $U^{1}AU$, the diagonal elements of which are the eigenvalues of A. A proof of this is given in Byron and Fuller [69].

The labelling f, g, E, ε is arbitrary and so the orthonormal eigenstates satisfying the radial equation are re-labelled f and E, and are referred to as such throughout the remainder of this document.

In summary, a finite difference method has been described that produces approximations of the eigenstates satisfying Eq. (3.64),

$$\left[\frac{d^2}{dr^2}+\frac{2}{r}\frac{d}{dr}-\frac{l(l+1)}{r^2}\right]f_{nl}(r)+\frac{2m}{\hbar^2}\left[E_{nl}-V(r)\right]f_{nl}(r)=0$$

the set of vectors $f_{nl}(r)$ orthonormal with respect to n for given l. Then with the spherical harmonics of the proceeding section satisfying Eq. (3.33),

$$L^2Y_l^m(\theta,\phi)=\hbar^2 l(l+1)Y_l^m(\theta,\phi)$$

$Y_l^m(\theta,\phi)$ orthonormal with respect to m and l, a fully defined approximation for the wave function of the Schrödinger equation Eq. (3.1) can be given for any potential $V(r)$,

$$\nabla^2\psi_{nlm}(r,\theta,\phi)+\frac{2m}{\hbar^2}\left(E_{nl}-V(r)\right)\psi_{nlm}(r,\theta,\phi)=0$$

$$\psi_{nlm}(r,\theta,\phi)=f_{nl}(r)Y_l^m(\theta,\phi)$$

ψ_{nlm} satisfying the requirements of orthonormality with respect to n, m and l.

3.3 Spherical Bessel Functions

By discretising the Schrödinger equation and solving for its eigenstates, the orbitals and energy values determined are those of the discrete Schrödinger equation rather than its continuous counterpart. For certain applications, discretisations preserving the Schrödinger equation can lead to results at odds with its continuous form, see for instance Flores [71], notwithstanding the effect of the mesh size on the accuracy of the eigenvalues/vectors. To judge the effectiveness then of the Schrödinger solver, the energy eigenvalues for a potential generated through the finite difference method should be compared to a set determined by alternate means from the continuous equation. A suitable test is the energy levels of a particle in a sphere with a finite square well potential, as represented in Figure 3-5, where the reference values can also be derived analytically.

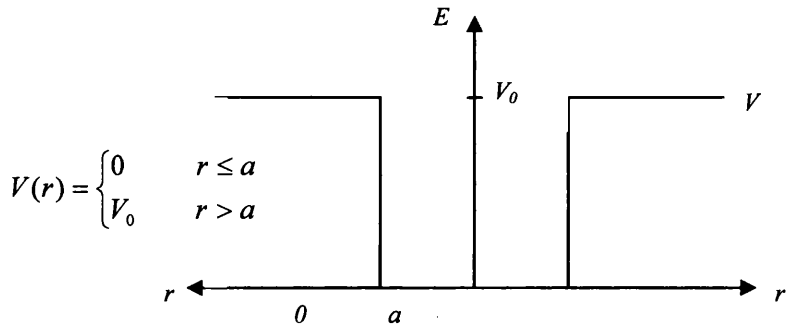


Figure 3-5 Schematic of finite square potential well for a particle in a sphere

Recalling the time-independent Schrödinger equation (TISE) in spherical polar co-ordinates as

$$\frac{-\hbar^2}{2m}\left(\frac{d^2 f_l}{dr^2}+\frac{2}{r}\frac{df_l}{dr}-\frac{l(l+1)}{r^2}f_l\right)+(V(r)-E)f_l=0$$

which for the square well, can be expressed such that

$$\frac{d^2 f_l}{dr^2}+\frac{2}{r}\frac{df_l}{dr}-\frac{l(l+1)}{r^2}f_l-q^2 f_l=0 \quad \text{where } q^2=\frac{2m}{\hbar^2}(V_0-E) \qquad \textit{outside well} \qquad (3.88a)$$

$$\frac{d^2 f_l}{dr^2}+\frac{2}{r}\frac{df_l}{dr}-\frac{l(l+1)}{r^2}f_l+k^2 f_l=0 \quad \text{where } k^2=\frac{2m}{\hbar^2}E \qquad \textit{inside well} \qquad (3.88b)$$

For the differential equation describing the well interior, the substitution $x = kr$ allows it to be represented as

$$\frac{d^2 f_l}{dx^2} + \frac{2}{x}\frac{df_l}{dx} + \left(1 - \frac{l(l+1)}{x^2}\right) f_l = 0 \tag{3.89}$$

For the solutions to this ODE, follow the powerfull and illuminating approach of Byron and Fuller [69] and consider the function $g_l(x)$

$$g_l(x) = \frac{C_l}{x^{l+1}} \oint_C \frac{e^{-ixz}}{(z+1)^{l+1}(z-1)^{l+1}} dz \tag{3.90}$$

on a closed contour containing both the points $z = \pm 1$ (else $g_l(x) = 0$ by Cauchy-Goursat's theorem[i]), C_l a constant. On differentiating,

$$g_l'(x) = -\frac{(l+1)C_l}{x^{l+2}} \oint \frac{e^{-ixz}}{(z^2-1)^{l+1}} dz + \frac{iC_l}{x^{l+1}} \oint \frac{ze^{-ixz}}{(z^2-1)^{l+1}} dz$$

$$g_l''(x) = \frac{(l+1)(l+2)C_l}{x^{l+3}} \oint \frac{e^{-ixz}}{(z^2-1)^{l+1}} dz + \frac{2i(l+1)C_l}{x^{l+2}} \oint \frac{ze^{-ixz}}{(z^2-1)^{l+1}} dz - \frac{C_l}{x^{l+1}} \oint \frac{z^2 e^{-ixz}}{(z^2-1)^{l+1}} dz$$

then by defining the function $q_l(x)$ as

$$q_l(x) = g_l''(x) + \frac{2}{x} g_l'(x) + \left(1 - \frac{l(l+1)}{x^2}\right) g_l(x) \tag{3.91}$$

to prove that $g_l(x)$ can indeed satisfy Eq. (3.89), it is sufficient to show that $q_l(x) = 0$. From g_l and its derivatives

$$q_l(x) = \frac{2ilC_l}{x^{l+2}} \oint \frac{ze^{-ixz}}{(z^2-1)^{l+1}} dz - \frac{C_l}{x^{l+1}} \oint \frac{e^{-ixz}}{(z^2-1)^l} dz$$

however

$$\frac{2lze^{-ixz}}{(z^2-1)^{l+1}} = -\frac{d}{dz}\left(\frac{e^{-ixz}}{(z^2-1)^l}\right) - ix\frac{e^{-ixz}}{(z^2-1)^l}$$

and so $q_l(x)$ can be written

$$q_l(x) = -\frac{iC_l}{x^{l+2}} \oint \frac{d}{dz}\left(\frac{e^{-ixz}}{(z^2-1)^l}\right) dz = 0$$

as the integral of a derivative over a closed contour vanishes[ii].

---

[i] Cauchy-Goursat Theorem: For a function, $f(z)$, analytic on and within a closed contour C, then

$$\oint_C f(z)dz = 0$$

[ii] A function is deemed analytic in a domain if it possesses a continuous derivative everywhere within that domain. Indeed, as shown through Cauchy's integral formula in Appendix II, Derivatives of Analytic Functions, this derivative will itself be analytic and so when integrated over a closed contour will equal zero as a consequence of the Cauchy-Goursat theorem

$g_l(x)$ is therefore a solution of Eq. (3.89) for any closed contour containing a singularity. The shape of the contour itself is immaterial. However to conform to conventional definitions, the choice of contours of Figure 3-6 and their corresponding multiplicative constants are used, resulting in

$$j_l(x) = \frac{1}{2\pi}\frac{(-2)^l l!}{x^{l+1}}\oint_{C_j}\frac{e^{-ixz}}{(z+1)^{l+1}(z-1)^{l+1}}dz \tag{3.92}$$

$$n_l(x) = \frac{1}{2\pi i}\frac{(-2)^l l!}{x^{l+1}}\oint_{C_n}\frac{e^{-ixz}}{(z+1)^{l+1}(z-1)^{l+1}}dz \tag{3.93}$$

the spherical Bessel function and spherical Neumann function of order l, although a little math is needed to cast these expressions in a more familiar form.

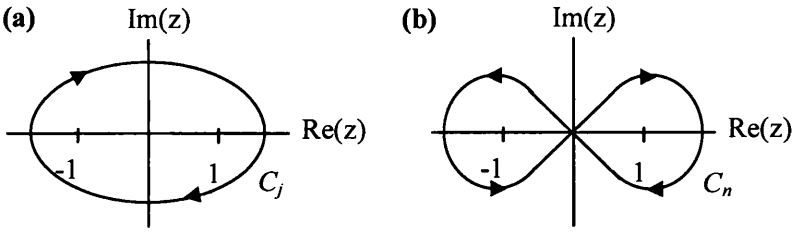


Figure 3-6 The contours of integration for the evaluation of (a) the spherical Bessel function (b) the spherical Neumann function

Let $y = xz$, then for $j_l(x)$

$$j_l(x) = \frac{1}{2\pi}\frac{(-2)^l l!}{x^{l+2}}\oint_{C_j}\frac{e^{-iy}}{\left((y/x)^2-1\right)^{l+1}}dy$$

$$= \frac{1}{2\pi}(-2)^l l! x^l \oint_{C_j}\frac{e^{-iy}}{y^{2l+2}}\left(1-\frac{x^2}{y^2}\right)^{-l-1}dy$$

C_j defined now as any anticlockwise contour enclosing the singularities $y = \pm x$.

Expanding $(1-x^2/y^2)^{-l-1}$ as a power series using the binomial theorem

$$\left(1-\frac{x^2}{y^2}\right)^{-l-1} = 1+(l+1)\frac{x^2}{y^2}+\frac{(l+1)(l+2)}{2}\frac{x^4}{y^4}+\ldots$$

$$= \sum_{m=0}^{\infty}\frac{(l+m)!}{m!l!}\left(\frac{x}{y}\right)^{2m} \tag{3.94}$$

which is only convergent in the region $|y| \geq x$, then a wise choice of C_j is a circle lying outside $y = x$. Employing Eq. (3.94) within the spherical Bessel function expression then yields

$$j_l(x) = \frac{1}{2\pi}(-2)^l l! x^l \left[\oint_{C_j} \frac{e^{-iy}}{y^{2l+2}} dy + x^2(l+1) \oint_{C_j} \frac{e^{-iy}}{y^{2l+4}} dy + \ldots \right]$$
$$= i(-2)^l l! x^l \left[\frac{(-i)^{2l+1}}{(2l+1)!} + (l+1)x^2 \frac{(-i)^{2l+3}}{(2l+3)!} + \frac{(l+1)(l+2)}{2} x^4 \frac{(-i)^{2l+5}}{(2l+5)!} + \ldots \right]$$
$$= 2^l x^l \sum_{m=0}^{\infty} (-1)^m \frac{(l+m)! x^{2m}}{m!(2l+2m+1)!} \qquad (3.95)$$

and so follows the conventional form of the spherical Bessel function [65]. The reduction of the integrals follows from the use of the residue theorem[iii], the residues determined from the b_l terms in the Laurent series expansions[iv] of each exponential term.

Due to the shape of the contour C_n the spherical Neumann function has to be approach a little differently, the binomial expansion, Eq. (3.94), used for its counterpart, $j_l(x)$, no longer convergent.

As demonstrated in Appendix II, Derivatives of Analytic Functions, a consequence of Cauchy's Integral formula is the expression

$$g^{(n)}(z_0) = \frac{n!}{2\pi i} \oint_C \frac{g(z)}{(z-z_0)^{n+1}} dz$$

which, when used to calculate the residue of a pole of order (l-1), in conjunction with Leibniz' theorem[v], allows Eq. (3.93) to be expressed as

---

[iii] The Residue Theorem: The integral of the function $f(z)$ around the closed contour C contain n singular points of $f(z)$, where n is finite, is equal to the sum of n integrals of $f(z)$ around circular contours containing only one singular point:

$$\oint_C f(z)dz = 2\pi i \sum_{j=1}^{n} R_j$$

where R_j is the residue at point z_j and defined as

$$R_j = \frac{1}{2\pi i} \oint_{C_j} f(z)dz$$

[iv] Laurent's Theorem: Let the function $f(z)$ be analytic through the closed annular region between the circles C_1 and C_2, each with common centre z_0, then at each point within the annulus, the function can be expanded as the convergent series

$$f(z) = \sum_{n=0}^{\infty} a_n (z-z_0)^n + \sum_{n=1}^{\infty} b_n (z-z_0)^{-n}$$

e.g. For the exponential e^z

$$e^z = \sum_{m=0}^{\infty} \frac{z^m}{m!}$$

[v] Leibniz theorem:

$$n_l(x)=\frac{(-2)^l l!}{l!x^{l+1}}\sum_{m=0}^{l}\frac{l!}{m!(l-m)!}\left(\left[\frac{d^{l-m}}{dz^{l-m}}e^{-ixz}\frac{d^m}{dz^m}\left(\frac{1}{z+1}\right)^{l+1}\right]_{z=1}-\left[\frac{d^{l-m}}{dz^{l-m}}e^{-ixz}\frac{d^m}{dz^m}\left(\frac{1}{z-1}\right)^{l+1}\right]_{z=-1}\right)$$
$$=\frac{(-2)^l}{x^{l+1}}\sum_{m=0}^{l}\frac{(-1)^m l!}{m!(l-m)!}(-ix)^{l-m}\frac{(l+m)!}{l!}\left(\frac{e^{ix}}{(-2)^{l+m+1}}-\frac{e^{-ix}}{2^{l+m+1}}\right)$$
$$=\frac{-1}{x^{l+1}}\sum_{m=0}^{l}\frac{(l+m)!}{m!(l-m)!}\frac{(-i)^{l-m}x^{l-m}}{2^{m+1}}\sum_{s=0}^{\infty}\frac{x^s}{s!}\left(i^s+(-1)^{m+l}(-i)^s\right)$$

where, again, the series expansions of the exponential terms has been used. With the removal of the common factor $1/x^{l+1}$ outside of the series expansion, the coefficients of remaining x^n terms, where n is odd, within the series expansion equals zero, and so $n_l(x)$ becomes

$$n_l(x)=\frac{-1}{x^{l+1}}\left(\frac{(l+1)(l+2)...2l}{2^l}+x^2\frac{l(l+1)(l+2)...(2l-2)}{2^l}+\frac{x^4}{2}\frac{(l-1)l(l+1)(l+2)...(2l-4)}{2^l}+...\right)$$
$$=\frac{(-1)^{l+1}}{2^l x^{l+1}}\sum_{m=0}^{\infty}\frac{(-1)^m(-1)^{l-m}(2l-2m)!}{m!(l-m)!}x^{2m} \qquad (3.96)$$
$$=\frac{(-1)^{l+1}}{2^l x^{l+1}}\sum_{m=0}^{\infty}\frac{(-1)^m(m-l)!}{m!(2m-2l)!}x^{2m}$$

the traditional expression for the spherical Neumann functions[vi].

Having two independent solutions to the well interior ODE, Eq. (3.88b), the general solution is then

$$f_l(kr)=a_1 j_l(kr)+a_2 n_l(kr) \qquad (3.97)$$

However, physically the wave function must remain regular as $kr \to 0$ and so, from the behaviour of each of the solutions in this limit,

$$j_l(x)\xrightarrow[x\to 0]{}\frac{2^l l!x^l}{(2l+1)!} \qquad n_l(x)\xrightarrow[x\to 0]{}-\frac{(2l)!}{2^l l!x^{l+1}} \qquad (3.98)$$

the constant a_2 must equal zero. Thus, for $r < a$,

$$\frac{d^n}{dx^n}[f(x)g(x)]=\sum_{m=0}^{n}\frac{n!}{m!(n-m)!}\frac{d^{n-m}}{dx^{n-m}}[f(x)]\frac{d^m}{dx^m}[g(x)]$$

[vi] The last step in Eq. (3.96) was achieved using the idenity

$$\frac{(s-n)!}{(2s-2n)!}=(-1)^{n-s}\frac{(2n-2s)!}{(n-s)!}$$

where s and n are integers where $s < n$. This follows from the relationship

$$\Gamma(z)\Gamma(1-z)=\frac{\pi}{\sin z\pi}$$

which can best be derived from Weierstrass' infinite product representation of the gamma function (see Appendix I, Gamma and Beta Functions) as demonstrated in Arfken and Weber [65].

$$f_l(kr) = a_1 j_l(kr) \tag{3.99}$$

For l = 0 to 3, this will have the form of the relevant spherical Bessel functions as illustrated in Figure 3-7.

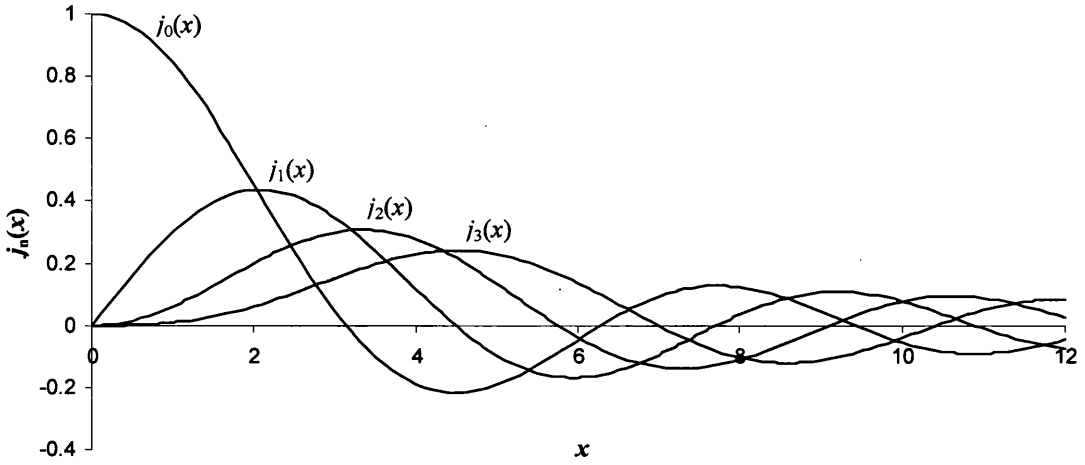


Figure 3-7 Plot of the spherical Bessel functions, $j_l(x)$, for l = 0 to 3

For the external ODE, Eq. (3.88a), the substitution $x = iqr$ reduces it to the familiar representation of Eq. (3.89) and so the solutions $j_l(iqr)$ and $n_l(iqr)$ hold. However, for the first solution, traditionally labelled $i_l(qr)$, convention requires multiplication by an additional factor of i^{-l}, ergo

$$i_l(qr) = i^{-l} j_l(iqr) \tag{3.100}$$

The second solution, traditionally labelled $k_l(qr)$, is more complex. As noted in the earlier discussion, in providing solutions to Eq. (3.89), the actual contour of integration of $g_l(x)$ in Eq. (3.90) is immaterial, provided that the said contour encloses the singularities at ±1. So should the contour C_j in Figure 3-6, be pinched at the origin, see Figure 3-8(a), and even separated into two circular contours, each enclosing a singularity in an anticlockwise direction, see Figure 3-8(b), then $j_l(x)$ will still form a valid solution.

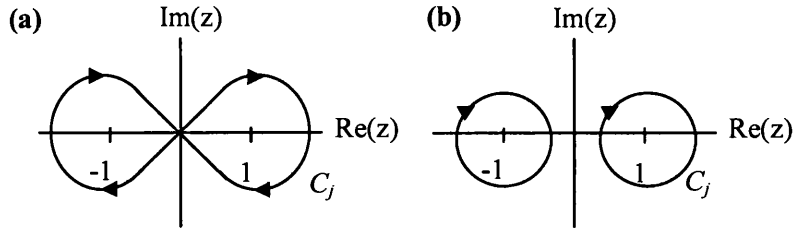


Figure 3-8 The contour of integration for the spherical Bessel function (a) deformed into an hour-glass (b) pinched into two circular contours

The contour of $n_l(x)$ can likewise be deformed into two circular contours, anti-clockwise about z = -1 but clockwise about $z = 1$.

In addition, if y_1 and y_2 are the two independent solutions of a homogeneous ODE, then, by the super-positioning principle [50], any linear combination Ay_1+By_2 is also a solution. Therefore, the functions

$$\begin{aligned} h_l^{(1)}(x) &= j_l(x) + in_l(x) \\ &= \frac{(-2)^l l!}{\pi x^{l+1}} \oint_{C_1} \frac{e^{-ixz}}{\left(z^2-1\right)^{l+1}} dz \end{aligned} \tag{3.101}$$

where C_1 is a contour, anticlockwise, about z = -1, and

$$\begin{aligned} h_l^{(2)}(x) &= j_l(x) - in_l(x) \\ &= \frac{(-2)^l l!}{\pi x^{l+1}} \oint_{C_2} \frac{e^{-ixz}}{\left(z^2-1\right)^{l+1}} dz \end{aligned} \tag{3.102}$$

where C_2 is a contour, clockwise, about $z = 1$, both also form solutions to Eq. (3.89). $h_l^{(1)}(x)$ and $h_l^{(2)}(x)$ are known as the spherical Hankel functions of the first and second kind respectively.

Returning then to $k_l(qr)$, traditionally the second solution of the exterior ODE, Eq. (3.88a), is based on the spherical Hankel of the first kind, and is set at

$$\begin{aligned} k_l(qr) &= -i^l h_l^{(1)}(iqr) \\ &= \frac{e^{-qr}}{qr} \sum_{m=0}^{\infty} \frac{1}{m!(2qr)^m} \frac{(l+m)!}{(l-m)!} \end{aligned} \tag{3.103}$$

and so the general solution outside of the well is given by

$$f_l(qr) = b_1 i_l(qr) + b_2 k_l(qr) \tag{3.104}$$

Asymptotically

$$\begin{aligned} i_l(x) &\xrightarrow[x\to\infty]{} \frac{e^x}{2x} \\ k_l(x) &\xrightarrow[x\to\infty]{} \frac{e^{-x}}{x} \end{aligned} \tag{3.105}$$

and so from the normalisation constraint $\int_{r=0}^{\infty} f_l^* f_l r^2 dr = 1, f_l \to 0$ as $r \to \infty$, then b_1 must equal zero. For $r > a$, the wave function is then

$$f_l(qr) = b_2 k_l(qr) \tag{3.106}$$

The function $k_l(x)$ is also known as a spherical modified Bessel function, and for l = 0 to 3 its general form can be seen in Figure 3-9.

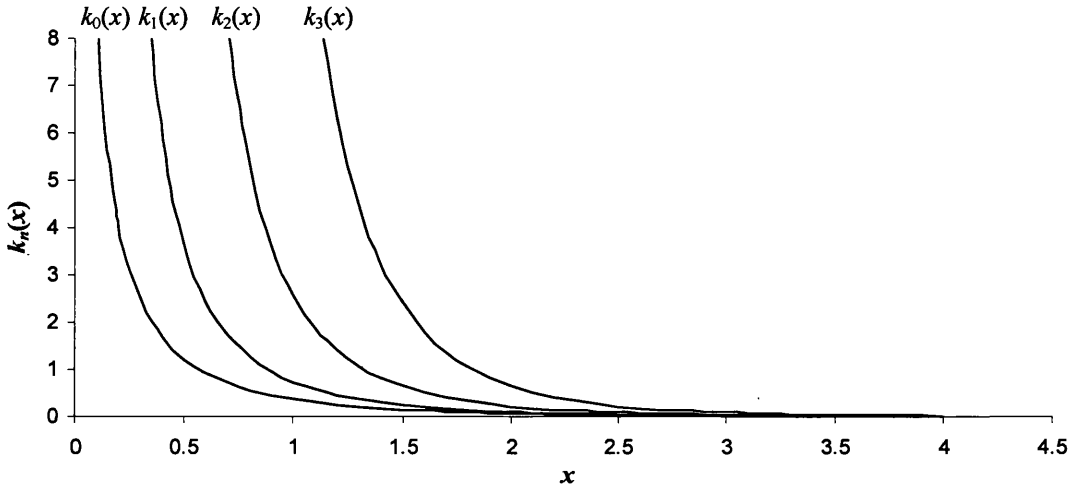


Figure 3-9 Plot of the spherical modified Bessel function, $k_l(x)$, for l = 0 to 3

Continuity at a is an important factor, and both f_l and its derivative must be continuous through this boundary. In mathematical terminology

$$0 = \lim_{\varepsilon \to 0} \left(\left. \frac{df_l}{dr} \right|_{r=a+\varepsilon} - \left. \frac{df_l}{dr} \right|_{r=a-\varepsilon} \right)$$
$$= \lim_{\varepsilon \to 0} \int_{a-\varepsilon}^{a+\varepsilon} \frac{d^2 f_l}{dr^2} dr \qquad (3.107)$$

which is true provided f_l'' is finite. For a physically realistic system, where f_l is required to satisfy Schrödinger's equation, f_l'' is never infinite. If f_l' is continuous, then f_l is also.

Cancelling the constants a_1 and b_2, then these two continuity requirements imply

$$\left. \frac{\frac{d}{dr} j_l(kr)}{j_l(kr)} \right|_{r=a} = \left. \frac{\frac{d}{dr} k_l(qr)}{k_l(qr)} \right|_{r=a} \qquad (3.108)$$

the solutions of which determine the allowed energy eigenvalues. Eq. (3.108) is a complicated transcendental equation dependant on l, V_0 and E, and a useful and relatively simple method of solution is via the intersection of the two equations [37]:

$$y_1^l(\beta) = \frac{\beta j_l'(\beta)}{a j_l(\beta)}$$
$$y_2^l(\beta) = \frac{\sqrt{\alpha^2 - \beta^2}\, k_l'(\sqrt{\alpha^2 - \beta^2})}{a k_l(\sqrt{\alpha^2 - \beta^2})} \qquad (3.109)$$

where

$$\beta = a\sqrt{\frac{2m}{\hbar^2}E}$$
$$\alpha = a\sqrt{\frac{2m}{\hbar^2}V_0}$$

with the prime on the k_l indicating the first derivative with respect to qr, evaluated at $r = a$, and similarly for j_l': i.e.

$$j_l'(\beta) = \frac{d}{d(kr)} j_l(kr)\Big|_{r=a}$$

For $l = 0$, the two equations are plotted in Figure 3-10, and their intersections circled. The radius of the well has been set at 5nm and V_0 at 2eV.

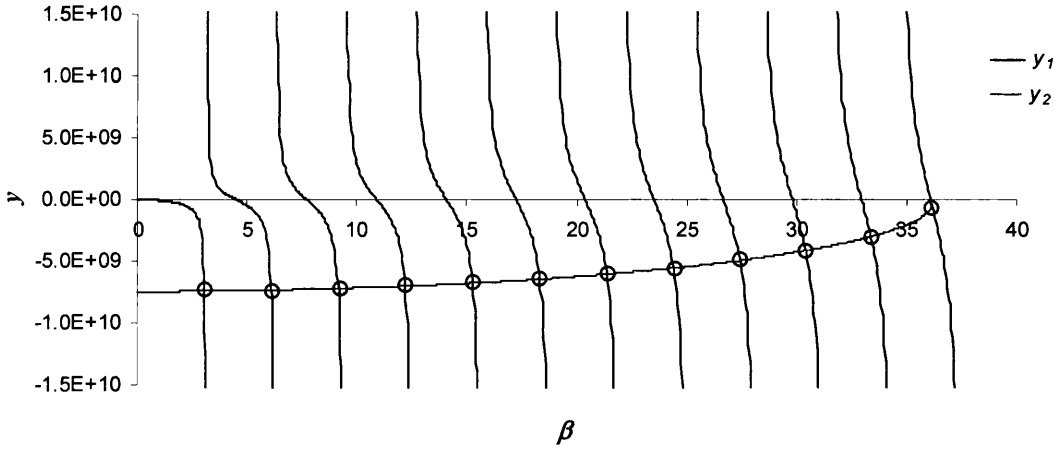


Figure 3-10 Plot of y_1 and y_2 for $l = 0$, determined for the 3D square well R =5nm, V_0 = 2eV. Points of intersection, circled, indicate the discrete energy eigenvalues of the potential well.

Maintaining the well radius and depth, the eigenvalues from the finite difference and analytical approaches are compared in Table 3-2 for $l = 0$ to 3, with favourable results.

For a 700-point mesh with a minimum inter-mesh spacing of 0.01nm, the discrete energy levels lie a mean ~0.23% above those determined analytically; an acceptable tolerance, validating the finite difference scheme.

Table 3-2 Table of energy eigenvalues for an R = 5nm, V_0 = 2eV square potential, spherically symmetric 3D well for l = 0 to 3. Eigenvalues, E^{FD}, are calculated with a finite difference method on a 700 point mesh and compared via $100\times(E^{FD} - E^{An})/E^{An}$ to the analytical energy eigenvalues, E^{An}, following from the roots of Eq. (2.108)

l = 0		1		2		3	
E^{FD} (eV)	*err* (%)	E^{FD} (eV)	*err* (%)	E^{FD} (eV)	*err* (%)	E^{FD} (eV)	*err* (%)
0.014	0.19	0.029	0.21	0.048	0.23	0.103	0.24
0.057	0.34	0.086	0.21	0.119	0.25	0.212	0.28
0.128	0.19	0.172	0.39	0.219	0.31	0.341	0.23
0.228	0.18	0.286	0.20	0.348	0.28	0.511	0.19
0.356	0.32	0.428	0.20	0.504	0.25	0.716	0.23
0.512	0.26	0.598	0.28	0.688	0.22	0.944	0.28
0.697	0.22	0.796	0.20	0.900	0.22	1.192	0.18
0.908	0.20	1.022	0.22	1.139	0.17	1.483	0.22
1.147	0.20	1.274	0.19	1.404	0.19	1.808	0.20
1.412	0.15	1.552	0.20	1.694	0.16		
1.701	0.14	1.850	0.24	1.995	0.22		
2.000	0.26						

The points of intersection of y^l_1 and y^l_2 of Eq. (3.109) were determined via a small intersection program, using a bisection method [50,51] (see also sub-section 2.3.2) to find the exact point of equality. For small l, $j_l(x)$ is best numerically computed by analytically expanding it in terms of sines and cosines through the recursion relation

$$j_l(x) = (-1)^l x^l \left(\frac{1}{x}\frac{d}{dx}\right)^l \frac{\sin x}{x} \tag{3.110}$$

However, for large l the recursion relationship

$$j_{n-1}(x) + j_{n+1}(x) = \frac{2n+1}{x} j_n(x) \tag{3.111}$$

can be useful. Unfortunately, for any extended range, the accumulation of rounding errors inherent with the use of any recursion relation can prove exceedingly problematic. To minimise their effects a technique similar to that employed in the FORTRAN subroutine BESJ, used to calculate Bessel functions of integer order, and discussed in Arfken and Weber [65] can be utilised. For given constant x, x_0, Eq. (3.111) is used working downwards from

$$j_{n+1}(x_0) = 0 \quad \text{and} \quad j_n(x_0) = \alpha$$

where $n \gg l$ and $n \gg x_0$, and finally normalising by comparing with results from the known form $j_0(x_0)$.

$k_l(x)$ can be approached in a similar manner. For small l, $k_l(x)$ is best considered through analytically determined Ae^{-B}/x^C factors, using for example

$$k_l(x) = (-1)^l x^l \left(\frac{1}{x}\frac{d}{dx}\right)^l \frac{e^{-x}}{x} \tag{3.112}$$

although for larger l, its decay as $r \to \infty$ implies that $k_l(x)$ is best treated working upward from known forms of $k_l(x)$ with the recurrence relation

$$k_{n-1}(x) - k_{n+1}(x) = -\frac{2n+1}{x} k_n(x) \tag{3.113}$$

3.4 Tight-Binding (TB)

An alternative method to the previous, essentially nearly free electron approach, where the band electrons are taken to be disassociated from their atomic sites and only weakly perturbed by the periodic ionic lattice, is the tight-binding (TB) approach. In the TB approximation, electrons are viewed as occupying the standard orbitals of their constituent atoms, then 'hopping' between atoms during conduction. This restricted Hilbert space, spanned by the atomic-like orbitals, is assumed to be sufficient to describe the wave function solutions of the Schrödinger equation. While the nearly free electron model is considered to be a good approximation for systems such as metals, the TB method can provide a better representation of systems where electrons are fairly localised, for instance in the chemical bonds of covalent semiconductors [29].

If the discretisation of the continuous Schrödinger equation for the nearly free electron approximation is an acceptable approach here, then its eigenstates should equal, to within a small tolerance, the tight binding model, formed for the same system.

Initially, begin in one dimension. Let $|x_i\rangle$ [i] be a state centred at site *i*, see Figure 3-11,

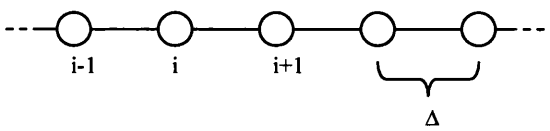


Figure 3-11 1-D chain of discrete sites

which obeys the relationship for orthonormality

$$\langle x_i | x_j \rangle = \delta_{ij} \tag{3.114}$$

and therefore the completeness, or closure, relation

$$\sum_i |x_i\rangle\langle x_i| = 1 \tag{3.115}$$

Then an arbitrary state $|\psi\rangle$ can be expanded in this basis[ii] as

---

[i] Introducing the Dirac notation, with the ket state vector $|m\rangle$ and the bra state vector $\langle n|$ such that

$$\langle n|\hat{a}|m\rangle = \int \psi_n^* \hat{a} \psi_m d^3r$$

[ii] In an *n*-dimensional vector space, an arbitrary vector **v** is described by its components $(v_1, v_2, \ldots v_n)$ via

$$|\psi\rangle = \sum_i \psi_i |x_i\rangle \tag{3.116}$$

where $|\psi_i|^2$ is the probability of finding an electron at site i.

For the 1-D chain in Figure 3-11, a generic tight-binding Hamiltonian can be expressed as

$$H = \sum_i |x_i\rangle \varepsilon_i \langle x_i| + \sum_{ij} |x_i\rangle u_{ij} \langle x_j| \tag{3.117}$$

where ε_i is the on-site energy and u_{ij} the hopping energy between sites i and j. Considering nearest neighbour interactions only on a uniform mesh, then $j = i \pm 1$, $u_{i,j+1} = u_{i,j-1} = u$ and so the Hamiltonian becomes

$$H = \sum_i |x_i\rangle (\varepsilon_0 + V_i) \langle x_i| + u \sum_i \left(|x_i\rangle \langle x_{i-1}| + |x_i\rangle \langle x_{i+1}| \right) \tag{3.118}$$

ε_i now separated into $\varepsilon_0 + V_i$, where V_i is an external potential and ε_0 is a residual energy, common to all the discrete spatial points within the space.

Applying the Hamiltonian to the state $|\psi\rangle = \sum_n \psi_n |x_n\rangle$ and recalling that $H|\psi\rangle = E|\psi\rangle$ then

$$H \sum_n \psi_n |x_n\rangle = \sum_n \psi_n |x_n\rangle (\varepsilon_0 + V_n) + u \sum_n \left(\psi_n |x_{n+1}\rangle + \psi_n |x_{n-1}\rangle \right) = E \sum_n \psi_n |x_n\rangle$$

Pre-multiply by $\langle x_k|$ and

$$\psi_k (\varepsilon_0 + V_k) + u(\psi_{k-1} + \psi_{k+1}) = E\psi_k \tag{3.119}$$

follows.

Consistent with preceding work, discretising the continuous Schrödinger equation on a uniform mesh gives

$$\left(\frac{\hbar^2}{m\Delta^2} + V_k \right) \psi_k - \frac{\hbar^2}{2m\Delta^2} (\psi_{k-1} + \psi_{k+1}) = E\psi_k \tag{3.120}$$

For Eq. (3.120)'s eigenstates to be acceptable, then Eq. (3.120) must be equivalent to Eq. (3.119), implying that

---

$$\mathbf{v} = \sum_{i=1}^{n} v_i \mathbf{e}_i$$

where $\mathbf{e}_i$ are a set of n linearly independent vectors which span the vector space. The $\mathbf{e}_i$ are said to form a basis and constitute a complete set (a set is complete if it is not contained within any larger set).

$$u = \frac{-\hbar^2}{2m\Delta^2}$$
$$\varepsilon_0 = \frac{\hbar^2}{m\Delta^2} \tag{3.121}$$

Taking the confining potential V to be that of a finite square well, $V = 0$ within the well walls, simple analytic solutions to the Schrödinger equation exist. From the continuous plane wave eigenfunction e^{ikx}, where

$$k = \sqrt{2mE/\hbar^2}$$

and E is an eigenvalue of the continuous Schrödinger equation, the solution of the discrete case is proposed as e^{ikx_i}, and so via Eq.(3.119):

$$\begin{aligned} Ee^{ikx_i} &= \varepsilon_0 e^{ikx_i} + u(e^{ikx_{i-1}} + e^{ikx_{i+1}}) \\ &= \varepsilon_0 e^{ikx_i} + ue^{ikx_i}(e^{-ik\Delta} + e^{ik\Delta}) \end{aligned}$$

i.e.

$$\begin{aligned} E &= \varepsilon_0 + u(e^{-ik\Delta} + e^{ik\Delta}) \\ &= \varepsilon_0 + 2u\cos(k\Delta) \end{aligned}$$

ergo

$$E = \varepsilon_0(1 - \cos(k\Delta)) \tag{3.122}$$

Eq. (3.122) is referred to as the discrete energy dispersion relation.

Three dimensionally the situation is slightly more complex. Retaining our spherically symmetric space, see Figure 3-12

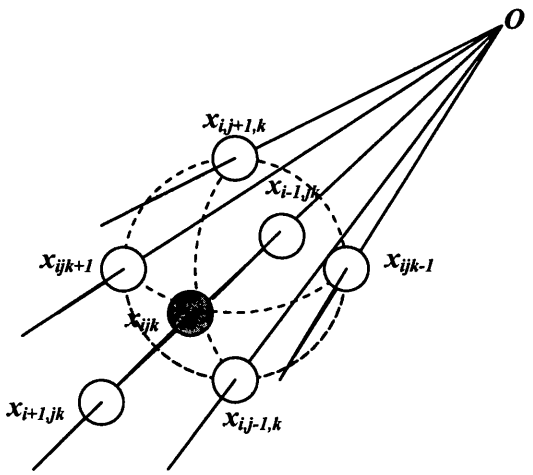


Figure 3-12 3-D spherical spatial grid of nearest neighbour discrete sites

then from the TB Hamiltonian

$$H = \sum_{ijk} |x_{ijk}\rangle \varepsilon_{ijk} \langle x_{ijk}| + \sum_{ijk} \sum_{i'j'k'} |x_{ijk}\rangle u_{ijki'j'k'} \langle x_{i'j'k'}| \tag{3.123}$$

taking nearest neighbour interactions only and once again separating ε_i into $\varepsilon_0 + V_{ijk}$, H becomes

$$H = \sum_{ijk} |x_{ijk}\rangle (\varepsilon_0 + V_{ijk}) \langle x_{ijk}| + \sum_{ijk} \left(u_{ijk,i-1} |x_{ijk}\rangle \langle x_{i-1,jk}| + u_{ijk,i+1} |x_{ijk}\rangle \langle x_{i+1,jk}| + u_{ijk,j-1} |x_{ijk}\rangle \langle x_{i,j-1,k}| \right.$$
$$\left. + u_{ijk,j+1} |x_{ijk}\rangle \langle x_{i,j+1,k}| + u_{ijk,k-1} |x_{ijk}\rangle \langle x_{ijk-1}| + u_{ijk,k+1} |x_{ijk}\rangle \langle x_{ijk+1}| \right)$$

The external potential is spherically symmetric and so can be reduced to one variable, V_i. Applying the Hamiltonian to the state $|\psi_{lm}\rangle = \sum_{i'j'k'} \psi^{lm}_{i'j'k'} |x_{i'j'k'}\rangle$, pre-multiplying by $\langle x_{abc}|$ and finally with the separation of variables $\psi^{lm}_{abc} = R^l_a \Phi^{lm}_{bc}$ then

$$\varepsilon_0 + V_a + u_{a-1} \frac{R^l_{a-1}}{R^l_a} + u_{a+1} \frac{R^l_{a+1}}{R^l_a} + A^{lm}_{abc} = E \tag{3.124}$$

where

$$A^{lm}_{abc} = \frac{1}{\Phi^{lm}_{bc}} \left(u_{b-1} \Phi^{lm}_{b-1c} + u_{b+1} \Phi^{lm}_{b+1c} + u_{c-1} \Phi^{lm}_{bc-1} + u_{c+1} \Phi^{lm}_{bc+1} \right)$$

with the simplifications $u_{a-1} = u_{abc,a-1}$ etc. The hopping integrals $u_{b\pm1}$, $u_{c\pm1}$ etc. however, still retain an intrinsic dependence on a with their b and c coordinates in accordance with the spherical discretisation of the space. As the RHS of Eq. (3.124) lacks any dependence on a, b or c, the LHS has to be constant regarding a, b and c, implying that A^{lm}_{abc} must be constant in b and c.

Comparing Eq. (3.124) with the discretisation of the continuous SE in spherical coordinates, analogous to preceding works

$$\left(\frac{\hbar^2}{m\Delta^2} + V_a + \frac{\hbar^2}{2m} \frac{l(l+1)}{x_a^2} \right) R^l_a - \frac{\hbar^2}{2m\Delta^2} \left(R^l_{a+1} + R^l_{a-1} \right) - \frac{\hbar^2}{2m\Delta} \left(\frac{R^l_{a+1} - R^l_{a-1}}{x_a} \right) = E R^l_a \tag{3.125}$$

then for equality

$$u_{a+1} = -\frac{\hbar^2}{2m\Delta^2} \left(1 + \frac{\Delta}{x_a} \right)$$
$$u_{a-1} = -\frac{\hbar^2}{2m\Delta^2} \left(1 - \frac{\Delta}{x_a} \right) \tag{3.126a}$$

$$\varepsilon_0 = \frac{\hbar^2}{m\Delta^2}$$
$$A_{abc}^{lm} = A_a^{lm} = \frac{\hbar^2}{2m}\frac{l(l+1)}{x_a^2} \tag{3.126b}$$

Taking once again the confining potential V to be that the finite square well of Figure 3-5, it is implied that within the well, the family of functions R^l are spherical Bessel functions if the tight-binding and nearly free models are equivalent. Proceeding as with the 1D case, Eq. (3.124) becomes

$$E = \frac{\varepsilon_0}{2}\left(2 - \left(1+\frac{\Delta}{r_a}\right)\frac{j_l(k(r_a+\Delta))}{j_l(kr_a)} - \left(1-\frac{\Delta}{r_a}\right)\frac{j_l(k(r_a-\Delta))}{j_l(kr_a)} + \frac{\Delta^2}{r_a^2}l(l+1)\right)$$

From the series expansion of $j_l(x)$ and the binomial theorem then

$$\begin{aligned}
&\frac{k(r_a+\Delta)j_l(k(r_a+\Delta)) + k(r_a-\Delta)j_l(k(r_a-\Delta))}{kr_a} \\
&\quad = 2^l \sum_{m=0}^{\infty} \frac{(-1)^m (l+m)!}{m!(2m+2l+1)!} \sum_{k=0}^{2m+l+1} \frac{(2m+l+1)!}{k!(2m+l+1-k)!} (kr_a)^{2m+l-k}\left((k\Delta)^k + (-k\Delta)^k\right) \\
&\quad = 2^{l+1} \sum_{m=0}^{\infty} \frac{(-1)^m (l+m)!}{m!(2m+2l+1)!} (kr_a)^{2m+l} \sum_{k=0}^{\infty} \frac{(-1)^k (k\Delta)^{2k}}{(2k)!} \frac{m!(l+m+k)!}{(m+k)!(l+m)!} \\
&\quad = 2^{l+1} \sum_{m=0}^{\infty} \frac{(-1)^m (l+m)!}{m!(2m+2l+1)!} (kr_a)^{2m+l}\left(1 + \sum_{k=1}^{\infty} \frac{(-1)^k (k\Delta)^{2k}}{(2k)!} \prod_{s=1}^{k}\left(1+\frac{l}{m+s}\right)\right)
\end{aligned}$$

and so Eq. (3.124) can be written

$$E = \varepsilon_0\left(1 - \frac{2^l \sum_{m=0}^{\infty} \frac{(-1)^m (l+m)!}{m!(2m+2l+1)!} (kr_a)^{2m+l}\left(1 + \sum_{k=1}^{\infty} \frac{(-1)^k (k\Delta)^{2k}}{(2k)!} \prod_{s=1}^{k}\left(1+\frac{l}{m+s}\right)\right)}{2^l \sum_{m=0}^{\infty} \frac{(-1)^m (l+m)!}{m!(2m+2l+1)!} (kr_a)^{2m+l}} + \frac{\Delta^2}{2r_a^2}l(l+1)\right) \tag{3.127}$$

When $l = 0$, the third term on the RHS vanishes and the second term reduces to the series expansion of $\cos(k\Delta)$, thus

$$E = \varepsilon_0\left(1 - \cos(k\Delta)\right)$$

identical to the 1D energy dispersion relation, Eq. (3.122). However, for $l \neq 0$ the k summation in the second term retains its dependence on m and so the RHS remains a function of $(kr_a)^n$. This implies that in general the tight-binding Hamiltonian is not satisfied exactly by the nearly free eigenstates. This said, an exact TB energy eigenvalue must be independent of r_a; thus in the limit $r_a \to \infty$, outside of the potential well and again assuming accordance with the nearly free electron

model, the family of functions R^l are spherical modified Bessel functions, asymptotically behaving as Eq. (3.105), namely

$$k_l(qr_a) \xrightarrow[qr_a \to \infty]{} (-1)^{l+1} \frac{e^{-qr_a}}{qr_a}$$

then in this limit Eq. (3.124) can be written as

$$E = \varepsilon_0 \left(1 - \frac{1}{2} \frac{q(r_a + \Delta)}{qr_a} \frac{k_l(q(r_a + \Delta))}{k_l(q\Delta)} - \frac{1}{2} \frac{q(r_a - \Delta)}{qr_a} \frac{k_l(q(r_a - \Delta))}{k_l(q\Delta)} \right) + V_0$$
$$= \varepsilon_0 \left(1 - \frac{e^{-q\Delta} + e^{q\Delta}}{2} \right) + V_0$$

and thus

$$E = \varepsilon_0 (1 - \cosh(q\Delta)) + V_0 \tag{3.128}$$

where q is

$$q = \sqrt{2m(V_0 - E)/\hbar^2}$$

Then to second order in Δ, $O(\Delta^2)$, both Eq. (3.122) and Eq. (3.128) are equal, i.e.

$$E = \frac{\hbar^2 k^2}{2m} + O(\Delta^2) \tag{3.129}$$

It is reasonable to state then, wave functions aside, should the energy eigenvalues of the nearly free electron model and the tight-binding model be exactly equivalent, then the LHS E's of Eq. (3.122) and Eq. (3.128) will be the same as the eigenvalues used in k and q on the RHSs. Using the nearly free eigenvalues to determine k and q, then any deviance between them and their resultant tight-binding Es is a measure of the difference between the two treatments.

For the square well potential of Figure 3-5, with V_0 set at 2eV and a well radius of 5nm, Table 3-3 compares the Eq. (3.122) (l = 0) and Eq. (3.128) (l > 0) LHS energies, labelled E^{TB}, with the analytic solutions, labelled E^{An}, for l = 0 to 3.

The percentage error increases with energy for Eq. (3.122) but decreases with increasing energy for Eq. (3.128), and lies in the range 3.21×10^{-7} % to 2.91 %, with its mean at 0.22 %. Despite the substantial percentage differences between the two approximations for the first energy eigenvalues of l = 1 and 2, overall the error is taken to be negligible for the energy spectrum of this size of nanocrystal. Indeed, should m be reduced to the effective electron mass of tin dioxide,

Table 3-3 Table of energy eigenvalues for an R = 5nm, V_0 = 2eV square potential, spherically symmetric 3D well for l = 0 to 3. Eigenvalues, E^{TB}, are calculated from the radial energy dispersion relations, Eq. (3.122) and Eq. (3.128) using the analytical energy eigenvalues, E^{An}, to determine k. E^{TB} is then compared to E^{An} via $100\times (E^{TB} - E^{An})/E^{An}$.

l = 0		1		2		3	
E^{TB} (eV)	*err* (%)	E^{TB} (eV)	*err* (%)	E^{TB} (eV)	*err* (%)	E^{TB} (eV)	*err* (%)
0.014	0.0003	0.028	2.9106	0.047	1.7513	0.102	0.7623
0.057	0.0012	0.085	0.9287	0.118	0.6486	0.211	0.3292
0.128	0.0028	0.170	0.4260	0.218	0.3165	0.339	0.1767
0.228	0.0050	0.285	0.2250	0.346	0.1719	0.510	0.0949
0.355	0.0078	0.427	0.1264	0.502	0.0972	0.714	0.0504
0.511	0.0112	0.596	0.0719	0.686	0.0547	0.941	0.0259
0.695	0.0152	0.794	0.0398	0.898	0.0294	1.190	0.0120
0.907	0.0199	1.019	0.0205	1.137	0.0142	1.480	0.0039
1.145	0.0251	1.272	0.0090	1.402	0.0055	1.804	0.0004
1.410	0.0309	1.549	0.0028	1.691	0.0012		
1.699	0.0372	1.846	0.0003	1.990	0.0000		
1.995	0.0437						

$0.275m$, then for l = 0 to 3, the percentage error between E^{TB} and E^{An}, is lowered to a mean ~0.06% with a maximum error of ~0.8 %. From Eq. (3.129) it is clear that as $\Delta \rightarrow 0$, E^{TB} tends to E^{An}. For non-zero Δ, replacing the analytic eigenvalues with finite difference ones makes no appreciable difference to the results of Table 3-3.

3.5 Self-Consistency

Accepting that the finite difference discretisation of the Schrödinger equation for the nearly free electron approximation gives suitable eigenstates for the systems under discussion, then the next issue is one of self-consistency between the solutions of Poisson's and Schrodinger's equations.

For a self-consistent solution, the electron density resulting from a potential must generate that same potential - that is $\rho_{out}(r) = \rho_{in}(r)$ or $V_{out}(r) = V_{in}(r)$. In the context of this work, this means that the the electron density formed from the eigenstates of Schrödinger's equation for a potential generated from Poisson's equation must produce that selfsame potential. This will generally not be the case for a spatially restricted system, with a small finite number of discrete energy levels and limited electron occupancy, where the initial potential was generated via a charge density estimated from bulk semiconductor properties.

To achieve this consistency, starting from a trial potential V^0, the obvious approach would be to then to determine the potential V^1 from V^0's charge density via

$$\begin{aligned} n_e(r,\theta,\varphi) &= \sum_{\sigma}\sum_{nlm} f(E_{nl} - \overline{V}^{0\,to\,R})\left|\psi_{nlm}(r,\theta,\varphi)\right|^2 \\ &= \sum_{\sigma}\sum_{nl} \frac{2l+1}{4\pi} f(E_{nl} - \overline{V}^{0\,to\,R})\left|f_{nl}(r)\right|^2 \\ &= n_e(r) \end{aligned} \tag{3.130}$$

and iterate until $V^{n+1} = V^n$. $f(E)$ is the thermally dependant Fermi fractional occupation factor where E is equal to the difference between the energy eigenvalue under consideration and the mean value of the potential within the well.

As the potential alters for each iteration naturally so too will its gradient at the boundary of the quantum well and thus the charge contained within the occupied surface states must also vary. As such, the interface dipole between this surface charge and the image charge in, say the STM tip (when applied in the charge writing scenario), and so the corresponding voltage drop, V_{int} in Figure 2-5, will also undergo variations. To maintain an electron affinity consistent with the measured S_b this implies that the the depth of the confining potential, see Eq. (2.49), and therefore the number of discrete energy levels, must also change. Naturally, this can have a profound effect on the charge density, particularly for the higher electron densities.

Under these circumstances, the simplistic iterative approach discussed above will generally fail to converge, the charge redistributions occurring between each iterative step driving the generating potentials into oscillation, each new potential being the extreme opposite to the old. This often leads to deep, non-convergent parabolic V's.

To overcome this, the charge displacement can be damped via a mixing of the input and output charge densities. As discussed in Kohanoff [29], the most basic strategy follows

$$\rho_{in}^{n+1}(r) = \alpha\rho_{out}^{n}(r) + (1-\alpha)\rho_{in}^{n}(r) \tag{3.131}$$

where α is an adjustable parameter, set to minimise the number of iterations. The greater α, the less is retained between cycles but the faster the rate of convergence, in theory at least. However, for some case, such as metallic and magnetic systems, α must lie at ~0.01 to avoid the divergence of the iterative procedure and a great number of iterations are necessary before self-consistency. More sophisticated approaches are reviewed in Kohnanoff [29], where generally the input and output charge densities of other proceeding iterations are included, although for the system under study here, the scheme developed below, dependant on just one previous iteration, is sufficient.

For a successful mixing strategy, it is assumed that as n approaches N, where N is the point at which the densities are taken as self-consistent, ρ^n becomes an increasingly better approximation of ρ^N. Should this be the case, then less damping is require as $n \rightarrow N$ since the charge redistribution between each iterative cycle is diminishing. Implied by this then, is that should the mixing parameter α of Eq. (3.131) increase as ρ^n tends to ρ^N then convergence can still be assured whilst avoiding the computationally prohibitive quantity of cycles of Eq. (3.131). Therefore, the charge density mixing procedure favoured for this work is

$$\begin{aligned} \rho_{in}^{n+1}(r) &= \alpha^n\rho_{out}^{n}(r) + (1-\alpha^n)\rho_{in}^{n}(r) \\ \alpha^n &= \frac{1}{1+\left|\frac{\rho_{in}^{n}-\rho_{out}^{n}}{\rho_{out}^{n}}\right|^{S}} \end{aligned} \tag{3.132}$$

where S is an adjustable parameter, set to ensure convergence. For tin dioxide, generally $S \approx 0.5 - 10$ is adequate, even when including electron-electron interactions. As $n \rightarrow N$, α^n tends to 1, and the input charge density equals the output charge density. Practically of course, equality is only to within a set numerical tolerance

To illustrate Eq. (3.132) on a discrete spatial mesh, a good measure of the self-consistency of the system is

$$\zeta^n = \sum_i \left(\frac{V_{out}^n(r_i) - V_{in}^n(r_i)}{V_{out}^n(r_i)} \right)^2 \tag{3.133}$$

applied at the end of each cycle. Figure 3-13 illustrates this self-consistency procedure for the parameters of a 15nm radius spherical tin dioxide nanoparticle on a 200 pt mesh. Beginning with a trial potential generated from Poisson's equation for the bulk electron density, the eigenstates are determined with the finite difference method, and iterated for ten cycles, each new potential formed from Poisson's equation using the wave function dependant electron density Eq. (3.130).

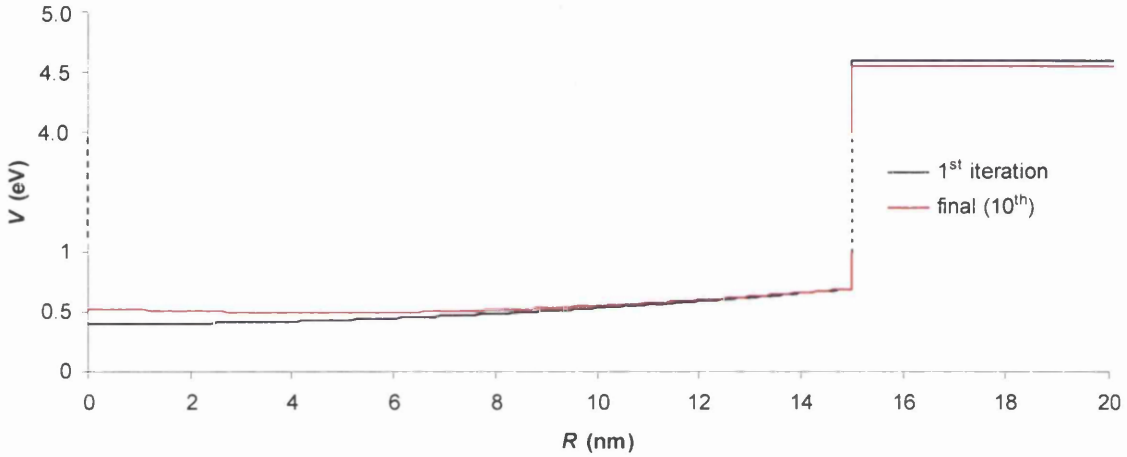


Figure 3-13 Plot of initial and final potentials for a 10 iteration self-consistency cycle applied to a 15nm SnO_2 grain at room temperature. Trial (1^{st}) potential generated from Poisson's equation using the bulk semiconductor electron density.

For the purposes of these illustrative examples, Figure's 3-13 to 3-15, the potential outside the well is set to the value of the tin dioxide work function, calculated to satisfy Eq. (2.49). Table 3-4 gives ζ for a selection of iterations and indicates the high degree of self-consistency than can be achieved after only a small number of cycles. $\alpha^n \approx 1$ by the final few iterations as required.

Table 3-4 Table of ζ for various iterations within a self-consistency cycle applied to a 15nm SnO_2 grain at room temperature. Trial potential generated from Poisson's equation using the bulk semiconductor electron density.

iteration	ζ
1^{st}	0.92
2^{nd}	0.38
3^{rd}	9.82×10^{-3}
4^{th}	6.94×10^{-4}
final (10^{th})	2.93×10^{-10}

Such is the power of Eq. (3.132) in coupling the solutions of Poisson's and Schrödinger's equations that even beginning from a square potential, discounting entirely the bulk semiconductor form of the electron density, within 10 iterations the same degree of self-consistency as Figure 3-13 and Table 3-4 can be achieved, see Figure 3-14 and Table 3-5.

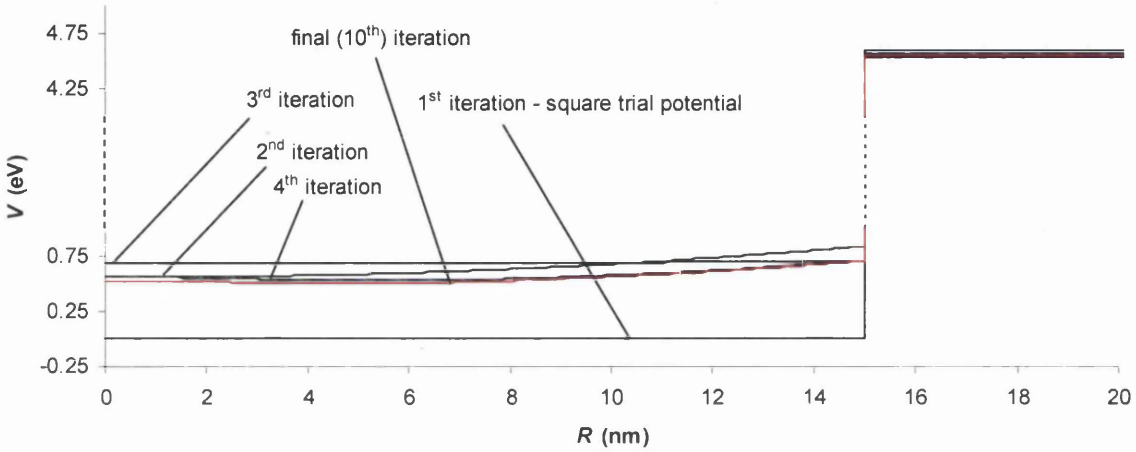


Figure 3-14 Plot of initial and final potentials for a 10 iteration self-consistency cycle applied to a 15nm SnO_2 grain at room temperature. The trial (1^{st}) potential is a 4.6eV deep square well.

Table 3-5 Table of ζ for various iterations within a self-consistency cycle applied to a 15nm SnO_2 grain at room temperature. The trial potential is a 4.6eV deep square well.

iteration	ζ
1^{st}	137.01
2^{nd}	2.29
3^{rd}	7.07
4^{th}	0.35
final (10^{th})	4.72×10^{-10}

Here, ζ^1 is two orders of magnitude greater than the ζ^1 of Table 3-4, though ζ rapidly decreases until V^{10} is equal to the V^{10} of Figure 3-13. Again $\alpha^n \rightarrow 1$ as n approaches 10.

Returning to Figure 3-13, the actual difference between the self-consistent potential and that of the bulk electron density may appear to be too small to warrant the extra computational effort, however, as the nanocrystals diminish in size the spacing between their discrete energy levels increases, and their electronic behaviour correspondingly diverges from that of their bulk counterparts'. On charging therefore, the exact forms of the individual orbitals containing the

added electrons become even more important, as the spatial localisation of the charge, and thus the resultant potential, is markedly different from the scenario where a continuum of states 'smears' the charge over the grain.

In way of example, should the effects of charging be simulated via the movement of the Fermi level up the potential well, although unlike an actual charged grain the overall charge is still taken as zero i.e. $Q = 0$ in Eq. (5.42). Then for the conduction band bottom of a 4nm radius SnO_2 grain as plotted in Figure 3-15(a), at E_f = 1.25eV the occupied surface state density switches from -6.52×10^{15} m^{-2} to 8.26×10^{16} m^{-2} between the bulk and self-consistent charge densities as the potential profile within the well changes from concave to convex. The maximum difference

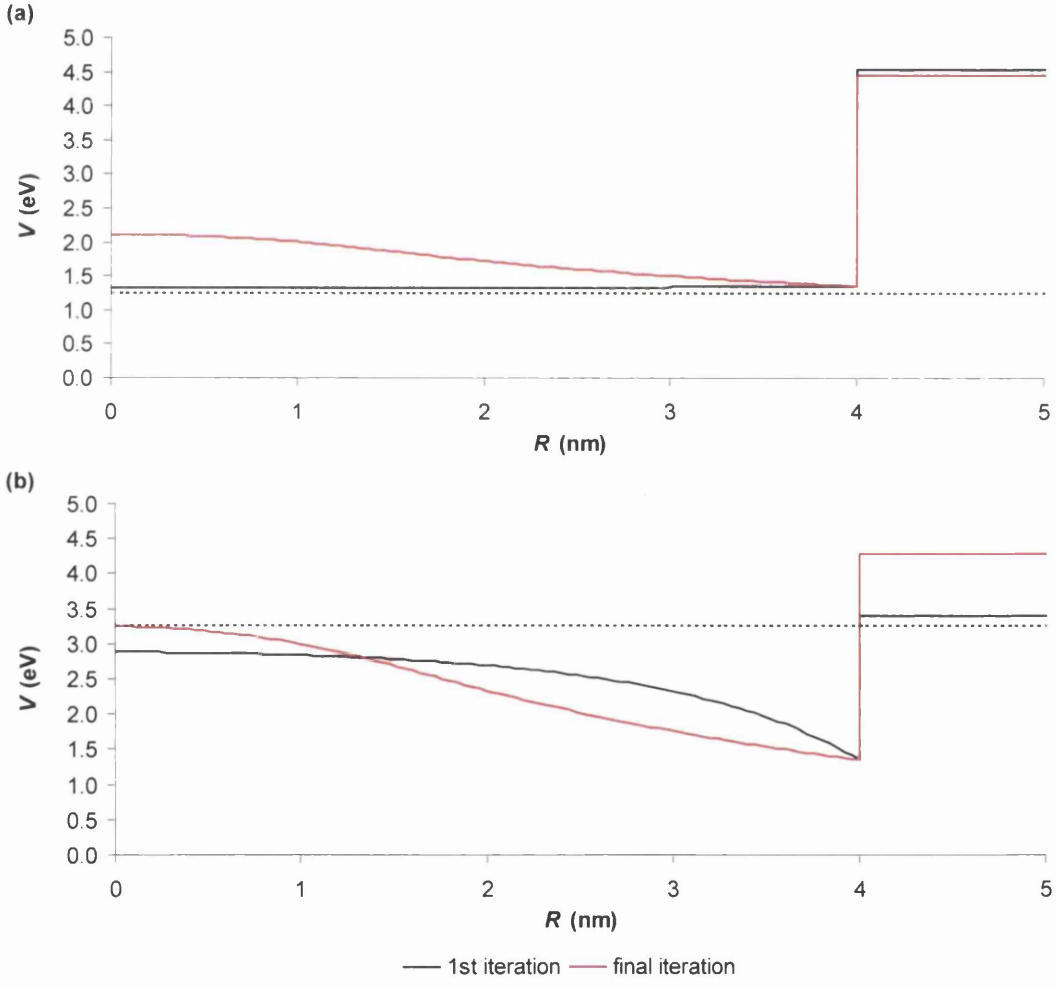


Figure 3-15 Plots of initial and final potentials for 25 iteration self-consistency cycles, applied to a 4nm SnO_2 grain at room temperature at Fermi levels of (a) 1.25eV and (b) 3.25eV, indicated on their respective graphs via a dotted line. Trial (1st) potentials are generated from Poisson's equation using the bulk semiconductor electron density.

between the two band bottoms, 0.77eV, lies at the well centre. This is in contrast to the E_f = 3.25eV case, Figure 3-15(b), where within the well, the maximum difference of 0.57eV is found at ~2.92nm, but with the trial potential now also convex, and additionally, outside the well the vacuum potential differs by 0.87eV as the density of charge contained within the surface states drops from $1.00\times10^{18}m^{-2}$ to $2.25\times10^{17}m^{-2}$. Continuing the self-consistency cycle over 25 iterations (although practically it could be halted far short of this), then for the two Fermi levels, ζ's of 1.83×10^{-18} and 1.19×10^{-26} can be achieved respectively.

Clearly then, self-consistency is of the utmost importance to the goals of this project, and the negelect of this issue would be in serious danger of rendering any charge writing simulation meaningless.

3.6 Conclusions

In this chapter, it has been shown that by solving Schrödinger's time-independent equation, the eigenstates of a finite potential can be determined, giving the discrete energy levels and orbitals of a quantum dot.

The mathematics and quantum mechanics were outlined that allow the electron wave function to be separated into its angular and radial components if the confining potential is solely dependant on the radial coordinate. It was demonstrated that this then allows Schrödinger's equation to be split into two separate equations, one angular one radial.

The angular equation was then solved exactly, its eigenvectors referred to as the spherical harmonics and whose dependence on two quantum numbers, m and l, was shown. Its eigenvalues are demonstrated to be $l(l+1)$ where l forms the allowed limit on the azimuthal quantum number m, $m \leq |l|$, and can take the values $l = 0, 1, 2, \ldots$. These eigenvalues are not only integral to the corresponding spherical harmonics, but are also required within the radial Schrödinger equation.

The radial component of the wave function cannot be determined exactly, dependant as it is on the form of V. To tackle this, a finite difference method is developed, discretising the radial equation on a uniform grid of spatial points. The whole radial operator is then expressed as a tri-diagonal matrix, reducing the ODE to a simpler algebraic problem of ascertaining the eigenstates of an $Ax = \lambda x$ type equation, solved using routines from the linear algebra package LAPACK (although simple QR-factorisation and inverse iteration techniques to solve this type of problem are discussed in the accompanying appendices). This was then coupled with a method of orthogonalising the eigenvectors to generate the final eigenstates of the radial equation. To assess the accuracy and suitability of this approach, firstly, the finite difference method was applied to a square potential where the eigenstates can be calculated analytically fairly simply, and the results of the two methods compared. For a 700-point mesh with an inter-mesh spacing of 0.01nm, the finite difference discrete energy levels lie a mean ~0.23% above those determined analytically, see Table 3-2, a more than acceptable discrepancy. Secondly, these nearly-free finite difference results are compared with those from a tight-binding Hamiltonian, where the electrons within the semiconductor are no longer taken to be essentially free, only being weakly perturbed by the periodic ionic lattice, but rather are viewed as occupying the standard orbitals of their constituent

atoms, then 'hopping' between atoms during conduction. Again using the finite square well potential as the reference potential, the results of the two approaches only differ by a mean ~0.22%, see Table 3-3, reducing to ~0.06% on using the effective mass of SnO_2.

Having shown that the radial component of the wave function can at least in principle be determined within an acceptable error tolerance, the next subject to be treated was that of self-consistency. For a self-consistent solution, the electron density resulting from a potential must generate that same potential - that is $\rho_{out}(r) = \rho_{in}(r)$ or $V_{out}(r) = V_{in}(r)$. To achieve this far from straightforward goal, a mixing scheme was developed to create an amalgamate charge density at each iteration from the charge densities of the current and preceding iterations. By doing so, the charge fluctuations between each iteration are damped and the system can converge to a self-consistent potential. The scheme is shown to be successful for a 15nm SnO_2 grain at room temperature even when the trial potential is square, see Figure 3-14, and the necessity of self-consistency is clearly demonstrated for higher electron densities in Figure 3-15.

So far however, to simplify the discussions a little, the mesh upon which the radial equation is discretised has been uniform. This is not ideal in treating scenarios where the decay of the wave function outside of the grain is slow, as the potentially vast external space, which is of little interest, has to be considered with the same resolution as the internal region, which is of great interest, but is possibly very small in extent. To overcome this, a variable mesh and the consequent new discretisation of the radial Schrödinger equation must be developed. This, incidentally, is essential for the logarithmic meshes used in Chapter 4 to model atoms in testing approximations of the exchange potential.

For a finite spatial mesh, as in Figure 3-16

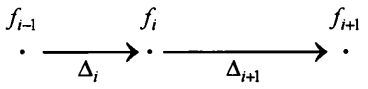


Figure 3-16 Schematic of variable mesh

a function f can be expanded via a Taylor series about site i such that

$$f_{i+1} = f_i + \Delta_{i+1} f_i' + \frac{\Delta_{i+1}^2}{2!} f_i'' + \frac{\Delta_{i+1}^3}{3!} f_i''' + \frac{\Delta_{i+1}^4}{4!} f_i^{iv} + \cdots$$
$$f_{i-1} = f_i - \Delta_i f_i' + \frac{\Delta_i^2}{2!} f_i'' - \frac{\Delta_i^3}{3!} f_i''' + \frac{\Delta_i^4}{4!} f_i^{iv} + \cdots \qquad (3.134)$$

(compare with Eq. (3.65) and Eq. (3.66)). Assuming that although $\Delta_i \neq \Delta_{i+1}$, the mesh spacings are of the same order of magnitude i.e. $O(\Delta_i) \approx O(\Delta_{i+1}) = O(\Delta)$, then

$$f_{i+1} + f_{i-1} = 2f_i + (\Delta_{i+1} - \Delta_i)f_i' + (\Delta_{i+1}^2 + \Delta_i^2)\frac{f_i''}{2!} + O(\Delta^3)$$
$$f_{i+1} - f_{i-1} = (\Delta_{i+1} + \Delta_i)f_i' + (\Delta_{i+1}^2 - \Delta_i^2)\frac{f_i''}{2!} + O(\Delta^3)$$

then the first and second derivatives of f can be written

$$\begin{aligned} f_i'' &= \frac{2}{\Delta_{i+1}^2 + \Delta_i^2}\left[f_{i+1} + f_{i-1} - 2f_i - (\Delta_{i+1} - \Delta_i)f_i'\right] + O(\Delta) \\ f' &= \frac{f_{i+1} - f_{i-1}}{\Delta_{i+1} + \Delta_i} - \frac{\Delta_{i+1}^2 - \Delta_i^2}{\Delta_{i+1} + \Delta_i}\frac{f_i''}{2!} + O(\Delta^2) \end{aligned} \tag{3.135}$$

therefore

$$f_i'' = \left[\frac{1}{\Delta_{i+1}\Delta_i}\right]\left[\left(1 - \frac{\Delta_{i+1} - \Delta_i}{\Delta_{i+1} + \Delta_i}\right)f_{i+1} + \left(1 + \frac{\Delta_{i+1} - \Delta_i}{\Delta_{i+1} + \Delta_i}\right)f_{i-1} - 2f_i\right] + O(\Delta) \tag{3.136}$$

and

$$\begin{aligned} f_i'' + \frac{2}{r_i}f_i' = &\left[1 - \frac{\Delta_{i+1}^2 - \Delta_i^2}{r_i(\Delta_{i+1} + \Delta_i)}\right]\left[\frac{1}{\Delta_{i+1}\Delta_i}\right] \times \\ &\left[\left(1 - \frac{\Delta_{i+1} - \Delta_i}{\Delta_{i+1} + \Delta_i}\right)f_{i+1} + \left(1 + \frac{\Delta_{i+1} - \Delta_i}{\Delta_{i+1} + \Delta_i}\right)f_{i-1} - 2f_i\right] + \frac{2}{r_i}\frac{f_{i+1} - f_{i-1}}{\Delta_{i+1} + \Delta_i} + O(\Delta) \end{aligned} \tag{3.137}$$

Then for a variable mesh, the equation analogous to the constant mesh Eq. (3.70) are

$$\begin{aligned} &\left(2A_i + \frac{2m}{\hbar^2}V_i + \frac{l(l+1)}{r_i^2}\right)f_i^{nl} - \left(A_iB_i + \frac{2}{r_i}\frac{1}{\Delta_{i+1} + \Delta_i}\right)f_{i+1}^{nl} - \left(A_iC_i - \frac{2}{r_i}\frac{1}{\Delta_{i+1} + \Delta_i}\right)f_{i-1}^{nl} \\ &\qquad = \frac{2m}{\hbar^2}E_n f_i^{nl} + O(\Delta) \end{aligned} \tag{3.138}$$

$$\begin{aligned} A_i &= \left[1 - \frac{\Delta_{i+1}^2 - \Delta_i^2}{r_i(\Delta_{i+1} + \Delta_i)}\right]\left[\frac{1}{\Delta_{i+1}\Delta_i}\right] \\ B_i &= \left[1 - \frac{\Delta_{i+1} - \Delta_i}{\Delta_{i+1} + \Delta_i}\right] \\ C_i &= \left[1 + \frac{\Delta_{i+1} - \Delta_i}{\Delta_{i+1} + \Delta_i}\right] \end{aligned} \tag{3.139}$$

Although the error in Eq. (3.138) is an order of magnitude greater than in Eq. (3.70), if the mesh is divided into regions of constant spacing, only at the points of change will the error be $\sim O(\Delta)$ else it will lie at the more satisfactory $\sim O(\Delta^2)$. In this way it is possible to retain a high degree of accuracy whilst making use of the advantages of a 'variable' mesh in modelling a region large in

spatial extent. The eigenstates satisfying Eq. (3.138) are obtained in exactly the same way as those of Eq. (3.70).

In modelling real systems, when the mesh extends outside the nanocrystal then it is likely that the material parameters will change. If the electron mass is not homogeneous over the mesh, and is taken to be dependant on the r coordinate only, then as suggested in Burt [72], the kinetic energy operator acting on ψ must be modified. If the Laplacian of Schrödinger's constant mass equation is replaced such that

$$\frac{\nabla^2\psi}{m^*} \longrightarrow \nabla\cdot\left(\frac{1}{m^*(r)}\nabla\psi\right)$$

then

$$\nabla\cdot\left(\frac{1}{m^*(r)}\nabla\psi\right) = \frac{1}{m^*(r)}\nabla\cdot\nabla\psi + \nabla\psi\cdot\nabla\left(\frac{1}{m^*(r)}\right) \tag{3.140}$$

since $\nabla\cdot(f\boldsymbol{A}) = f\nabla\cdot\boldsymbol{A} + \boldsymbol{A}\cdot\nabla f$ [73]. Now $\nabla\cdot\nabla\psi$ is simply $\nabla^2\psi$, and given

$$\nabla f = \frac{\partial f}{\partial r}\hat{r} + \frac{1}{r}\frac{\partial f}{\partial\theta}\hat{\theta} + \frac{1}{r\sin\theta}\frac{\partial f}{\partial\phi}\hat{\phi} \tag{3.141}$$

the second scalar product in the RHS of Eq. (3.141) reduces to the product of two partial derivates with respect to r, since m^* is solely dependant on r, then

$$\nabla\cdot\left(\frac{1}{m^*(r)}\nabla\psi\right) = \frac{\nabla^2\psi}{m^*(r)} - \frac{1}{m^*(r)^2}\frac{\partial\psi}{\partial r}\frac{dm^*(r)}{dr} \tag{3.142}$$

The second term in the RHS of Eq. (3.142) is zero except at the boundaries between two different media, and, being dependant on only the r coordinate, effects only the radial Schrödinger equation Eq. (3.64) and its discretisations.

The radial component, $f_{nl}(r)$, of the total wave function for a non-homogeneous media, must therefore satisfy the equation

$$\frac{\hbar^2}{2m^*(r)}\left[\frac{d^2}{dr^2} + \left(\frac{2}{r} - \frac{1}{m^*(r)}\frac{dm^*(r)}{dr}\right)\frac{d}{dr} - \frac{l(l+1)}{r^2}\right]f_{nl}(r) + \left[E_n - V(r)\right]f_{nl}(r) = 0 \tag{3.143}$$

which on a variable mesh, is discretised so that

$$\left(\frac{\hbar^2}{m_i^*}A_i + V_i + \frac{\hbar^2 l(l+1)}{2m_i^* r_i^2}\right) f_i^{nl} - \frac{\hbar^2}{2m_i^*}\left(A_i B_i + \frac{1}{\Delta_{i+1}+\Delta_i}\left(\frac{2}{r_i} - \frac{1}{m_i^*}\frac{m_{i+1}^* - m_{i-1}^*}{\Delta_{i+1}+\Delta_i}\right)\right) f_{i+1}^{nl}$$
$$- \frac{\hbar^2}{2m_i^*}\left(A_i C_i - \frac{1}{\Delta_{i+1}+\Delta_i}\left(\frac{2}{r_i} - \frac{1}{m_i^*}\frac{m_{i+1}^* - m_{i-1}^*}{\Delta_{i+1}+\Delta_i}\right)\right) f_{i-1}^{nl} = E_n f_i^{nl} + O(\Delta) \quad (3.144)$$

A_i, B_i and C_i as defined in Eq. (3.138). Once again this is solved for its eigenstates in the same manner as Eq. (3.70).

This Chapter then has developed techniques to accurately provide the single-particle eigenstates of a given potential formed across inhomogeneous media, and introduced the iterative scheme to self-consistently reconcile these eigenstates with the potential generating them. Up to this point, that potential has not made any allowance for the influence of each electron on its companions, other than through their net charge density in Poisson's equation. Forward then to Chapter 4 and the dissection and modelling of these electron-electron interactions.

Chapter 4 Electron-Electron Interaction Effects

In this Chapter, the complex effects of electron-electron interaction, essential in modelling any multi-electron system, are discussed, and it is shown how the single-particle techniques of the preceding Chapters can be extended and adapted, to provide a self-consistent reference potential possessing the ground state electronic density of the many body interacting system.

Unlike earlier sections, atomic units will be adopted for this Chapter, in order to ease the passage of the reader slightly through the plethora of symbols present in the forthcoming arguments. Thus

$$e = h = m = 1$$

with energies measured in Hartree, H,

$$1\,\mathrm{H} = \frac{\hbar^2}{m a_0^2} \approx 4.36 \times 10^{-18}\,\mathrm{J}$$

and unit of length now the Bohr radius, a_0,

$$1\,a_0 = \frac{\varepsilon_0 h^2}{\pi m e^2} \approx 5.29 \times 10^{-11}\,\mathrm{m}$$

Section 4.1 introduces the difficulties associated with any many bodied system, develops the Hartree and Hartree-Fock approaches, discusses the antisymmetry of the many bodied wave function and Slater determinants, and then proceeds to cover density functional theory (DFT) and the Kohn-Sham method, with the introduction of the Hartree and exchange-correlation potentials. The task of finding the Kohn-Sham single-particle eigenstates is shown to be the same as minimising the total ground state multi-particle energy and the section concludes with a brief outline of the available exchange-correlation approximations.

In Section 4.2 the exact Hartree-Fock exchange term returns in a DFT context, the exchange only optimised effective potential (OEP) method is discussed and the Krieger, Li and Iafrate (KLI) approximation to the exchange potential is introduced. The KLI method is implemented in modelling atomic systems, in good accord with literature examples, and the appropriateness of using the KLI exchange in place of the exact OEP exchange is debated.

Section 4.3 is devoted to the electron-electron interaction of smallest magnitude, correlation, and the uniform electron gas density expansions of Perdew and Wang.

Section 4.4 covers the numerical implementation of the Hartree, exchange and correlation potentials, and discusses the simplifications allowed by the nature of the spherical quantum dot under study. Within the closed subshell approximation, these three potentials are shown to be spherically symmetric and as such, compatible with the finite difference techniques outlined in Chapter 3 to determine the allowed eigenstates of a spherically symmetric quantum well. Applied in conjunction with each other on a homogeneous 10nm radii SnO_2 grain for integer orbital occupancy, several self-consistent potentials resulting from these techniques are shown.

Section 4.5 concludes the Chapter and summarises the developments therein.

4.1 The Many Body Electron Problem and the Kohn-Sham Equations

The essential issue in studying and analysing the electronic structure of matter is not the solution of the Schrödinger equation for isolated electrons, but rather its solution for a system of N interacting electrons. A task of much greater difficulty, with the electrons now free to interact amongst each other: the presence of an electron in one particular region of space influencing the behaviour of the other electrons in the surrounding regions through Coulombic repulsions and the effects of the Pauli exclusion principle. The wave function of this many electron system is then not merely the product of the wave functions of the individual electrons in isolation, but a rather hazier entity altogether. The difficulties associated with accurate calculations of this nature are in essence the quantum mechanical many-body problem.

The Hamiltonian for such a multi-electron system, be it atom, molecule or quantum dot, is

$$\hat{H} = \sum_{i=1}^{N}\left[-\frac{\nabla_i^2}{2} + v(\mathbf{r}_i)\right] + \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1 \\ j\neq i}}^{N}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \tag{4.1}$$

where the first term represents the kinetic energy, the second, the 'external' potential i.e. the potential due to the nuclei, any magnetic fields etc., and the third term is the electron-electron repulsion.

The wave function describing the total electron cloud is then the many bodied wave function $\Psi_k(\mathbf{r}_1, \mathbf{r}_2, \ldots \mathbf{r}_N)$ satisfying Schrödinger's equation

$$\hat{H}\Psi_k(\mathbf{r}_1,\ldots,\mathbf{r}_N) = E_k\Psi_k(\mathbf{r}_1,\ldots,\mathbf{r}_N) \tag{4.2}$$

The electron probability distribution follows from

$$P_k(\mathbf{r}_1,\ldots,\mathbf{r}_N) = |\Psi_k(\mathbf{r}_1,\ldots,\mathbf{r}_N)|^2 \tag{4.3}$$

and determines the probability $P_k(\mathbf{r}_1, \mathbf{r}_2, \ldots \mathbf{r}_N)$ of the existence of an electron at $\mathbf{r}_1$, another at $\mathbf{r}_2$ and so forth.

The ground state of the system is the label applied to the lowest energy eigenstates satisfying Eq. (4.2).

4.1.1 The Hartree and the Hatree-Fock (HF) Approximations

The complexity of the many-electron wave function Ψ_k makes it practically unsuitable, and so one method of approximation, the Hartree approach, is to form a product of single particle wave functions such that

$$\Psi(\mathbf{r}_1,\cdots,\mathbf{r}_N)=\prod_{i=1}^{N}\psi_i(\mathbf{r}_i) \tag{4.4}$$

each of the functions $\psi_i(\mathbf{r}_i)$ satisfying a one-electron Schrödinger equation

$$\begin{aligned}\hat{h}_i\psi_i(\mathbf{r}_i)&=\left(-\frac{\nabla_i^2}{2}+v(\mathbf{r}_i)+v_{eff}(\mathbf{r}_i)\right)\psi_i(\mathbf{r}_i)\\&=\varepsilon_i\psi_i(\mathbf{r}_i)\end{aligned} \tag{4.5}$$

known as the Hartree equation, where the new term $v_{eff}(\mathbf{r}_i)$ represents an averaged potential

$$v_{eff}(\mathbf{r}_i)=\int\frac{\sum_{j\neq i}^{N}\left|\psi_j(\mathbf{r}_j)\right|^2}{\left|\mathbf{r}_i-\mathbf{r}_j\right|}d\mathbf{r}_j \tag{4.6}$$

which is the classical electrostatic potential felt by particle i due to the charge distribution of the other N-1 particles. The product of single particle wave functions, Eq. (4.4), is referred to as a Hartree product.

The total energy is then the sum of the eigenstates, ε_i, minus a term correcting for the double counting of the electron-electron interaction [19,29]:

$$E=\sum_{i=1}^{N}\varepsilon_i-\frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}J_{ij} \tag{4.7}$$

$$J_{ij}=\iint\frac{\left|\psi_i(\mathbf{r}_i)\right|^2\left|\psi_j(\mathbf{r}_j)\right|^2}{\left|\mathbf{r}_i-\mathbf{r}_j\right|}d\mathbf{r}_i d\mathbf{r}_j \tag{4.8}$$

The Hartree Hamiltonian is spin-independent, and as such, its eigenstates are degenerate with respect to spin and therefore spin can be neglected.

Nevertheless, for electrons, and indeed all fermions[i], the wave functions satisfying Eq. (4.4) must be antisymmetric with respect to the interchange of the space and spin coordinates of any pair of

[i] Fermions are particles with half integer values of spin (e.g. protons, neutrons, electrons (all spin-1/2) and Ω^- (spin-3/2)). They are represented by antisymmetric wave functions and obey Fermi-Dirac statistics.

particles. This is the generalisation of the Pauli exclusion principle[ii] to many interacting particles [70,29]. Redefining ψ to include a spin component by extending the single particle wave functions into single particle spin orbitals[iii] such that the variable $\mathbf{x}_i$ condenses into a single variable the spatial coordinates $\mathbf{r}_i$ and the spin coordinates σ_i [29], then it is found that the Hartree product Eq. (4.4) does not fulfil this antisymmetric requirement.

Consider for instance the two electron system

$$\Psi(\mathbf{x}_1,\mathbf{x}_2)=\psi_1(\mathbf{x}_1)\psi_2(\mathbf{x}_2)$$

then clearly

$$\psi_1(\mathbf{x}_1)\psi_2(\mathbf{x}_2)\neq-\psi_1(\mathbf{x}_2)\psi_2(\mathbf{x}_1)$$

However, forming the many electron wave function from linear combinations of Hartree products such that

$$\Psi(\mathbf{x}_1,\mathbf{x}_2)=\frac{1}{\sqrt{2}}\left[\psi_1(\mathbf{x}_1)\psi_2(\mathbf{x}_2)-\psi_1(\mathbf{x}_2)\psi_2(\mathbf{x}_1)\right]$$

does provide a two electron wave function that is antisymmetric under coordinate exchange

$$\left[\psi_1(\mathbf{x}_1)\psi_2(\mathbf{x}_2)-\psi_1(\mathbf{x}_2)\psi_2(\mathbf{x}_1)\right]=-\left[\psi_1(\mathbf{x}_2)\psi_2(\mathbf{x}_1)-\psi_1(\mathbf{x}_1)\psi_2(\mathbf{x}_2)\right]$$

Generally, this can be extended to encompass a system of N electrons, the many bodied wave function now defined with a mathematical construct called a Slater determinant:

$$\begin{aligned}\Psi(\mathbf{x}_1,\ldots,\mathbf{x}_N)&=\frac{1}{\sqrt{N}}\begin{vmatrix}\psi_1(\mathbf{x}_1) & \psi_2(\mathbf{x}_1) & \cdots & \psi_N(\mathbf{x}_1)\\ \psi_1(\mathbf{x}_2) & \psi_2(\mathbf{x}_2) & \cdots & \psi_N(\mathbf{x}_2)\\ \vdots & \vdots & \cdots & \vdots\\ \psi_1(\mathbf{x}_N) & \psi_2(\mathbf{x}_N) & \cdots & \psi_N(\mathbf{x}_N)\end{vmatrix}\\ &=SD\{\psi_1(\mathbf{x}_1)\psi_2(\mathbf{x}_2)\cdots\psi_N(\mathbf{x}_N)\}\end{aligned}\qquad(4.9)$$

Symmetric wave functions describe Bosons (e.g. pions (spin-0), photons and gluons (spin-1), and the theorised gravitons (spin-2)), particles with integral values of spin obeying Bose-Einstein statistics.

[ii] The Pauli exclusion principle states that no two independent fermions can occupy the same state.

[iii] A general spin orbital $a(x)$ consists of the product of a spatial function $\phi(\mathbf{r})$ and a spin function $\xi(\sigma)$. For the electron orbitals, σ is either +½ or -½, usually referred to as spin-up (↑) in the positive case or spin-down in the negative (↓). Then

$$a(\mathbf{x}_i)=\phi_a(\mathbf{r}_i)\xi_i^a(\sigma)$$

The spin functions are required to be orthonormal, and so for the general matrix element

$$\begin{aligned}\left\langle a(\mathbf{x}_1)b(\mathbf{x}_2)\right|1/|\mathbf{r}_1-\mathbf{r}_2|\left|c(\mathbf{x}_1)d(\mathbf{x}_2)\right\rangle&=\left\langle\phi_a(\mathbf{r}_1)\phi_b(\mathbf{r}_2)\right|1/|\mathbf{r}_1-\mathbf{r}_2|\left|\phi_c(\mathbf{r}_1)\phi_d(\mathbf{r}_2)\right\rangle\left\langle\xi_1^a(\sigma)\middle|\xi_1^c(\sigma)\right\rangle\left\langle\xi_2^b(\sigma)\middle|\xi_2^d(\sigma)\right\rangle\\&=\left\langle\phi_a(\mathbf{r}_1)\phi_b(\mathbf{r}_2)\right|1/|\mathbf{r}_1-\mathbf{r}_2|\left|\phi_c(\mathbf{r}_1)\phi_d(\mathbf{r}_2)\right\rangle\delta(\sigma_a,\sigma_c)\delta(\sigma_b,\sigma_d)\end{aligned}$$

A function of the form of Eq. (4.9) is inherently antisymmetric since the interchange of two particles corresponds to an interchange of two rows within the determinant which changes the sign of the determinant (see Appendix II, Introduction to Eigenvalues and Eigenvectors). In addition, if two single particle eigenvectors are identical then two columns of the determinant are identical, the determinant vanishes (again see Appendix II), and therefore one electron and one electron only can occupy a given orbital at any one time (the aforementioned Pauli exclusion principle).

The modification of the Hartree approximation to encompass the Slater determinant representation of the many-bodied wave function leads to the Hartree-Fock approximation.

Separating the Hamiltonian, Eq. (4.1), such that

$$\hat{H} = \sum_{i=1}^{N} \hat{h}_i + \frac{1}{2} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \qquad (4.10)$$

$$\hat{h}_i = \frac{\nabla_i^2}{2} + v(\mathbf{r}_i)$$

where v is again the external potential, then the wave function Eq. (4.9) satisfying the Schrödinger equation $\hat{H}\Psi = E\Psi$ is formed from single orthonormal electron wave functions such that the variational principle [70] (see Appendix III, Introduction to the Calculus of Variations) is satisfied i.e.

$$\delta \langle \Psi | H | \Psi \rangle = 0 \qquad (4.11)$$

Represented as a matrix, the operator $\hat{H}$ will have the diagonal elements $\hat{h}_i$ and the off-diagonal entries $\hat{v}_2(\mathbf{r}_i, \mathbf{r}_j)$ where

$$\hat{v}_2(\mathbf{r}_i, \mathbf{r}_j) = \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \qquad (4.12)$$

and the expectation value of the Hamiltonian is then simply

$$\langle \Psi | \hat{H} | \Psi \rangle = \sum_{i=1}^{N} \langle \psi_i | \hat{h} | \psi_i \rangle + \frac{1}{2} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \left[\langle \psi_i(\mathbf{x}_1) \psi_j(\mathbf{x}_2) | \hat{v}_2(\mathbf{r}_1, \mathbf{r}_2) | \psi_i(\mathbf{x}_1) \psi_j(\mathbf{x}_2) \rangle \right. \qquad (4.13)$$

$$\left. - \langle \psi_i(\mathbf{x}_1) \psi_j(\mathbf{x}_2) | \hat{v}_2(\mathbf{r}_1, \mathbf{r}_2) | \psi_j(\mathbf{x}_1) \psi_i(\mathbf{x}_2) \rangle \right]$$

With the constraint on the single-particle orbitals of orthonormality, the variational equation Eq. (4.11) can be implemented with the Lagrange multipliers λ_{ij}, where $\lambda_{ij}^* = \lambda_{ji}$, yielding

$$\delta\langle\Psi|\hat{H}|\Psi\rangle = \delta\left[\langle\Psi|\hat{H}|\Psi\rangle - \sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_{ij}\langle\psi_i|\psi_j\rangle\right] = 0 \tag{4.14}$$

and the resulting Euler equations given by

$$\hat{h}_i\psi_i(\mathbf{x}_1) + \sum_{j\neq i}^{N}\left[\int\psi_j^*(\mathbf{x}_2)\hat{v}_2(\mathbf{r}_1,\mathbf{r}_2)\psi_j(\mathbf{x}_2)d\mathbf{x}_2\right]\psi_i(\mathbf{x}_1) - \sum_{j\neq i}^{N}\left[\int\psi_j^*(\mathbf{x}_2)\hat{v}_2(\mathbf{r}_1,\mathbf{r}_2)\psi_i(\mathbf{x}_2)d\mathbf{x}_2\right]\psi_j(\mathbf{x}_1) = \sum_{j=1}^{N}\lambda_{ij}\psi_j(\mathbf{x}_1) \tag{4.15}$$

Forming the matrix Λ from the Lagrange multipliers so that the element Λ_{ij} is equal to λ_{ij}, then Λ will be Hermitian since $\lambda_{ij}^* = \lambda_{ji}$, and therefore, there must exist an unitary operator A that will diagonalise Λ via

$$A\Lambda A^{-1} = \varepsilon \tag{4.16}$$

where ε is a diagonal matrix, its non zero entries the eigenvalues ε_i (see footnote [iii] of Section 3.2 and also Apendix II, Introduction to Eigenvalues and Eigenvectors). Transforming the orbitals with $\psi' = A\psi$ then

$$\langle\psi|\Lambda|\psi\rangle = \langle\psi'|\varepsilon|\psi'\rangle \tag{4.17}$$

and the Lagrange multipliers λ_{ij} $i \neq j$ can be said to have been eliminated [70]. The Slater determinant is invariant under unitary transform [29], and therefore the Euler equations can be taken to be also. Thus

$$\hat{F}_i\psi_i'(\mathbf{x}_1) = \varepsilon_i\psi_i'(\mathbf{x}_1) \tag{4.18}$$

where

$$\begin{aligned}\hat{F}_i &= \hat{h}_i + \sum_{j=1}^{N}\left(\hat{J}_j - \hat{K}_j\right)\\ \hat{J}_j\psi_i'(\mathbf{x}_1) &= \left[\int\psi_j'^*(\mathbf{x}_2)\hat{v}_2(\mathbf{r}_1,\mathbf{r}_2)\psi_j'(\mathbf{x}_2)d\mathbf{x}_2\right]\psi_i'(\mathbf{x}_1)\\ \hat{K}_j\psi_i'(\mathbf{x}_1) &= \left[\int\psi_j'^*(\mathbf{x}_2)\hat{v}_2(\mathbf{r}_1,\mathbf{r}_2)\psi_i'(\mathbf{x}_2)d\mathbf{x}_2\right]\psi_j'(\mathbf{x}_1)\end{aligned} \tag{4.19}$$

with the restriction $j \neq i$ dropped as $\hat{J}_i$ and $\hat{K}_i$ are equal at $j = i$ and will therefore cancel each other i.e. a particle in a spin orbital ψ_i does not react with itself. $\hat{J}_i$ and $\hat{K}_i$ are referred to as the Coulomb and exchange operators, $\hat{F}_i$ is known as the Fock operator, and collectively Eq. (4.18) and Eq. (4.19) are known as the Hartree-Fock equations. Most importantly, in Eq. (4.18) the Lagrange multipliers ε_i have become the single electron eigenvalues.

Furthermore, defining

$$J_{ij} = \langle \psi_i(\mathbf{x}_1) | \hat{J}_j | \psi_i(\mathbf{x}_1) \rangle = \int\int \psi_i^*(\mathbf{x}_1)\psi_j^*(\mathbf{x}_2)\hat{v}_2(\mathbf{r}_1,\mathbf{r}_2)\psi_i(\mathbf{x}_1)\psi_j(\mathbf{x}_2)d\mathbf{x}_1 d\mathbf{x}_2$$
$$K_{ij} = \langle \psi_i(\mathbf{x}_1) | \hat{K}_j | \psi_i(\mathbf{x}_1) \rangle = \int\int \psi_i^*(\mathbf{x}_1)\psi_j^*(\mathbf{x}_2)\hat{v}_2(\mathbf{r}_1,\mathbf{r}_2)\psi_j(\mathbf{x}_1)\psi_i(\mathbf{x}_2)d\mathbf{x}_1 d\mathbf{x}_2 \quad (4.20)$$

where the arbitrary prime has been dropped on the orbitals for clarity, then the total energy E, equal to the expectation value of the Hamiltonian Eq. (4.10), may be written

$$E = \sum_{i=1}^{N} E_{ii} + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(J_{ij} - K_{ij}) \quad (4.21)$$

where

$$E_{ii} = \langle \psi_i | \hat{h}_i | \psi_i \rangle \quad (4.22)$$

Pre-multiplying $\hat{F}_i\psi_i$ by the wave function's complex conjugate and integrating yields an explicit expression for ε_i

$$\varepsilon_i = E_{ii} + \sum_{j=1}^{N}(J_{ij} - K_{ij}) \quad (4.23)$$

and therefore the total Hartree-Fock energy can be expressed as

$$E = \sum_{i=1}^{N}\varepsilon_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(J_{ij} - K_{ij}) \quad (4.24)$$

Now, consider the energy difference between this system of N electrons and the energy of a system of N+1 electrons, labelled j =1, 2, ..., N, i. Then

$$\begin{aligned} E_{N+1} - E_N &= \sum_{j=1}^{N} E_{jj} + \frac{1}{2}\sum_{j=1}^{N}\sum_{k=1}^{N}(J_{jk} - K_{jk}) + E_{ii} + 2\frac{1}{2}\sum_{k=1}^{N}(J_{ik} - K_{ik}) - E_N \\ &= E_{ii} + \sum_{k=1}^{N}(J_{ik} - K_{ik}) \\ &= \varepsilon_i \end{aligned} \quad (4.25)$$

and as such the single particle eigenvalue, ε_i can be interpreted as an ionisation energy i.e. the energy required to remove an electron from orbital i assuming that the other orbitals remain unchanged by the process. This is known as Koopman's theorem.

Computationally, the spin orbitals need to be considered separately in terms of their spin and spatial components. In order to remain consistent with the earlier terminology and that of of the main reference of this section, Kohanoff [29], $\psi(\mathbf{r})$ will refer to the spatial dependant part of $\psi(\mathbf{x})$.

From the orthonormality of the spin part of $\psi(\mathbf{x})$, as discussed in footnote [iii] of this section, the Coulomb, J_{ij}, and exchange, K_{ij}, integrals in the Hartree-Fock energy can be written

$$J_{ij} = \int\int \psi_i^*(\mathbf{r}_1)\psi_j^*(\mathbf{r}_2)v_2(\mathbf{r}_1,\mathbf{r}_2)\psi_i(\mathbf{r}_1)\psi_j(\mathbf{r}_2)d\mathbf{r}_1 d\mathbf{r}_2 \qquad (4.26)$$
$$K_{ij} = \delta(\sigma_i,\sigma_j)\int\int \psi_i^*(\mathbf{r}_1)\psi_j^*(\mathbf{r}_2)v_2(\mathbf{r}_1,\mathbf{r}_2)\psi_j(\mathbf{r}_1)\psi_i(\mathbf{r}_2)d\mathbf{r}_1 d\mathbf{r}_2$$

and the Fock operator $\hat{F}_i$ applied to $\psi_i(\mathbf{r})$ gives the Hartree-Fock equations for spatial orbitals:

$$\hat{F}_i\psi_i(\mathbf{r}_1) = \varepsilon_i\psi_i(\mathbf{r}_1) \qquad (4.27)$$

with

$$\hat{F}_i = \hat{h}_i + \sum_{j=1}^{N}\left(\hat{J}_j - \hat{K}_j\right)$$
$$\hat{J}_j = \int \psi_j^*(\mathbf{r}_2)\hat{v}_2(\mathbf{r}_1,\mathbf{r}_2)\psi_j(\mathbf{r}_2)d\mathbf{r}_2 \qquad (4.28)$$
$$\hat{K}_j = \delta(\sigma_i,\sigma_j)\int \psi_j^*(\mathbf{r}_2)\hat{v}_2(\mathbf{r}_1,\mathbf{r}_2)\psi_i(\mathbf{r}_2)d\mathbf{r}_2$$

The Hartree-Fock equations are similar then to the Hartree equations, except that they contain the exchange integral operators which introduce the coupling terms between the different single electron states. These effective exchange potentials are non-local, dependent as they are on the location of the other electrons, and exchange can be said to be fully taken into account within the HF approximation. In contrast, the Hartree approximation only accounts for the other electrons as a mean field i.e. with a local potential. Importantly, the electron-electron self-interaction cancels exactly within the HF method.

It must always be remembered however that the foundation of the Hartree-Fock approach, the anti-symmetric Slater determinant, is only an approximation. In the true many bodied wave function, the electrons will not only experience the effects of exchange and electrostatic repulsion, but will also always act to correlate their movements in order to minimise their total energy. There are many other possible anti-symmetric wave functions which cannot be written as Slater determinants and as such are inaccessible through the HF approximation. There are methods of improving the HF approximation, for instance including correlation effects via linear combinations of Slater determinants or using perturbative methods. Unfortunately, these schemes are quite computationally costly however, and so to avoid these problems, the alternative approach of density functional theory (DFT) is applied here.

4.1.2 Density Functional Theory (DFT)

Using DFT it is possible to determine the electronic ground state density, and, as such, the governing potential and the total ground state energy. Extending this knowledge to cover excited states is discussed in Chapter 7.

Let $\hat{H}$ be a Hamiltonian such that

$$\hat{H} = \hat{T} + \hat{V} + \hat{U}_{ee} \tag{4.29}$$

where $\hat{T}$ is the kinetic energy operator

$$\hat{T} = -\frac{1}{2}\sum_i \nabla_i^2 \tag{4.30}$$

V, the interaction with external fields

$$\hat{V} = \sum_i v(\mathbf{r}_i) \tag{4.31}$$

and $\hat{U}_{ee}$ is the operator for the electron-electron interaction. Using the approach of Levy [74] define the universal functional

$$F[n] = \min_{\Phi \to n} \left\langle \Phi \middle| \hat{T} + \hat{U}_{ee} \middle| \Phi \right\rangle \tag{4.32}$$

for all densities n obtainable from some antisymmetric wave function Φ, these densities consequentially deemed N-representable[iv] . The functional F is referred to as universal since it is explicitly independent of the external potential [29].

Denoting the ground state energy, wave function and density by E_{GS}, Φ_{GS}, and $n_{GS}(\mathbf{r})$ respectively, then the two theorems of density functional theory for N-representable $n(\mathbf{r})$ are [75]:

$$E[n] = F[n] + \int v(\mathbf{r}) n(\mathbf{r}) d\mathbf{r} \geq E_{GS} \tag{4.33}$$

$$E_{GS} = F[n_{GS}] + \int v(\mathbf{r}) n_{GS}(\mathbf{r}) d\mathbf{r} \tag{4.34}$$

To prove Eq. (4.33), let the wave function which minimises Eq. (4.32) be denoted as $\Phi^n_{\min}$ then Eq. (4.33) becomes

$$\left\langle \Phi^n_{\min} \middle| T + U_{ee} \middle| \Phi^n_{\min} \right\rangle + \int v(\mathbf{r}) n(\mathbf{r}) d\mathbf{r} = \left\langle \Phi^n_{\min} \middle| \hat{T} + \hat{V} + \hat{U}_{ee} \middle| \Phi^n_{\min} \right\rangle \geq E_{GS}$$

[iv] Another commonly encountered term is V-representable and refers to the subspace of N-representabile densities that follow from a local potential V.

therefore Eq. (4.33) is true by the definition of the ground state i.e. the minimum of the expectation value of the kinetic and potential energy operators.

From the expectation value of the Hamiltonian for the ground state

$$E_{GS} = \left\langle \Phi_{GS} \middle| \hat{T} + \hat{V} + \hat{U}_{ee} \middle| \Phi_{GS} \right\rangle \tag{4.35}$$

then using Eq. (4.33)

$$\left\langle \Phi_{GS} \middle| \hat{T} + \hat{V} + \hat{U}_{ee} \middle| \Phi_{GS} \right\rangle \leq \left\langle \Phi_{\min}^{n_{GS}} \middle| \hat{T} + \hat{V} + \hat{U}_{ee} \middle| \Phi_{\min}^{n_{GS}} \right\rangle$$

The external potential can be cancelled, so

$$\left\langle \Phi_{GS} \middle| \hat{T} + \hat{U}_{ee} \middle| \Phi_{GS} \right\rangle \leq \left\langle \Phi_{\min}^{n_{GS}} \middle| \hat{T} + \hat{U}_{ee} \middle| \Phi_{\min}^{n_{GS}} \right\rangle \tag{4.36}$$

However, the definition of $\Phi_{\min}^{n_{GS}}$ as the wave function that minimises F requires that

$$\left\langle \Phi_{GS} \middle| \hat{T} + \hat{U}_{ee} \middle| \Phi_{GS} \right\rangle \geq \left\langle \Phi_{\min}^{n_{GS}} \middle| \hat{T} + \hat{U}_{ee} \middle| \Phi_{\min}^{n_{GS}} \right\rangle \tag{4.37}$$

Eq. (4.36) and Eq. (4.36) can only both be simultaneously true if

$$\left\langle \Phi_{GS} \middle| \hat{T} + \hat{U}_{ee} \middle| \Phi_{GS} \right\rangle = \left\langle \Phi_{\min}^{n_{GS}} \middle| \hat{T} + \hat{U}_{ee} \middle| \Phi_{\min}^{n_{GS}} \right\rangle$$

then

$$\begin{aligned} E_{GS} &= \left\langle \Phi_{GS} \middle| \hat{T} + \hat{U}_{ee} \middle| \Phi_{GS} \right\rangle + \int v(\mathbf{r}) n_{GS}(\mathbf{r}) d\mathbf{r} \\ &= \left\langle \Phi_{\min}^{n_{GS}} \middle| \hat{T} + \hat{U}_{ee} \middle| \Phi_{\min}^{n_{GS}} \right\rangle + \int v(\mathbf{r}) n_{GS}(\mathbf{r}) d\mathbf{r} \\ &= F[n_{GS}] + \int v(\mathbf{r}) n_{GS}(\mathbf{r}) d\mathbf{r} \end{aligned}$$

proving Eq. (4.34).

The ground state electronic density then determines the ground state multi-particle wave function, from which all ground state properties can be calculated. This implies that all the ground state properties are in fact functionals of the ground state density. Henceforth the ground state many-electron wave function will simply be represented by the symbol Φ. This discussion of the fundamental properties and theorems of DFT is a reformulation of Hohenberg and Kohn's work [76] to guarantee the antisymmetry of Φ.

How now to actually determine these groundstate entities? The electron-electron interaction energy U_{ee}, the expectation value of the electron-electron interaction operator $\hat{U}_{ee}$, is generally decomposed into two parts: the classical electrostatic energy, E_H, often called the Hartree term (compare with the Hartree Coulomb integral J_{ij})

$$E_H = \frac{1}{2}\int\int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}' \tag{4.38}$$

and exchange and correlation energies, $\tilde{E}_{xc}$. The main problem is the assessment of the many particle kinetic energy term

$$T = \langle \Phi | \hat{T} | \Phi \rangle \tag{4.39}$$

In the Kohn-Sham scheme, the true kinetic energy is split into the sum of the non-interacting kinetic energy i.e. the energy of single-electron wave functions in the absence of electron-electron interactions

$$T_s = -\frac{1}{2}\sum_i f_i \int \psi_i^*(\mathbf{r})\nabla^2\psi_i(\mathbf{r})d\mathbf{r} \tag{4.40}$$

ignoring for the moment any possible spin dependence, and a correlation component. This additional term is included with the exchange and correlation contributions from U_{ee}, this sum labelled E_{xc}. The exact exchange, calculable as in HF theory, accounts for the energy reduction due to antisymmetrisation and corrects the self-interaction of the Hartree energy. The correlation energy is now the energy difference between the full ground state energy and the one obtained from the approximate Slater determinant. It allows for the kinetic energy difference between the interacting and non-interacting systems and accounts for the energy reduction as the electrons coordinate their movements to minimise their Coulomb energy.

The Kohn-Sham approach is then conceptually simple. Assume that a system of non-interacting electrons exists, whose ground state density is identical to that of the interacting system of electrons. As their kinetic energy and exchange energy can be calculated exactly, provided a realistic approximation can be made of the correlation effects, then in principle, by minimising the non-interacting system with respect to its density, the true ground state density of the interacting system can be found.

Therefore, separating the energy functional into its constituent parts such that

$$E_{KS}[n] = T_s[n] + \int n(\mathbf{r})v(\mathbf{r})d\mathbf{r} + \frac{1}{2}\int\int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + E_{xc}[n] \tag{4.41}$$

and applying the variational principle, with the constraint that the density integrates to N electrons:

$$\frac{\delta}{\delta n(\mathbf{r})}\left[E_{KS}[n]-\mu\int n(\mathbf{r})d\mathbf{r}\right]=0 \tag{4.42}$$

then

$$\frac{\delta E_{KS}[n]}{\delta n(\mathbf{r})}=\frac{\delta T_S[n]}{\delta n(\mathbf{r})}+v(\mathbf{r})+\int\frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}dr'+\frac{\delta E_{xc}[n]}{\delta n(\mathbf{r})}=\mu \tag{4.43}$$

where μ is the Lagrange multiplier. The functional derivative of the exchange-correlation energy with respect to the electronic density is called the exchange-correlation potential, $v_{xc}(\mathbf{r})$, and its preceding term in Eq. (4.43) is referred to as the Hartree potential, $v_H(\mathbf{r})$ – note the similarity to the effective potential v_{eff} of Eq. (4.6) in the Hartree approximation. Both these potentials and the external potential are local.

Next consider a system of non-interacting particles confined within a local potential $v_s(\mathbf{r})$. In the absence of the electron-electron interactions, Hartree and exchange-correlation, the minimisation condition requires

$$\frac{\delta T_S[n]}{\delta n(\mathbf{r})}+v_s(\mathbf{r})=\mu \tag{4.44}$$

Denoting the density that solves this as $n_s(\mathbf{r})$, then these equations, Eq. (4.43) and Eq. (4.44), have identical solutions, $n_s(\mathbf{r}) = n(\mathbf{r})$, provided that the potential v_s is chosen to satisfy

$$v_s(\mathbf{r})=v(\mathbf{r})+v_H(\mathbf{r})+v_{xc}(\mathbf{r}) \tag{4.45}$$

Ergo, by solving the Schrödinger equation

$$\left[-\frac{\nabla^2}{2}+v_s(\mathbf{r})\right]\psi_i(\mathbf{r})=\varepsilon_i\psi_i(\mathbf{r}) \tag{4.46}$$

for the non interacting single body eigenvectors in a potential $v_s(\mathbf{r})$, the density, $n(\mathbf{r})$, of the interacting many bodied system in a potential $v(\mathbf{r})$[v] is ascertained:

$$n(\mathbf{r})=\sum_i f_i|\psi_i(\mathbf{r})|^2 \tag{4.47}$$

Along with expressions for the Hartree and exchange-correlation potential, the two equations Eq. (4.45) and Eq. (4.46) form the Kohn-Sham equations, the self-consistent solutions of which can be determined using the procedures of Chapter 3 – see Section 4.4. The use of a local exchange-

[v] The external potential $v(\mathbf{r})$ for the quantum dot system of interest here follows from the solution of Poisson's equation, see for example Eq. (2.10). In this way, v will in general be refered to as v_p if it is generated from Poisson's equation alone, or v_{sp} if it follows from some iterative procedure coupling Poisson's equation and a Schrodinger-like equation.

correlation potential is in contrast to the non-local exchange potentials of the Hartree-Fock approach, and this reduction from integral exchange operators to a relatively simple multiplicative exchange potential makes the KS equations far quicker to solve than older HF ones.

It is usual to state that the KS approach maps the interacting many bodied problem onto an auxiliary single-particle problem, and that the task of minimising $E_{KS}[n]$ has been replaced by that of solving a non-interacting Schrödinger equation. However, these single-particle eigenstates have no physical meaning by themselves in the strictest sense[vi], and only form part of the mathematical construction devised to yield the correct density. Yet they can provide good approximations to the excitation energies of physical systems [23], and indeed have mathematically been shown to be the excitation energies to zeroth order in the electron-electron interaction [28]. Detailed theoretical calculations with comparisons to experimental data [24,25,26,27] demonstrate that the KS eigenvalues approximate very well the vertical ionisation potentials. For valence orbitals, Gritsenko *et al.* [27] suggest that for closed-shell molecules the KS eigenvalue - ionisation potential deviation can be as little as 0.08eV (<0.1%). For the lower valence and core levels, the deviation can be substantially greater, although Chong *et al.* [26] indicates that over 64 different molecules, over all the 406 ionisation potentials, the maximum error is only ~3.6 % and the average energy difference only 0.4eV.

Whilst the KS eigenstates are taken as acceptable approximations to physical eigenstates for the main results contained in Chapter 6 of this work, Chapter 7 discusses a method of accurately including self-energy using Green's functions in order to assess the true electron addition and removal energies of the many body system for future investigations.

The Kohn-Sham method and the DFT ground state density, although often considered concepts applicable only at absolute zero with integer orbital occupancies, hold at elevate temperatures, see for example Kohn *et al.* [77] or the work of Mermin [78], and with fractional occupancies, see for instance Perdew *et al.* [30], although some articles contend this, see for example Kohanoff [29]. Fractional occupancies appear in a KLI context (see Section 4.2) in Krieger *et al.* [21], and most relevantly to the methods employed within this treatise, are calculated by Torsti *et al.* [20] with Fermi-Dirac statistics to simulate a Na quantum dot system at 1200 K (see Section 1.1).

[vi] An exception to this is the highest occupied KS eigenvalue, which does satisfy Koopman's theorem when the exact functional form of the exchange energy is used. This is discussed in more detail in Section 4.1

Naturally, maintaining a thermal electron component in the total electron density in this way makes this total density very sensitive to the accuracy of all the Kohn-Sham energy levels, including those that would be considered the 'excited' eigenstates of a more traditional integer occupancy scheme. As such, it is imperative that the highest accuracy functionals alone are used within the Kohn-Sham method [23] if these excited states, and so their Fermi-Dirac dependent fractional occupancies and consequentially the variationally minimising density, are to have meaning in this treatise.

The next stage is to treat the spin of the electrons. The extension of the Kohn-Sham theory, and DFT in general, to include spin is not however completely straightforward. As discussed in Eschrig and Pickett [79] for instance, the spin-DFT ground state density does not uniquely determine the potential, although Kohn *et al.* [80] estimates that the effects of this are not significant.

It is sufficient to separate the total density into two independent spin densities such that

$$\begin{aligned} n(\mathbf{r}) &= n_{\uparrow}(\mathbf{r}) + n_{\downarrow}(\mathbf{r}) \\ &= \sum_i n_{i\uparrow}(\mathbf{r}) + n_{i\downarrow}(\mathbf{r}) \end{aligned} \tag{4.48}$$

where

$$n_{i\sigma}(\mathbf{r}) = f_{i\sigma}\psi^*_{i\sigma}(\mathbf{r})\psi_{i\sigma}(\mathbf{r}) \tag{4.49}$$

then the single-particle Kohn-Sham spin orbitals must self-consistently satisfy

$$\left[-\frac{\nabla^2}{2} + v_{s\sigma}(\mathbf{r})\right]\psi_{i\sigma}(\mathbf{r}) = \varepsilon_{i\sigma}\psi_{i\sigma}(\mathbf{r}) \tag{4.50}$$

where

$$v_{s\sigma}(\mathbf{r}) = v(\mathbf{r}) + v_H(\mathbf{r}) + v_{xc\sigma}(\mathbf{r}) \tag{4.51}$$

with the spin dependant exchange-correlation potential

$$v_{xc\sigma}([n_\sigma];\mathbf{r}) = \frac{\delta E_{xc}[n_\sigma]}{\delta n_\sigma(\mathbf{r})} \tag{4.52}$$

The Coulomb potential v_H remains unchanged, dependant as it is solely on the total electronic density

$$v_H([n];\mathbf{r}) = \int \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}' \tag{4.53}$$

The total energy of the system in spin DFT is therefore [29]

$$E_{KS}[n_\sigma] = T_s[n_\sigma] + \int n(\mathbf{r})v(\mathbf{r})d\mathbf{r} + \frac{1}{2}\int\int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + E_{xc}[n_\sigma]$$
$$= \sum_\sigma \sum_i f_{i\sigma}\varepsilon_{i\sigma} - \frac{1}{2}\int\int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}' - \int v_{xc}(\mathbf{r})n(\mathbf{r})d\mathbf{r} + E_{xc}[n_\sigma] \quad (4.54)$$

where the spin-polarised expressions for the exchange-correlation potential have been replaced with the un-polarised potential, and

$$T_s[n_\sigma] = -\frac{1}{2}\sum_\sigma \sum_i f_{i\sigma}\int \psi^*_{i\sigma}(\mathbf{r})\nabla^2\psi_{i\sigma}(\mathbf{r})d\mathbf{r} \quad (4.55)$$

has been used for the kinetic energy.

But what of E_{xc} and its functional derivative v_{xc}? Up to this point DFT has been exact and no approximations introduced, however the precise form of E_{xc} is unknown. As a consequence of previous discussions (specifically those following Eq. (4.34)), E_{xc} is guaranteed to be a density functional – the exchange correlation energy is a component of the ground state energy and all ground state properties are functionals of the ground state density, but it is not explicitly known.

Broadly, there are three levels of approximation in the literature to address the exchange-correlation energy [81].

The first is the local density approximation (LDA). Here the general inhomogeneous electronic system is treated as locally homogeneous electron gas, the simplest system of correlated electrons, substituting in the variable electron density into known uniform-density expressions.

For instance, dividing E_{xc} into individual exchange and correlation components, the homogeneous exchange energy per electron is given exactly by the expression [29]

$$\varepsilon_x^{LDA}[n] = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3} n^{1/3} \quad (4.56)$$

and so the LDA exchange energy is merely

$$E_x^{LDA}[n] = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3}\int n(\mathbf{r})\varepsilon_x^{LDA}[n(\mathbf{r})]d\mathbf{r} \quad (4.57)$$

For spin dependent systems Eq. (4.56) becomes

$$\varepsilon_x^{LDA}[n,\zeta] = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3} n^{1/3}\frac{(1+\zeta)^{4/3}+(1-\zeta)^{4/3}}{2} \quad (4.58)$$

with the introduction of the relative spin polarisation parameter ζ of Perdew and Wang [22]

$$\zeta = \frac{n_\uparrow - n_\downarrow}{n_\downarrow + n_\uparrow} \tag{4.59}$$

The correlation contribution can be excellently approximated. Several schemes exist, the most accurate based on the quantum Monte Carlo calculations of Ceperley and Alder [82] for a homogeneous electron gas, which are exact within numerical accuracy. The parameterisation of these results by Perdew and Wang [22] are discussed in Section 4.3.

In general however, the use of a homogeneous electron gas as a reference system can introduce considerable errors, especially when treating systems with densities that are considerably non-uniform, such as molecules for example. In addition, the Hartree self-interaction terms are not completely cancelled by the LDA exchange-correlation terms and most problematically, the LDA v_{xc} decays exponentially instead of proportionally to $-1/r$, particularly affecting ionisation energies. The LDA is also notoriously poor in its calculation of band gaps, see for example Muscat *et al.* [83]. Although in its favour, the LDA does satisfy the sum rule for the exchange-correlation hole[vii], and can model systems with strong bonds (metallic, ionic and covalent) well, in particular finding considerable success when applied to bulk metals. The model's successes and limits are covered in some detail in Kohanoff [29].

The second level of exchange-correlation approximation, gradient expansions (GEs), addresses the LDAs' neglect of inhomogeneities in the electron density, making a series expansion of E_{xc} in terms of the density and its gradients. In general,

$$E_{xc}^{GE}[n] = \int n(\mathbf{r}) \varepsilon_{xc}^{LDA}[n(\mathbf{r})] F_{xc}[n(\mathbf{r}), \nabla n(\mathbf{r}), \nabla^2 n(\mathbf{r}), \ldots] \tag{4.60}$$

where the enhancement factor F_{xc} modifying the LDA expression is given to fourth order by

$$F_{xc}(p,q) = 1 + \frac{10}{81} p + \frac{146}{2025} q^2 - \frac{73}{405} qp + Dp^2 + O(\nabla n^6) \tag{4.61}$$

where

[vii]The exchange-correlation hole represents the reduction in probability of finding a second electron in the immediate vicinity of an existing electron. The exchange hole follows from the tendency of spin alike electrons to avoid each other and the correlation hole arises from the movements of electrons to minimise their Coulomb energy. Importantly, taken together the exchange-correlation hole about a given electron contains exactly one displaced electron – this is known as the sum rule.

$$p = \frac{|\nabla n|^2}{4(3\pi^2)^{2/3} n^{8/3}}$$
$$q = \frac{\nabla^2 n}{4(3\pi^2)^{2/3} n^{5/3}} \tag{4.62}$$

The first two coefficients are exact, whilst the coefficient of qp has been estimated to an accuracy of ~20% and the final coefficient D has not been calculated explicitly, though is estimated to be zero [29].

Though improving on the LDA for binding energies and general atomic energies, and with some improvement in the calculation of band gaps, gradient expansion approximations unfortunately offer no improvement in general for the properties of semiconductors. They retain the incorrect exponential decay of the exchange correlation potential in the asymptotic limit of the local density approximation, and still do not compensate completely for the self-interactions of the Hartree term.

Whilest there are many schemes to improve on these avenues of approximation, see Kohanoff [29], the third major level in the estimation of ground state functionals are orbital dependant expressions, which form implicit, rather than explicit, functionals of the density. In this way, the KS theory can be reformulated to use the HF representation of the exact exchange, although substituting the Kohn-Sham single electron orbitals in place of the HF wave functions. Then, with a suitable approximation of the correlation contribution, the total energy is minimised under the constraint that these orbitals must form solutions of a non-interacting problem in a local potential. This is known as the optimised effective potential method (OEP). As discussed in the next section, Section 4.4, the self-consistent generation of a potential and its corresponding eigenstates is not a simple task, so for practicality, the KLI approximation to this optimised effective potential is used here.

In the ideal case, due to the rather arbitrary division of E_{xc} into E_x and E_c, it is likely that it is best to treat both terms, exchange and correlation, in a consistent manner to achieve the optimal compensation of errors introduced in the approximation of the correlation term. Thus, by treating exchange as a functional of the orbitals then, although the correlation is of a significantly smaller magnitude, the most accurate results are likely to follow if the correlation is treated in a similar way. However, orbital dependant representations of the correlation energy are not exact and are complex objects indeed. The method of Engel and Dreizler [81] will be briefly discussed in

Chapter 7, but for now such procedures are un-practical. For the work undertaken here, the exchange energy will be in its exact form, to allow self-interaction corrections and in order to achieve the correct asymptotic behaviour, but the correlation energy will be approximated via the parameterisation of Perdew and Wang [22]. Indeed, a similar pairing – KLI exchange and a polynomial approximation of E_c by Tantar and Cerpely [84] – was used by Mucucci, Hess and Iafrate [19] to model 2-dimensional circular quantum dots. This coupling is accurate enough for the current purposes of the author.

4.2 The Optimised Effective Potential (OEP), Exact Exchange and the KLI Exchange Approximation

Defining the Kohn-Sham exchange-energy functional in order that in conjunction with the expression for the Hartree or Coulomb energy, Eq. (4.38), this functional will cancel any self-interaction effects [81], ergo

$$E_x[n_\sigma] = -\frac{1}{2}\sum_\sigma \sum_{ij} f_{i\sigma} f_{j\sigma} \int\int \frac{\psi_{i\sigma}^{KS*}(\mathbf{r}')\psi_{j\sigma}^{KS*}(\mathbf{r})\psi_{i\sigma}^{KS}(\mathbf{r})\psi_{j\sigma}^{KS}(\mathbf{r}')}{|\mathbf{r}'-\mathbf{r}|} d\mathbf{r}'d\mathbf{r} \tag{4.63}$$

with the exchange potential related to this via

$$v_{x\sigma}([n_\sigma];\mathbf{r}) = \frac{\delta E_x[n_\sigma]}{\delta n_\sigma(\mathbf{r})} \tag{4.64}$$

then an avenue is available for the exact treatment of exchange within DFT, although this is not as straightforward at might first appear.

4.2.1 The Optimised Effective Potential (OEP)

The functional derivative of Eq. (4.63) with respect to n_σ does not follow immediately, but must be evaluated as a chain expansion

$$v_{x\sigma}(\mathbf{r}) = \int d\mathbf{r} \frac{\delta v_{s\sigma}(\mathbf{r}')}{\delta n_\sigma(\mathbf{r})} \sum_i \left(\int d\mathbf{r}'' \left[\frac{\delta \psi_{i\sigma}^{KS*}(\mathbf{r}'')}{\delta v_{s\sigma}(\mathbf{r}')} \frac{\delta E_x}{\delta \psi_{i\sigma}^{KS*}(\mathbf{r}'')} + c.c. \right] + \frac{\delta \varepsilon_{i\sigma}^{KS}}{\delta v_{s\sigma}(\mathbf{r}')} \frac{\partial E_x}{\partial \varepsilon_{i\sigma}^{KS}} \right) \tag{4.65}$$

where *c.c.* indicates the complex conjugate of the first term.

The derivatives of the energy with respect to the eigenvalues and eigenvectors can be calculated directly from Eq. (4.63)

$$\begin{aligned} \frac{\partial E_x}{\partial \varepsilon_{i\sigma}^{KS}} &= -\frac{\partial f_{i\sigma}}{\partial \varepsilon_{i\sigma}^{KS}} \sum_j f_{j\sigma} \int\int \frac{\psi_{i\sigma}^{KS*}(\mathbf{r}')\psi_{j\sigma}^{KS*}(\mathbf{r})\psi_{i\sigma}^{KS}(\mathbf{r})\psi_{j\sigma}^{KS}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}'d\mathbf{r} \\ \frac{\delta E_x}{\delta \psi_{i\sigma}^{KS}(\mathbf{r})} &= -\sum_j f_{i\sigma} f_{j\sigma} \int \frac{\psi_{i\sigma}^{KS*}(\mathbf{r}')\psi_{j\sigma}^{KS*}(\mathbf{r})\psi_{j\sigma}^{KS}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}' \end{aligned} \tag{4.66}$$

while the remaining derivatives can be assessed by the response of the system to a small perturbation; should the potential $v_{s\sigma}$ in Eq. (4.50) be perturbed by $\delta v_{s\sigma}$ then

$$\begin{aligned} \delta \psi_{i\sigma}^{KS}(\mathbf{r}) &= -\int d\mathbf{r}' G_{i\sigma}(\mathbf{r},\mathbf{r}') \psi_{i\sigma}^{KS}(\mathbf{r}') \delta v_{s\sigma}(\mathbf{r}') \\ \delta \varepsilon_{i\sigma}^{KS} &= \int d\mathbf{r}\, \psi_{i\sigma}^{KS*}(\mathbf{r}) \psi_{i\sigma}^{KS}(\mathbf{r}) \delta v_{s\sigma}(\mathbf{r}) \end{aligned} \tag{4.67}$$

where

$$G_{i\sigma}(\mathbf{r},\mathbf{r}') = \sum_{j\neq i} \frac{\psi_{j\sigma}^{KS}(\mathbf{r})\psi_{j\sigma}^{KS*}(\mathbf{r}')}{\varepsilon_{j\sigma}^{KS} - \varepsilon_{i\sigma}^{KS}} \tag{4.68}$$

see Appendix III, Introduction to Green's Functions for further detail, and consequently

$$\begin{aligned} \frac{\delta\psi_{i\sigma'}^{KS}(\mathbf{r}')}{\delta v_{s\sigma}(\mathbf{r})} &= -G_{i\sigma}(\mathbf{r}',\mathbf{r})\psi_{i\sigma}^{KS}(\mathbf{r})\delta_{\sigma\sigma'} \\ \frac{\delta\varepsilon_{i\sigma'}^{KS}}{\delta v_{s\sigma}(\mathbf{r})} &= \psi_{i\sigma}^{KS*}(\mathbf{r})\psi_{i\sigma}^{KS}(\mathbf{r})\delta_{\sigma\sigma'} \end{aligned} \tag{4.69}$$

Furthermore, the variation of the density with respect to the potential, often called the linear response function, $\chi_{s\sigma}(\mathbf{r}, \mathbf{r}')$, can now be determined such that

$$\begin{aligned} \chi_{s\sigma}(\mathbf{r},\mathbf{r}') &= \frac{\delta n_\sigma(\mathbf{r})}{\delta v_{s\sigma}(\mathbf{r}')} \\ &= \sum_i \int d\mathbf{r}'' \left(\frac{\delta n_\sigma(\mathbf{r})}{\delta\psi_{i\sigma}^{KS}(\mathbf{r}'')} \frac{\delta\psi_{i\sigma}^{KS}(\mathbf{r}'')}{\delta v_{s\sigma}(\mathbf{r}')} + c.c. \right) \\ &= -\sum_i f_{i\sigma}\psi_{i\sigma}^{KS*}(\mathbf{r})G_{i\sigma}(\mathbf{r},\mathbf{r}')\psi_{i\sigma}^{KS}(\mathbf{r}') + c.c. \end{aligned} \tag{4.70}$$

Eq. (4.65) actually requires the inverse of this, however, Eq. (4.70) is very ill conditioned with respect to inversion, so in an effort to avoid this, multiply Eq. (4.65) by $\chi_{s\sigma}(\mathbf{r}, \mathbf{r}'')$ and integrate over $\mathbf{r}''$. Then since

$$\int dr'' \chi_{s\sigma}^{-1}(\mathbf{r},\mathbf{r}'')\chi_{s\sigma}(\mathbf{r}'',\mathbf{r}') = \delta(\mathbf{r}-\mathbf{r}') \tag{4.71}$$

the OEP integral equation

$$\int d\mathbf{r}' \chi_{s\sigma}(\mathbf{r},\mathbf{r}')v_{x\sigma}(\mathbf{r}') = \Lambda_{x\sigma}(\mathbf{r}) \tag{4.72}$$

is finally obtained, where

$$\Lambda_{x\sigma}(\mathbf{r}) = \sum_i \left(\int d\mathbf{r}' \left[\psi_{i\sigma}^{KS*}(\mathbf{r})G_{i\sigma}(\mathbf{r},\mathbf{r}') \frac{\delta E_x}{\delta\psi_{i\sigma}^{KS*}(\mathbf{r}')} + c.c. \right] + \psi_{i\sigma}^{KS*}(\mathbf{r})\psi_{i\sigma}^{KS}(\mathbf{r}) \frac{\partial E_x}{\partial\varepsilon_{i\sigma}^{KS}} \right) \tag{4.73}$$

However,

$$\int d\mathbf{r}\, \psi_{i\sigma}^{KS*}(\mathbf{r})G_{i\sigma}(\mathbf{r},\mathbf{r}') = 0 \tag{4.74}$$

as a consequence of the orthonormality of the wave functions, therefore any solution $v_{x\sigma}$ of Eq. (4.72) is only determined up to a constant and its exact magnitude must be set from the requirement that $v_{x\sigma}$ must minimise the total energy of the system. That in mind, let the total energy, E, be given by

$$E[n_e] = -\frac{1}{2}\sum_{i,\sigma} f_{i\sigma}\int d\mathbf{r}\,\psi_{i\sigma}^{KS*}(\mathbf{r})\nabla^2\psi_{i\sigma}^{KS}(\mathbf{r}) + \int d\mathbf{r}\,v(\mathbf{r})n_e(\mathbf{r}) + \frac{1}{2}\int\int\frac{n_e(\mathbf{r})n_e(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}d\mathbf{r}' + E_x[n_e] \tag{4.75}$$

where the eigenstates satisfy

$$h_\sigma\psi_{i\sigma}^{KS}(\mathbf{r}) = \left(-\frac{\nabla^2}{2} + v_{s\sigma}(\mathbf{r})\right)\psi_{i\sigma}^{KS}(\mathbf{r}) = \varepsilon_{i\sigma}^{KS}\psi_{i\sigma}^{KS}(\mathbf{r}) \tag{4.76}$$

Then, should the spin fractional occupancy of an eigenstate p be altered slightly such that $f_{p\sigma} \rightarrow f_{p\sigma} + \delta f_{p\sigma}$, where $f_{i\sigma}$ remains constant for all eigenstates $i \neq p$, it follows that the self-consistent potential will change by some small amount, from $v_{s\sigma}$ to $v_{s\sigma} + \delta v_{s\sigma}$ in order that the total energy remains minimised. This change in total energy can be related to the alterations in the other two quantities via

$$\delta E = \left.\frac{\partial E}{\partial f_{p\sigma}}\right|_{v_{s\sigma'}} \delta f_{p\sigma} + \sum_{\sigma'}\int d\mathbf{r}\left.\frac{\delta E}{\delta v_{s\sigma'}}\right|_{f_{p\sigma}} \delta v_{s\sigma'} \tag{4.77}$$

For the total energy to be minimized with respect to the potential $v_{s\sigma}$, then

$$\frac{\delta E[n_e]}{\delta v_{s\sigma}(\mathbf{r})} = \sum_{i,\sigma'}\int d\mathbf{r}'\left(\frac{\delta E}{\delta\psi_{i\sigma'}^{KS}(\mathbf{r}')}\frac{\delta\psi_{i\sigma'}^{KS}(\mathbf{r}')}{\delta v_{s\sigma}(\mathbf{r})} + c.c.\right) = 0 \tag{4.78}$$

implying that Eq, (4.77) reduces to

$$\frac{\delta E}{\delta f_{p\sigma}} = \frac{\partial E}{\partial f_{p\sigma}} \tag{4.79}$$

so

$$\begin{aligned}\int d\mathbf{r}\frac{\delta n_e(\mathbf{r})}{\delta f_{p\sigma}}\frac{\delta E}{\delta n_e(\mathbf{r})} &= \int d\mathbf{r}\,\psi_{p\sigma}^{KS*}(\mathbf{r})h_\sigma\psi_{p\sigma}^{KS}(\mathbf{r}) \\ &= \int d\mathbf{r}\frac{\psi_{p\sigma}^{KS}(\mathbf{r})}{f_{p\sigma}}\frac{\delta E}{\delta\psi_{p\sigma}^{KS}(\mathbf{r})} \\ &= \varepsilon_{p\sigma}^{KS}\end{aligned} \tag{4.80}$$

This provides the normalisation constraint on $v_{x\sigma}$, since the exact $v_{x\sigma}$ is needed in $v_{s\sigma}$ for Eq. (4.78) to hold and Eq. (4.80) to be true. Eq. (4.80) can be written as

$$\int d\mathbf{r}\,\psi_p^{KS*}(\mathbf{r})\left(\psi_p^{KS}(\mathbf{r})v_x(\mathbf{r}) - \frac{1}{f_p}\frac{\delta E_x}{\delta\psi_p^{KS*}(\mathbf{r})}\right) + c.c. = 0 \tag{4.81}$$

However, for m the highest occupied energy orbital

$$\begin{aligned}\varepsilon_{m\sigma}^{KS} &= \int \psi_{m\sigma}^{KS*}(\mathbf{r}) h_\sigma \psi_{m\sigma}^{KS}(\mathbf{r}) d\mathbf{r} \\ &= \int \psi_{m\sigma}^{KS*}(\mathbf{r}) \left(-\frac{\nabla^2}{2} + v(\mathbf{r}) + v_H(\mathbf{r}) + v_{xc\sigma}(\mathbf{r}) \right) \psi_{m\sigma}^{KS}(\mathbf{r}) \\ &= \int \psi_{m\sigma}^{KS*}(\mathbf{r}) \left(-\frac{\nabla^2}{2} + v(\mathbf{r}) + v_H(\mathbf{r}) + v_{m\sigma}(\mathbf{r}) \right) \psi_{m\sigma}^{KS}(\mathbf{r})\end{aligned}$$

employing Eq. (4.97) in conjunction with footnote [i] of this section to arrive at the last line, the potential $v_{m\sigma}$ to be defined presently in Eq. (4.85). Importantly $\bar{v}_{xom} = \bar{v}_{m\sigma}$ only holds for the m orbital as discussed in some detail later. Then

$$\begin{aligned}\varepsilon_{m\sigma}^{KS} &= \int \psi_{m\sigma}^{KS*}(\mathbf{r}) \left(-\frac{\nabla^2}{2} + v(\mathbf{r}) \right) \psi_{m\sigma}^{KS}(\mathbf{r}) d\mathbf{r} \\ &\quad + \int \psi_{m\sigma}^{KS*}(\mathbf{r}) \sum_j^N f_{j\sigma} \int \psi_{j\sigma}^{KS*}(\mathbf{r}_2) \hat{v}_2(\mathbf{r}, \mathbf{r}_2) \psi_{j\sigma}^{KS}(\mathbf{r}_2) d\mathbf{r}_2 \psi_{m\sigma}^{KS}(\mathbf{r}) d\mathbf{r} \\ &\quad - \int \sum_j f_{j\sigma} \int \psi_{m\sigma}^{KS*}(\mathbf{r}_2) \psi_{j\sigma}^{KS*}(\mathbf{r}) \hat{v}_2(\mathbf{r}, \mathbf{r}_2) \psi_{j\sigma}^{KS}(\mathbf{r}_2) \psi_{m\sigma}^{KS}(\mathbf{r}) d\mathbf{r}_2 d\mathbf{r} \\ &= \int \psi_{m\sigma}^{KS*}(\mathbf{r}) \hat{h}_m^{HF} \psi_{m\sigma}^{KS}(\mathbf{r}) d\mathbf{r} \\ &\quad + \int \psi_{m\sigma}^{KS*}(\mathbf{r}) \sum_j^N f_{j\sigma} \hat{J}_{j\sigma}^{HF} \psi_{m\sigma}^{KS}(\mathbf{r}) d\mathbf{r} - \int \psi_{m\sigma}^{KS*}(\mathbf{r}_2) \sum_j^N f_{j\sigma} \hat{K}_{j\sigma}^{HF} \psi_{m\sigma}^{KS}(\mathbf{r}_2) d\mathbf{r}_2 \\ &= \int \psi_{m\sigma}^{KS*}(\mathbf{r}) \left[\hat{h}_m^{HF} + \sum_j^N f_{j\sigma} \left(\hat{J}_{j\sigma}^{HF} - \hat{K}_{j\sigma}^{HF} \right) \right] \psi_{m\sigma}^{KS}(\mathbf{r}) d\mathbf{r} \\ &= \int \psi_{m\sigma}^{KS*}(\mathbf{r}) \varepsilon_{m\sigma}^{HF} \psi_{m\sigma}^{KS}(\mathbf{r}) d\mathbf{r} \qquad (4.82)\end{aligned}$$

generalising the Hartree-Fock equations (compare with Eq. (4.27) and Eq. (4.28)) for spin fractional occupancy and using the new notation of including the spin as a subscript. The Fock operator (square parenthesise, see Eq. (4.19)) however has been evaluated with the OEP orbitals. Then, from Eq. (4.82) in conjunction with Eq. (4.25) it is possible to write

$$\varepsilon_{m\sigma}^{KS} = \bar{E}^{HF} - \bar{E}_-^{HF} \qquad (4.83)$$

where $\bar{E}^{HF}$ represents the expectation value of the total energy, which has been determined in the manner of Eq. (4.24), and the minus on the subscript indicates that the m orbital has been discounted in the energetic summation. As such, Koopman's theorem (see Section 4.1) can be said to be satisfied, but only for the highest occupied KS orbital [21].

Therefore, unless p is equal to m, then although Eq. (4.81) should be obeyed by any correctly normalized self-consistent KS solution, it will lack any physical significance, in the strictest sense, with regard to the correct rate of change of E with respect to f, since

(4.84)

$$\varepsilon_p^{KS} \neq \bar{\varepsilon}_p^{HF} \text{ unless } p = m$$

although E^{KS} will of course equal E^{HF}. To be exact, the Hartree-Fock eigenvalues are recognized as the unrelaxed electron-removal energies, whereas only the highest occupied KS eigenvalue is considered an ionisation energy (but inclusive of relaxation effects).

Consequently, for best results, $v_{x\sigma}$ is only required to satisfy Eq. (4.81) for $p = m$, and then Eq. (4.81) can be considered to provide a generalisation of Engel *et al.* [81]'s normalisation constraint to the fractional occupancy regime.

Although at first sight the OEP integral equation Eq. (4.72) appears straightforward to solve, it is very numerically demanding. In Engel and Vosko [85], the interested reader can find a technical discussion for spherical systems (generally meaning atoms, but Kotani [86,87] demonstrates the method within band structure calculations), where two main refinements are added to the preceding arguments. Firstly, the Green's function of Eq. (4.68) are replaced with Green's functions formed from products of the first and second solutions to the radial Kohn-Sham equation, then secondly, the analytic asymptotic form of $v_{x\sigma}$ is used to enhance accuracy of the constructed numerical exchange potential. However, this still leaves the solution the integral equation a far from trivial task.

4.2.2 The Krieger, Li and Iafrate (KLI) Approximation

An alternative, semi-analytical approximate scheme to solving the full, numerically demanding integral equation was proposed by Krieger, Li and Iafrate. Indeed, there are three approaches to the KLI approximation, briefly outlined in Engel *et al.* [81], but here, only the exchange-only derivation most clearly complimentary to the preceding work is selected from Krieger *et al.* [21] and discussed.

Forming the potential

$$v_{i\sigma}(\mathbf{r}) = \frac{1}{f_{i\sigma}\psi_{i\sigma}^{KS*}(\mathbf{r})}\frac{\delta E_x}{\delta \psi_{i\sigma}^{KS}(\mathbf{r})} \tag{4.85}$$

then from the total energy, Eq. (4.75), it is possible to form the derivative

$$\begin{aligned}\frac{\delta E[n_e]}{\delta \psi_{i\sigma}^{KS}(\mathrm{r}')} &= f_{i\sigma}\left[-\frac{\nabla^2}{2} + v(\mathrm{r}') + v_H(\mathrm{r}') + v_{i\sigma}(\mathrm{r}')\right]\psi_{i\sigma}^{KS*}(\mathrm{r}') \\ &= f_{i\sigma}\left[\varepsilon_{i\sigma}^{KS} - v_{x\sigma}(\mathrm{r}') + v_{i\sigma}(\mathrm{r}')\right]\psi_{i\sigma}^{KS*}(\mathrm{r}')\end{aligned} \tag{4.86}$$

As such, Eq. (4.78), the derivative of the total energy with respect to the potential $v_{s\sigma}$ becomes

$$\frac{\delta E[n_e]}{\delta v_{s\sigma}(\mathbf{r})} = \sum_i f_{i\sigma} \int d\mathbf{r}' \left(\left[\varepsilon_{i\sigma}^{KS} - v_{x\sigma}(\mathbf{r}') + v_{i\sigma}(\mathbf{r}') \right] G_{i\sigma}(\mathbf{r}', \mathbf{r}) \psi_{i\sigma}^{KS*}(\mathbf{r}') \psi_{i\sigma}^{KS}(\mathbf{r}) + c.c. \right) = 0 \tag{4.87}$$

using Eq. (4.69). This can be simplified further as a consequence of Eq. (4.74), thus

$$\sum_i f_{i\sigma} \int d\mathbf{r}' \left(\left[v_{x\sigma}(\mathbf{r}') - v_{i\sigma}(\mathbf{r}') \right] G_{i\sigma}(\mathbf{r}', \mathbf{r}) \psi_{i\sigma}^{KS*}(\mathbf{r}') \psi_{i\sigma}^{KS}(\mathbf{r}) + c.c. \right) = 0 \tag{4.88}$$

If $\varepsilon_{j\sigma}$ in the definition of the Green's function, Eq. (4.68), is approximated with a mean energy $\bar{\varepsilon}_{i\sigma}$ then

$$\begin{aligned} G_{i\sigma}(\mathbf{r}', \mathbf{r}) &\approx \sum_{j \neq i} \frac{\psi_{j\sigma}^{KS}(\mathbf{r}') \psi_{j\sigma}^{KS*}(\mathbf{r})}{\bar{\varepsilon}_{i\sigma}^{KS} - \varepsilon_i^{KS}} \\ &\approx \frac{\delta(\mathbf{r}' - \mathbf{r}) - \psi_{i\sigma}^{KS}(\mathbf{r}') \psi_{i\sigma}^{KS*}(\mathbf{r})}{\bar{\varepsilon}_{i\sigma}^{KS} - \varepsilon_i^{KS}} \end{aligned} \tag{4.89}$$

Integrating Eq. (4.88) over $\mathbf{r}'$ then yields

$$\sum_i f_{i\sigma} \left[\left(v_{x\sigma}(\mathbf{r}) - v_{i\sigma}(\mathbf{r}) \right) \psi_{i\sigma}^{KS*}(\mathbf{r}) \psi_{i\sigma}^{KS}(\mathbf{r}) - \left(\bar{v}_{x\sigma i} - \bar{v}_{i\sigma} \right) \psi_{i\sigma}^{KS*}(\mathbf{r}) \psi_{i\sigma}^{KS}(\mathbf{r}) \right] = 0 \tag{4.90}$$

where

$$\begin{aligned} \bar{v}_{x\sigma i}(\mathbf{r}) &= \int \psi_{i\sigma}^{KS*}(\mathbf{r}) \psi_{i\sigma}^{KS}(\mathbf{r}) v_{x\sigma}(\mathbf{r}) d\mathbf{r} \\ \bar{v}_{i\sigma}(\mathbf{r}) &= \int \psi_{i\sigma}^{KS*}(\mathbf{r}) \psi_{i\sigma}^{KS}(\mathbf{r}) v_{i\sigma}(\mathbf{r}) d\mathbf{r} \end{aligned} \tag{4.91}$$

The approximate exchange potential, $v_{x\sigma}^{Sl}(\mathbf{r})$, attributed to Slater [88], is written as

$$v_{x\sigma}^{Sl}(\mathbf{r}) = \frac{\sum_i n_{i\sigma}(\mathbf{r}) v_{i\sigma}(\mathbf{r})}{\sum_i n_{i\sigma}(\mathbf{r})} \tag{4.92}$$

allowing Eq. (4.90) to be arranged so that

$$v_{x\sigma}(\mathbf{r}) = v_{x\sigma}^{Sl}(\mathbf{r}) + \frac{\sum_i n_{i\sigma}(\mathbf{r}) \left[\bar{v}_{x\sigma i} - \bar{v}_{i\sigma} \right]}{\sum_i n_{i\sigma}(\mathbf{r})} \tag{4.93}$$

Let m signify the index of the highest occupied orbital i.e. $f_{i\sigma} = 0$ when $i > m$, then realizing that $\psi_{m\sigma}$ will tend to zero as $r \to \infty$ slower than any other orbital, there must be a point R_m such that

$$\left| \frac{f_{i\sigma} \psi_{i\sigma}(\mathbf{r})}{f_{m\sigma} \psi_{m\sigma}(\mathbf{r})} \right| < \delta \qquad i < m \quad \text{and} \quad r > R_m \tag{4.94}$$

where δ is arbitrarily small. Dividing Eq. (4.90) by $f_{m\sigma}\psi_{m\sigma}(\mathbf{r})$ and neglecting terms $O(\delta)$

$$\left[\left(v_{x\sigma}(\mathbf{r})-v_{m\sigma}(\mathbf{r})\right)\psi_{m\sigma}^{KS*}(\mathbf{r})-\left(\bar{v}_{x\sigma m}-\bar{v}_{m\sigma}\right)\psi_{m\sigma}^{KS*}(\mathbf{r})\right]=0 \qquad r>R_m$$

i.e.

$$v_{x\sigma}(\mathbf{r})=v_{m\sigma}(\mathbf{r})+\left(\bar{v}_{x\sigma m}-\bar{v}_{m\sigma}\right) \qquad r>R_m \tag{4.95}$$

Now, Eq. (4.90) is determined only up to a constant – if $v_{x\sigma}$ is a solution then so is $v_{x\sigma}+C$, the eigenfunctions remaining unchanged. In the asymptotic limit,

$$v_{m\sigma}(\mathbf{r})\to-\frac{f_{m\sigma}}{r} \tag{4.96}$$

from Eq. (4.85), and choosing C so that $v_{x\sigma}(\mathbf{r})\to 0$ then

$$\bar{v}_{x\sigma m}=\bar{v}_{m\sigma} \tag{4.97}$$

and

$$v_{x\sigma}(\mathbf{r})\to-\frac{f_{m\sigma}}{r} \tag{4.98}$$

as $r\to\infty$[i]. Therefore

---

[i] Eq. (4.97) and Eq. (4.98) are not just properties of the KLI approximation to the exchange potential, but the exact KS exchange in general. This can be seen from the alternative derivation below. Beginning at the universal Eq. (88) and recalling the argument which led to Eq. (94), then given

$$\sum_i f_{i\sigma}\int d\mathbf{r}'\left(\left[v_{x\sigma}(\mathbf{r}')-v_{i\sigma}(\mathbf{r}')\right]G_{i\sigma}(\mathbf{r}',\mathbf{r})\psi_{i\sigma}^{KS*}(\mathbf{r}')\psi_{i\sigma}^{KS}(\mathbf{r})+c.c.\right)=0$$

dividing throughout by $f_{m\sigma}\psi_{m\sigma}(\mathbf{r})$ and neglecting terms $O(\delta)$, yields

$$\int d\mathbf{r}'\left[v_{x\sigma}(\mathbf{r}')-v_{m\sigma}(\mathbf{r}')\right]G_{m\sigma}(\mathbf{r}',\mathbf{r})\psi_{m\sigma}^{KS*}(\mathbf{r}')=0$$

for $r>R_{m\sigma}$. Applying the operator $\varepsilon_{m\sigma}-h_\sigma(\mathbf{r})$ to this expression, $h_\sigma(\mathbf{r})$ defined as in Eq. (4.76), and integrating over $\mathbf{r}'$, then an expression identical to Eq. (4.95) is arrived at:

$$v_{x\sigma}(\mathbf{r})=v_{m\sigma}(\mathbf{r})+\left(\bar{v}_{x\sigma m}-\bar{v}_{m\sigma}\right) \qquad r>R_m \tag{*}$$

using

$$\left(\varepsilon_{m\sigma}-h_\sigma(\mathbf{r})\right)G_{m\sigma}(\mathbf{r}',\mathbf{r})=-\sum_{i\neq m}\psi_{i\sigma}^{KS}(\mathbf{r}')\psi_{i\sigma}^{KS*}(\mathbf{r})$$
$$=-\delta(\mathbf{r}'-\mathbf{r})+\psi_{m\sigma}^{KS}(\mathbf{r}')\psi_{m\sigma}^{KS*}(\mathbf{r})$$

As before, $v_{x\sigma}$ is determined only up to a constant i.e. both $v_{x\sigma}$ and $v_{x\sigma}+C$ are solutions of (*), then C can be chosen so that $v_{x\sigma}(\mathbf{r})\to 0$ as $r\to\infty$. In the same limit $v_{m\sigma}(\mathbf{r})\to -f_{m\sigma}/r$, from Eq. (4.85), and (*) reduces to

$$\bar{v}_{x\sigma m}=\bar{v}_{m\sigma}$$

or can be used to give

$$v_{x\sigma}(\mathbf{r})\to-\frac{f_{m\sigma}}{r}$$

confirming the earlier claims that Eq. (4.97) and Eq. (4.98) are properties of the Kohn-Sham exact exchange.

$$v_{x\sigma}(\mathbf{r}) = v_{x\sigma}^{SI}(\mathbf{r}) + \frac{\sum_{i}^{m-1} n_{i\sigma}(\mathbf{r})[\bar{v}_{x\sigma i} - \bar{v}_{i\sigma}]}{\sum_{i} n_{i\sigma}(\mathbf{r})} \tag{4.99}$$

To use this expression practically, then the set of constants $\bar{v}_{x\sigma i}$, i extending from 1 to m-1 has to first be ascertained. Multiplying Eq. (4.99) by $n_{j\sigma}(\mathbf{r})$ and integrating over $\mathbf{r}$ yields

$$f_{j\sigma}\bar{v}_{x\sigma j} = f_{j\sigma}\bar{v}_{x\sigma j}^{SI} + \sum_{i=1}^{m-1}\int \frac{n_{j\sigma}(\mathbf{r})n_{i\sigma}(\mathbf{r})}{n_{\sigma}(\mathbf{r})} d\mathbf{r}[\bar{v}_{x\sigma i} - \bar{v}_{i\sigma}] \qquad j = 1,2,\ldots,m-1 \tag{4.100}$$

then on defining

$$M_{ij}^{\sigma} = \int \frac{n_{j\sigma}(\mathbf{r})n_{i\sigma}(\mathbf{r})}{n_{\sigma}(\mathbf{r})} d\mathbf{r} \qquad i,j = 1,2,\ldots,m-1 \tag{4.101}$$

$\bar{v}_{x\sigma i}$ can be obtained from solving the system of linear equations

$$\bar{v}_{x\sigma j}^{SI} - \bar{v}_{j\sigma} = \sum_{i=1}^{m-1}\left(\delta_{ij} - f_{j\sigma}^{-1}M_{ji}^{\sigma}\right)[\bar{v}_{x\sigma i} - \bar{v}_{i\sigma}] \tag{4.102}$$

The actual methods of numerical implementation of the above are discussed in Section 4.4, where due to the nature of the semiconductor nanocrystal under investigation, some simplifications can be introduced and the KLI exchange potential can be shown to be spherically symmetric in these circumstances and so compatible with the computational methods developed in the preceding Chapters. Although at first the task of solving Eq. (4.99) appears to be considerably more complex than simply solving the OEP integral equation, it really does provide a fast and effective means of evaluating the exchange potential.

4.2.3 Evaluation of the Exchange Models

Extensive discussions and comparisons have been made within the literature between the various methodologies and approximations in calculating the potentials of atoms within the exchange only limit, see for instance Ref.'s [21,81,85,89].

Engel and Dreizler [81] provide OEP ground-state energies satisfying the Levy-Perdew theorem [90]

$$E_x[\psi_{i\sigma}^{KS}] = -\int n(\mathbf{r})\mathbf{r}\cdot\nabla v_x(\mathbf{r})d\mathbf{r} \tag{4.103}$$

to within 0.14 mH (0.0038 eV) up to Radon (^{86}Rn, E = 21 866.75 H (~595.03 keV) E_x = 387.45 H (~10.54 keV)) for closed-subshell atoms. With these as reference values, the KLI ground-state

energies are found to deviate between ~0 to a maximum of 10 mH (~0 to 0.27 eV) while the LDA values lie a substantial ~138 to 14 424 mH (~3.76 to 392.50 eV) adrift. Compared to non-local HF, the OEP values rest between 0 and 40 mH (~0 to 1.09 eV) higher, corresponding to the difference between minimising the same energetic expression to produce a common potential as opposed to minimising it for a set of orbital-dependent potentials.

Krieger, Li and Iafrate [21] report that direct calculation of the highest occupied atomic eigenvalues, ε_m, are accurate to within 0.4% of the OEP results and show that the Slater and LSD (spin polarised LDA) potentials over and underestimate ε_m to ~10% and ~40% respectively.

Clearly then, the KLI method is substantially better than comparable approximate methods and compares very favourably with highly precise OEP results for a fraction of the numerical cost.

In order to assess the accuracy of the actual implementation of the preceding equations by the author before progressing onto calculating the band structure of the SnO_2 grains, it is sensible to test against known data. Therefore, on a 700-point logarithmic mesh, a selection of exchange potentials for Neon (^{10}Ne) have been plotted in Figure 4-1, and likewise for Cadmium (^{48}Cd) in Figure 4-2, appearing to recreate Figures 1 and 2 of Ref. [21] exactly.

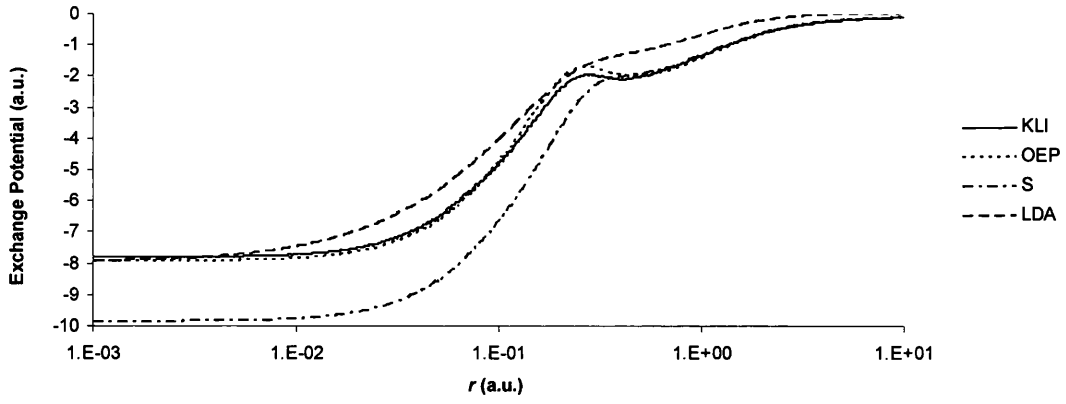


Figure 4-1 Plot of exchange potential, in Hartree, versus the radial distance from the nucleus, in Bohr, for Neon using the KLI approximation (KLI), the optimised effective potential method (OEP), the Slater potential of Eq. (4.92) (S), and the local density approximation (LDA).

The KLI exchange potential is given by Eq. (4.99), the OEP potential satisfies Eq. (4.72), the Slater potential follows from Eq. (4.92), and lastly, the local density exchange potential is derived from Eq. (4.58).

Whilst not quite achieving the very high accuracy of Ref. [81] on their 1600 point mesh, the KLI ground state energies of ^{10}Ne and ^{48}Cd calculated here, do lie within an acceptable ~0.02% and ~0.05% of their reference OEP values, 128.54(54) H (~3.49 keV) and 5 465.11(44) H (~0.15 MeV) respectively. The highest occupied eigenvalues differ from Ref. [21]'s KLI -0.8494 H

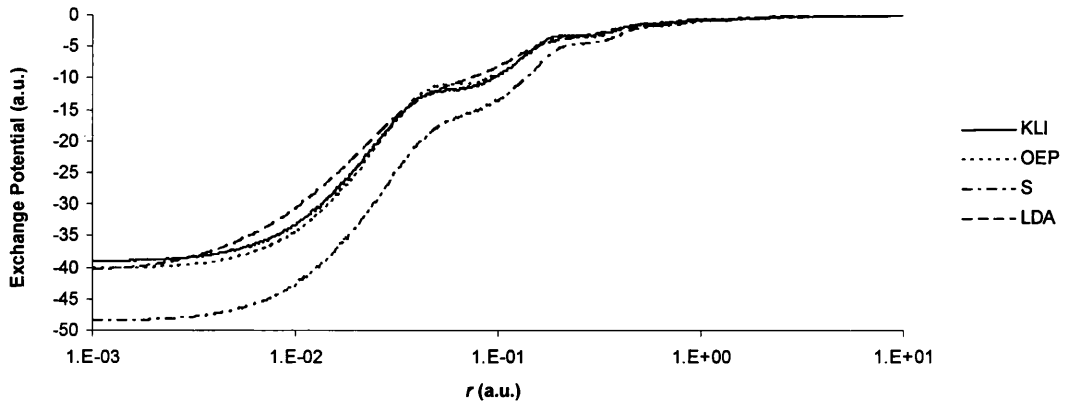


Figure 4-2 Plot of exchange potential, in Hartree, versus the radial distance from the nucleus, in Bohr, for Cadmium using the KLI approximation (KLI), the optimised effective potential method (OEP), the Slater potential of Eq. (4.92) (S), and the local density approximation (LDA).

(-23.11 eV) for ^{10}Ne, and -0.2651 H (-7.21 eV) for ^{48}Cd, by the slightly higher percentages of ~0.74% and ~0.9% respectively - still an adequate tolerance, especially given the sensitivity of the atomic energy levels to small fluctuations in the Coulomb and exchange potentials.

In principle then, the KLI method has been amply shown to be satisfactory for the purposes of calculating the exchange potential within this document, and actual numerical implementation of the procedure favourably reproduces the results of literature.

4.3 Correlation Within the Random Phase Approximation for a Uniform Electron-Gas

The correlation potential is calculated using the parameterised random phase (RPA) uniform electron gas approximation of Perdew and Wang [22]. They use high and low electron density expansions for analytic expressions of the correlation energy fitted to the Green's function Monte Carlo results of Cerpley and Alder [82]. The correlation potential follows from the functional derivative of the correlation energy. In the quantum dot system under study here, this method appears to very adequately allow for the very small magnitude effects of electron correlation.

For the uniform electron gas, the density parameter, r_s, and the relative spin polarisation, ζ (encountered previously in Section 1.1, Eq. (4.59)), are defined as

$$r_s = \left(\frac{3}{4\pi (n_\downarrow + n_\uparrow)} \right)^{1/3}$$
$$\zeta = \frac{n_\uparrow - n_\downarrow}{n_\downarrow + n_\uparrow} \tag{4.104}$$

with the correlation energy per electron, ε_c, expressed as a function of these in both a high density expansion,

$$\varepsilon_c(r_s,\zeta) = c_0(\zeta)\ln r_s - c_1(\zeta) + c_2(\zeta) r_s \ln r_s - c_3(\zeta) r_s + \ldots\ldots \tag{4.105}$$

and a low density expansion

$$\varepsilon_c(r_s,\zeta) = \frac{-d_0(\zeta)}{r_s^p} + \frac{d_1(\zeta)}{r_s^{2p-1/2}} + \ldots\ldots \tag{4.106}$$

In the random phase approximation (RPA)[i] the correct exponent of the density parameter is p = 3/4 (improving on the p = 1 of alternative schemes [91,92]) and on fitting to Cerpley and Alder data which is exact, the parameters of Table 4-1 follow.

---

[i] Discussed in more detail in Chapter 7, the random phase approximation (RPA) relates to the effect of a charge, such as an electron, on its surrounding space. In essence, the screened coulomb potential is the electrostatic potential due to an electron and its associated polarisation cloud. This is related to the bare coulomb potential of the unscreened electron via the dielectric function. In the random phase approximation, this dielectric function is calculated using the lowest order term of the polarisation propagator only i.e. the polarisation propagator between two space-time points, 1 and 2, is based solely on the Green's function propagators $G(1,2)$ and $G(2,1)$ and neglects further contributions including the screened coulomb interactions between the propagators themselves.

Table 4-1 Parameters for the density expansion approximations of correlation energy. Energies in Hartree

	$\varepsilon_c(r_s, 0)$	$\varepsilon_c(r_s, 1)$
c_0	0.031091	0.015545
c_1	0.070823	0.049778
c_2	0.00256	0.00055
c_3	0.00936	0.0023
d_0	0.4001	0.4296
d_1	0.4590	0.7918

To link between the $\zeta = 0$ and $\zeta = 1$ cases, the spin interpolation formula first proposed by Vosko, Wilk and Nusair [91] is used:

$$\varepsilon_c(r_s,\zeta) = \varepsilon_c(r_s,0) + \alpha_c(r_s)\frac{f(\zeta)}{f''(0)}\left(1-\zeta^4\right) + \left(\varepsilon_c(r_s,1) - \varepsilon_c(r_s,0)\right)f(\zeta)\zeta^4$$
$$f(\zeta) = \frac{(1+\zeta)^{4/3} + (1-\zeta)^{4/3} - 2}{2^{4/3} - 2} \tag{4.107}$$

where clearly $f''(0)$=1.709921. The spin stiffness, $\alpha_c(r_s)$, is defined as $\partial^2\varepsilon_c(r_s,\zeta)/\partial\zeta^2$ evaluated at $\zeta = 0$, and is approximated by

$$\alpha_c(r_s) = f''(0)\left(\varepsilon_c(r_s,1) - \varepsilon_c(r_s,0)\right) \tag{4.108}$$

The correlation potential for electrons of spin σ then follows from the derivative

$$v_{corr}^{\sigma}(r_s,\zeta) = \frac{\partial(n_e\varepsilon_c)}{\partial n_\sigma}$$
$$= \varepsilon_c(r_s,\zeta) - \frac{r_s}{3}\frac{\partial\varepsilon_c(r_s,\zeta)}{\partial r_s} - (\zeta - \operatorname{sgn}\sigma)\frac{\partial\varepsilon_c(r_s,\zeta)}{\partial\zeta} \tag{4.109}$$

where

$$\operatorname{sgn}\sigma = \begin{cases} +1 & \sigma = \uparrow \\ -1 & \sigma = \downarrow \end{cases} \tag{4.110}$$

and

$$\frac{\partial\varepsilon_c(r_s,\zeta)}{\partial r_s} = \left(1 - f(\zeta)\zeta^4\right)\frac{\partial\varepsilon_c(r_s,0)}{\partial r_s} + f(\zeta)\zeta^4\frac{\partial\varepsilon_c(r_s,1)}{\partial r_s} + \left(1-\zeta^4\right)\frac{f(\zeta)}{f''(0)}\frac{d\alpha_c(r_s)}{dr_s}$$
$$\frac{\partial\varepsilon_c(r_s,\zeta)}{\partial\zeta} = 4f(\zeta)\zeta^3\left(\varepsilon_c(r_s,0) - \varepsilon_c(r_s,1) - \frac{\alpha_c(r_s)}{f''(0)}\right) \tag{4.111}$$
$$- f'(\zeta)\left(\zeta^4\varepsilon_c(r_s,0) - \zeta^4\varepsilon_c(r_s,1) - \left(1-\zeta^4\right)\frac{\alpha_c(r_s)}{f''(0)}\right)$$

In the general case of the SnO_2 grain relevant here, excepting Figure 4-5, the electron spin densities are taken to be equal and therefore spin-unpolarised i.e. ζ is equal to 0, allowing the correlation potential to be simply approximated as:

$$v_{corr}^{hd}(r_s) = c_0(0)\left(\ln r_s - \frac{1}{3}\right) - c_1(0) + \frac{c_2(0)r_s}{3}(2\ln r_s - 1) - \frac{2r_s c_3(0)}{3}$$
$$v_{corr}^{ld}(r_s) = \frac{-5d_0(0)}{4r_s^{3/4}} + \frac{4d_1(0)}{3r_s} \qquad (4.112)$$

These two forms are equal at $r_s \approx 25.84$ (roughly 25 to 26 stored electrons) and so are both required to span the total range of electron densities involved in modelling charge storage.

As with all local density approximations, the key idea is to consider the general inhomogeneous electronic system as locally homogeneous at each specific spatial point, so for example, the electron density $n_e(\mathbf{r})$ is taken to be uniform at $\mathbf{r}$. Thus, in Eq. (4.112) the density parameter r_s is actually dependent on $\mathbf{r}$ and as such, so are the correlation potentials.

4.4 Numerical Implementation

The computation requirements of calculating the Coulomb, KLI exchange and RPA correlation potentials in three dimensions can be reduced greatly by exploiting the spherical symmetry of the system under study and by modelling the electronic spectrum with closed sub-shells, summing over the azimuthal quantum number m for given l.

This second quality is not an approximation as such, since within the limits of the assumptions of the model, the energy eigenvalues, and therefore the thermal fractional orbital occupancies, are independent of m. In other words, thermal filling occurs evenly across degenerate orbitals, and even the direct addition of individual electrons via quantum mechanical tunnelling during charge writing, has no direct effect on this, since at non zero temperatures, maintaining an electronic equilibrium, the added electron is just absorbed into the electron sea already present in the conduction band. Its charge is felt, and in conjunction with the techniques of Section 5.6, the Fermi level is seen to move, corresponding to an increase in the electron population in the partially filled and empty energy levels, but its actual physical presence must be treated as being 'smeared' over the complete spectrum of occupied states.

Importantly, in order to separate the wave functions into their radial and angular components and so utilise the techniques of Chapter 3 to actually compute the eigenstates of the effective potential, then that potential must be spherically symmetric; that is, a function of r only. Therefore, the electron-electron interaction potentials must also be shown to be solely radially dependant.

In order to achieve these goals as expeditely as possible, the first part of this section analyses some of the common individual components of the three potentials before proceeding to the second part of the section where the specifics of each case are treated. The section concludes with a selection of graphs of the potentials, formed on a homogeneous SnO_2 quantum dot system for a variety of integer occupancies, placing these electron-electron interaction potentials into practise with the self-consistent methods developed in Chapter 3.

To begin then, take the frequently encountered term

$$\hat{v}_2(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|}$$

This is, of course, the reciprocal of the magnitude of the vector $\mathbf{r}_{12}$ in Figure 4-3

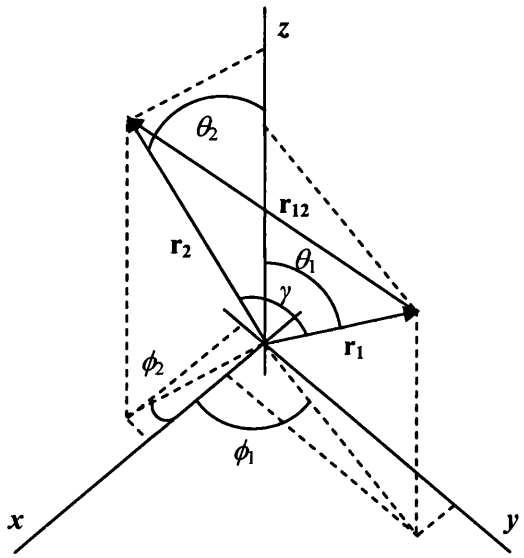


Figure 4-3 Representation of the two vectors $\mathbf{r}_1$ and $\mathbf{r}_2$, connected by the vector $\mathbf{r}_{12}$, the magnitude of which is $|\mathbf{r}_1-\mathbf{r}_2|$.

which can be expressed in terms of the angle γ between the two vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ with the aid of the cosine formula of elementary geometry such that

$$\begin{aligned}\frac{1}{|\mathbf{r}_1-\mathbf{r}_2|} &= \left(r_1^2+r_2^2-2r_1r_2\cos\gamma\right)^{-1/2} \\ &= \frac{1}{r_>}\left(1+\frac{r_<^2}{r_>^2}-2\frac{r_<}{r_>}\cos\gamma\right)^{-1/2} \\ &= \frac{1}{r_>}\sum_{n=0}^{\infty}\left(\frac{r_<}{r_>}\right)^n P_n(\cos\gamma)\end{aligned} \tag{4.113}$$

with

$$\left.\begin{aligned}r_> &= r_1 \\ r_< &= r_2\end{aligned}\right\} r_1 > r_2 \qquad \left.\begin{aligned}r_> &= r_2 \\ r_< &= r_1\end{aligned}\right\} r_2 > r_1 \tag{4.114}$$

and where the last line of Eq. (4.113) follows from expanding the bracketed term of the line above in a binomial series then collating powers of $(r_</r_>)$, the Legendre function $P_n(\cos\gamma)$ (recall Section 3.1) defined as the coefficient of the nth power.

From the addition theorem for spherical harmonics (see Appendix III, The Addition Theorem), the Legendre function can be expanded

$$P_n(\cos\gamma) = \frac{4\pi}{2n+1}\sum_{m=-n}^{n} Y_n^m(\theta_1,\phi_1)Y_n^{m*}(\theta_2,\phi_2) \tag{4.115}$$

allowing $1/|\mathbf{r}_1-\mathbf{r}_2|$ to be written terms of Y_m^n

$$\frac{1}{|\mathbf{r}_1-\mathbf{r}_2|} = \sum_{n=0}^{\infty}\frac{4\pi}{2n+1}\frac{r_<^n}{r_>^{n+1}}\sum_{m=-n}^{n} Y_n^m(\theta_1,\phi_1)Y_n^{m*}(\theta_2,\phi_2) \tag{4.116}$$

Usefully, if γ is equal to zero i.e. (θ_1, ϕ_1) is equal to (θ_2, ϕ_2), then $P_n(1) = 1$, and the product of a spherical harmonic and its complex conjugate summed over all m is simply

$$\sum_{m=-n}^{n} Y_n^m(\theta,\phi)Y_n^{m*}(\theta,\phi) = \sum_{m=-n}^{n}\left|Y_n^m(\theta,\phi)\right|^2 = \frac{2n+1}{4\pi} \tag{4.117}$$

For the product of two different spherical harmonics, then as seen in Section 3.1 Eq. (3.44), their orthogonality with respect to their quantum numbers requires that

$$\int Y_{n_1}^{m_1}{}^*(\theta,\phi)Y_{n_2}^{m_2}(\theta,\phi)d\Omega = \delta_{n_1,n_2}\delta_{m_1,m_2}$$

where

$$d\Omega = \sin\theta d\theta d\phi$$

Setting n_2 and m_2 to zero in Eq. (3.44) yields an expression for the solid angle integral of one spherical harmonic

$$\int Y_n^m(\theta,\phi)d\Omega = \sqrt{4\pi}\delta_{n,0}\delta_{m,0} \tag{4.118}$$

The integral of the product of three spherical harmonics, often written $\langle l_1 m_1 | Y_{l_2}^{m_2} | l_2 m_2 \rangle$ where

$$\langle l_1 m_1 | Y_{l_2}^{m_2} | l_2 m_2 \rangle = \int Y_{l_1}^{m_1*}(\theta,\phi)Y_{l_2}^{m_2}(\theta,\phi)Y_{l_3}^{m_3}(\theta,\phi)d\Omega \tag{4.119}$$

is something which has received a great deal of attention due to its frequent appearances in quantum mechanical problems and has undergone extensive analytical analysis. While of course it is possible to analytically or numerically integrate the product after determining the explicit forms of the spherical harmonics, it is simpler and usually more accurate (if integrating numerically), to either utilise ready tabulated data for common orbital combinations or the Gaunt formula, where the integral is expressed in terms of 3*j* symbols [70]:

$$\langle l_1 m_1 | Y_{l_2}^{m_2} | l_2 m_2 \rangle = (-1)^{m_1} \sqrt{\frac{(2l_1+1)(2l_2+1)(2l_3+1)}{4\pi}} \begin{pmatrix} l_1 & l_2 & l_3 \\ -m_1 & m_2 & m_3 \end{pmatrix} \begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix} \tag{4.120}$$

These coefficients are closely related to the common Clebsch-Gordan (CB) coefficients, such that

$$\begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = \frac{(-1)^{l_1 - l_2 - m_3}}{\sqrt{2l_3+1}} \langle l_1 l_2 m_1 m_2 | l_1 l_2 l_3 - m_3 \rangle \tag{4.121}$$

where

$$\langle l_1 l_2 m_1 m_2 | l_1 l_2 l_3 m_3 \rangle = \delta(m_3, m_1 + m_2) \sqrt{\frac{(l_1+l_2-l_3)!(l_3+l_1-l_2)!(l_3+l_2-l_1)!(2l_3+1)}{(l_1+l_1+l_2+1)!}}$$
$$\times \sum_k \frac{(-1)^k \sqrt{(l_1+m_1)!(l_1-m_1)!(l_2+m_2)!(l_2-m_2)!(l_3+m_3)!(l_3-m_3)!}}{k!(l_1+l_2-l_3-k)!(l_1-m_1-k)!(l_2+m_2-k)!(l_3-l_2+m_1+k)!(l_3-l_1-m_2+k)!} \tag{4.122}$$

The 3*j* symbols have several useful qualities [70], and in particular

$$\begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = \begin{pmatrix} l_3 & l_1 & l_2 \\ m_3 & m_1 & m_2 \end{pmatrix} = \begin{pmatrix} l_2 & l_3 & l_1 \\ m_2 & m_3 & m_1 \end{pmatrix} \tag{4.123}$$
$$(-1)^{l_1+l_2+l_3} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = \begin{pmatrix} l_2 & l_1 & l_3 \\ m_2 & m_1 & m_3 \end{pmatrix} = \begin{pmatrix} l_1 & l_3 & l_2 \\ m_1 & m_3 & m_2 \end{pmatrix} = \begin{pmatrix} l_1 & l_2 & l_3 \\ -m_1 & -m_2 & -m_3 \end{pmatrix}$$

$$\sum_{m_1 m_2} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \begin{pmatrix} l_1 & l_2 & l_4 \\ m_1 & m_2 & m_4 \end{pmatrix} = \frac{\delta_{l_1 l_4} \delta_{m_1 m_4}}{2l_3+1}$$
$$\sum_{l_3 m_3} (2l_3+1) \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_4 & m_5 & m_3 \end{pmatrix} = \delta_{m_1 m_4} \delta_{m_2 m_5} \tag{4.124}$$

$$\begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = 0 \quad \text{unless} \quad \begin{matrix} m_1 + m_2 + m_3 = 0 \\ \Delta(l_1 l_2 l_3) \end{matrix}$$
$$\begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix} = 0 \quad \text{if } l_1 + l_2 + l_3 \text{ is odd} \tag{4.125}$$

The symbol $\Delta(l_1 l_2 l_3)$ is known as the triangle condition, see Figure 4-4, and is the requirement that the vector sum of the angular momenta is zero i.e. $\mathbf{L}_3 = \mathbf{L}_1 + \mathbf{L}_2$. This can expressed as

$$\begin{matrix} l_1 + l_2 + l_3 \text{ is an integer} \\ |l_1 - l_2| \le l_3 \le l_1 + l_2 \end{matrix} \tag{4.126}$$

as discussed in Arfken and Weber [65]. This means that the integral, Eq. (4.120), will vanish unless

$$\begin{matrix} l_1 + l_2 + l_3 \text{ is an even integer} \\ |l_1 - l_2| \le l_3 \le l_1 + l_2 \\ m_2 + m_3 = m_1 \end{matrix} \tag{4.127}$$

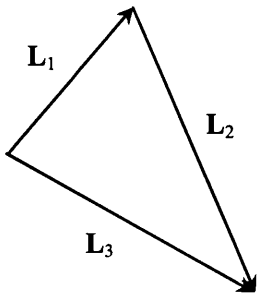


Figure 4-4 Schematic representation of the triangle condition $\Delta(l_1 l_2 l_3)$ i.e. the vector sum of the angular momenta is zero

Furthermore, when $\Delta(l_1 l_2 l_3)$ is satisfied, the identities

$$Y_{l_1}^{m_1*}(\theta,\phi)Y_{l_2}^{m_2}(\theta,\phi) = \sum_{LM}(-1)^{m_2}\sqrt{\frac{(2l_1+1)(2l_2+1)(2L+1)}{4\pi}} \times \begin{pmatrix} l_1 & l_2 & L \\ m_1 & -m_2 & M \end{pmatrix}\begin{pmatrix} l_1 & l_2 & L \\ 0 & 0 & 0 \end{pmatrix} Y_L^M(\theta,\phi) \tag{4.128}$$

$$Y_{l_1}^{m_1}(\theta,\phi)Y_{l_2}^{m_2}(\theta,\phi) = \sum_{LM}\sqrt{\frac{(2l_1+1)(2l_2+1)(2L+1)}{4\pi}}\begin{pmatrix} l_1 & l_2 & L \\ m_1 & m_2 & M \end{pmatrix}\begin{pmatrix} l_1 & l_2 & L \\ 0 & 0 & 0 \end{pmatrix} Y_L^{M*}(\theta,\phi)$$

also hold [70].

Applying all these relations to the electron-electron interaction potentials, then considerable simplifications can be made.

4.4.1 Implementing the Coulomb Potential

For the Coulomb potential of Eq. (4.53)

$$v_H([n_e];\mathbf{r}_2) = \int \frac{n_e(\mathbf{r}_1)}{|\mathbf{r}_2 - \mathbf{r}_1|} d\mathbf{r}_1$$

the total electronic density is given by Eq. (4.47)

$$n_e(\mathbf{r}_1) = \sum_i f_i |\psi_i(\mathbf{r}_1)|^2$$

which in spherical polar coordinates can be expressed as

$$n_e(r_1,\theta_1,\varphi_1) = \sum_{nlm} f_{nl}|R_{nl}(r_1)|^2 |Y_l^m(\theta_1,\phi_1)|^2 \tag{4.129}$$

The radial part of the wave function written temporarily with the notation R_{nl} to avoid confusion with the fractional occupancy factor f_{nl}, and the subscript e has been reintroduced on n to denote the total electron density and so avoiding confusion with the index n. Then, using Eq. (4.117), the

sum over all m of the product of the spherical harmonic and its complex conjugate can be reduced to a single term dependent on l, removing the angular dependence of n_e

$$n_e(r_1) = \sum_{nl} f_{nl} \frac{2l+1}{4\pi} |R_{nl}(r_1)|^2 \tag{4.130}$$

recall Eq. (3.130).

Next, the $1/|\mathbf{r}_1\text{-}\mathbf{r}_2|$ term in Eq. (4.53) can be expanded in terms of (θ_1, ϕ_1) and (θ_2, ϕ_2) dependant spherical harmonics such that

$$v_H(r_2,\theta_2,\phi_2) = \int\int \sum_{nl} f_{nl} \frac{2l+1}{4\pi} |R_{nl}(r_1)|^2 \sum_{k=0}^{\infty} \frac{4\pi}{2k+1} \frac{r_<^k}{r_>^{k+1}} \sum_{m=-k}^{k} Y_k^m(\theta_1,\phi_1) Y_k^{m*}(\theta_2,\phi_2) r_1^2 dr_1 d\Omega_1 \tag{4.131}$$

Then using Eq. (4.118), the angular Ω_1 integral can be performed, and employing the resultant Kronecker delta functions within the summations to remove the (θ_2, ϕ_2) angular dependence, the Coulomb potential reduces to

$$v_H(r_2) = \int \sum_{nl} f_{nl} \frac{2l+1}{\sqrt{4\pi}} \frac{|R_{nl}(r_1)|^2}{r_>} r_1^2 dr_1 \tag{4.132}$$

and, as such, the requirement that v_H be spherically symmetric has been met.

To perform the radial integral, the integration interval is simply divided into two, so that

$$v_H(r_2) = \int_0^{r_2} \sum_{nl} f_{nl} \frac{2l+1}{\sqrt{4\pi}} \frac{|R_{nl}(r_1)|^2}{r_2} r_1^2 dr_1 + \int_{r_2}^{\infty} \sum_{nl} f_{nl} \frac{2l+1}{\sqrt{4\pi}} \frac{|R_{nl}(r_1)|^2}{r_1} r_1^2 dr_1 \tag{4.133}$$

Naturally, in practise these continuous potentials, radii and so forth are mapped onto a uniform or variable mesh as detailed in Chapter 3. Over a finite discrete grid of N spatial points, the integral of a continuous function can be estimated in a number of ways. Here, the composite Simpson's formula [50] is used

$$\int_{\alpha}^{\alpha+2M\Delta} g(x)dx \approx \frac{\Delta}{3}\big(g(x_\alpha) + 4g(x_{\alpha+1}) + 2g(x_{\alpha+2}) + 4g(x_{\alpha+3}) + \cdots + 4g(x_{\alpha+2M-1}) + g(x_{\alpha+2M})\big) + O(\Delta^4) \tag{4.134}$$

over each even sub-interval of equal mesh spacing, extending from point x_α to point $x_{\alpha+2M}$, and where each discrete value $g(x_{\alpha+\beta})$ is exactly equal to the continuous $g(x)$ evaluated at $x = \alpha + \beta\Delta$. If the interval is odd, $2M + 1$, then Eq. (4.134) can be used to estimate the integral between points

α and $\alpha + 2M$ while the integral from point $\alpha + 2M$ to point $\alpha + 2M + 1$ can be calculated using an expression generated by Hamming's direct method [51], analogous to Simpson's half-formula:

$$\int_{2M\Delta}^{(2M+1)\Delta} g(x)dx \approx \frac{\Delta}{12}\left(-g(x_{2M-1}) + 8g(x_{2M}) + 5g(x_{2M+1})\right) + O(\Delta^3) \tag{4.135}$$

As discussed in Chapter 3, the spatial mesh simulating the quantum dot system under study is usually made up of several predefined inter-mesh spacings, with the only constraint that for the total extent of the spatial grid, N is set large enough that $R_{nl}(r_N)$ is approximately equal to zero.

4.4.2 Implementing the Exchange Potential

The KLI exchange potential is given by Eq. (4.99),

$$v_{x\sigma}([n_{i\sigma}];\mathbf{r}) = \frac{\sum_i n_{i\sigma}(\mathbf{r})v_{i\sigma}(\mathbf{r})}{\sum_i n_{i\sigma}(\mathbf{r})} + \frac{\sum_i^{h-1} n_{i\sigma}(\mathbf{r})[\bar{v}_{x\sigma i} - \bar{v}_{i\sigma}]}{\sum_i n_{i\sigma}(\mathbf{r})}$$

substituting h for m as the label of the highest occupied eigenstate to avoid confusion with the symbol used for the azimuthal quantum number.

The degeneracy of each eigenstate of given n and l in the spherical space of the problem implies that the occupancy of spin up and spin down orbitals of equal energy will be the same. Ergo, $n\uparrow = n\downarrow$ and as such the spin up and spin down exchange potentials will be identical, so it is sufficient to only calculate one potential.

Thus to begin, identify the common denominator of the two terms as half the total electron density and so independent of angular position via Eq. (4.130), then

$$v_x(r,\theta,\phi) = \frac{\sum_i n_{i\sigma}(\mathbf{r})v_{i\sigma}(\mathbf{r})}{\sum_{nl} f_{nl}\frac{2l+1}{8\pi}|R_{nl}(r)|^2} + \frac{\sum_i^{h-1} n_{i\sigma}(\mathbf{r})[\bar{v}_{x\sigma i} - \bar{v}_{i\sigma}]}{\sum_{nl} f_{nl}\frac{2l+1}{8\pi}|R_{nl}(r)|^2} \tag{4.136}$$

For the sake of clarity, the spin subscripts will only be gradually removed as each factor is individually considered in the proceeding calculations.

For the numerator of the first term, referred to as $T_{1\sigma}$, where $v_{i\sigma}$ can be expressed as

$$v_{i\sigma}(\mathbf{r}_1) = -\sum_j f_{j\sigma}\int \frac{\psi^*_{i\sigma}(\mathbf{r}_2)\psi^*_{j\sigma}(\mathbf{r}_1)\psi_{j\sigma}(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|\psi^*_{i\sigma}(\mathbf{r}_1)}d\mathbf{r}_2 \tag{4.137}$$

from Eq. (4.85), $T_{1\sigma}$ can be written

$$\begin{aligned}
T_{1\sigma}(\mathbf{r}_1) &= \sum_i n_{i\sigma}(\mathbf{r}_1) v_{i\sigma}(\mathbf{r}_1) \\
&= -\sum_i \sum_j f_{i\sigma} f_{j\sigma} \int \frac{\psi_{i\sigma}(\mathbf{r}_1)\psi^*_{i\sigma}(\mathbf{r}_2)\psi^*_{j\sigma}(\mathbf{r}_1)\psi_{j\sigma}(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_2 \\
&= -\frac{1}{4} \sum_{n_i l_i m_i} \sum_{n_j l_j m_j} f_{n_i l_i} f_{n_j l_j} \sum_{LM} \frac{4\pi}{2L+1} G(n_i l_i, n_j l_j; r_1) Y_L^{M*}(\theta_1,\phi_1) Y_{l_i}^{m_i}(\theta_1,\phi_1) Y_{l_j}^{m_j *}(\theta_1,\phi_1) \\
&\qquad \times \int Y_L^M(\theta_2,\phi_2) Y_{l_i}^{m_i *}(\theta_2,\phi_2) Y_{l_j}^{m_j}(\theta_2,\phi_2) d\Omega_2
\end{aligned} \tag{4.138}$$

where

$$G(n_i l_i, n_j l_j; r_1) = R_{n_i l_i}(r_1) R^*_{n_j l_j}(r_1) \int \frac{r_<^L}{r_>^{L+1}} R^*_{n_i l_i}(r_2) R_{n_j l_j}(r_2) r_2^2 dr_2 \tag{4.139}$$

using Eq. (4.116) and the equality of the spin states and their occupancies. The spin subscript is now dropped from $T_{1\sigma}$, and replacing the product $Y_L^{M*} Y_{l_i}^{m_i}$ using Eq. (4.128) and evaluating the angular integral using Eq. (4.120), yields

$$\begin{aligned}
T_1(\mathbf{r}_1) = -\frac{1}{4} \sum_{n_i l_i m_i} \sum_{n_j l_j m_j} f_{n_i l_i} f_{n_j l_j} \sum_{LM} \sum_{L'M'} \frac{4\pi}{2L+1} G(n_i l_i, n_j l_j; r_1) \frac{(2L+1)(2l_j+1)}{4\pi} \\
\times \sqrt{(2L'+1)(2l_j+1)} \begin{pmatrix} L & l_i & L' \\ M & -m_i & M' \end{pmatrix} \begin{pmatrix} L & l_i & l_j \\ M & -m_i & m_j \end{pmatrix} \begin{pmatrix} L & l_i & L' \\ 0 & 0 & 0 \end{pmatrix} \\
\times \begin{pmatrix} L & l_i & l_j \\ 0 & 0 & 0 \end{pmatrix} Y_{L'}^{M'}(\theta_1,\phi_1) Y_{l_j}^{m_j *}(\theta_1,\phi_1)
\end{aligned} \tag{4.140}$$

where the orders of elements within the 3*j* symbols can be adjusted using Eq. (4.123).

Then using Eq. (4.124) and finally Eq. (4.117) it is possible to write

$$T_1(\mathbf{r}_1) = -\frac{1}{4} \sum_{n_i l_i} \sum_{n_j l_j} f_{n_i l_i} f_{n_j l_j} \frac{(2l_i+1)(2l_j+1)}{4\pi} \sum_L G(n_i l_i, n_j l_j; r_1) \begin{pmatrix} L & l_i & l_j \\ 0 & 0 & 0 \end{pmatrix}^2 = T_1(r_1) \tag{4.141}$$

demonstrating that the first term, in its entirety, of Eq. (4.136) is independent of the angular coordinates.

The second numerator is a little more complex. As discussed in Section 4.2, the terms in the square parenthesis are determined from the solution of the system of linear equations Eq. (4.102), namely

$$\bar{v}^{SI}_{x\sigma j} - \bar{v}_{j\sigma} = \sum_{i=1}^{m-1} \left(\delta_{ij} - f_{j\sigma}^{-1} \int \frac{n_{j\sigma}(\mathbf{r}) n_{i\sigma}(\mathbf{r})}{n_\sigma(\mathbf{r})} d\mathbf{r} \right) [\bar{v}_{x\sigma i} - \bar{v}_{i\sigma}]$$

The Slater potential, given by the first term of Eq. (4.99), has been shown to have no angular or spin dependence, Eq. (4.130) and Eq. (4.141), and as such its average, $\bar{v}_{x\sigma i}^{Sl} \to \bar{v}_{xi}^{Sl}$, is simply

$$\begin{aligned}\bar{v}_{xi}^{Sl} &= \int \psi_i^*(\mathbf{r})\psi_i(\mathbf{r})v_x^{SL}(\mathbf{r})d\mathbf{r} \\ &= \int R_{n_i l_i}^*(r)R_{n_i l_i}(r)v_x^{SL}(r)r^2 dr \end{aligned} \tag{4.142}$$

exploiting the orthogonality of the spherical harmonics.

For the average of the potential $v_{i\sigma}$

$$\begin{aligned}\bar{v}_{i\sigma} &= \int \psi_{i\sigma}^*(\mathbf{r}_1)\psi_{i\sigma}(\mathbf{r}_1)v_{i\sigma}(\mathbf{r}_1)d\mathbf{r}_1 \\ &= -\sum_j f_{j\sigma}\int \frac{\psi_{i\sigma}(\mathbf{r}_1)\psi_{i\sigma}^*(\mathbf{r}_2)\psi_{j\sigma}^*(\mathbf{r}_1)\psi_{j\sigma}(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}d\mathbf{r}_2 d\mathbf{r}_1 \\ &= -\frac{1}{2}\sum_{n_j l_j m_j} f_{n_j l_j}\sum_{LM}\frac{4\pi}{2L+1}\int G(n_i l_i, n_j l_j; r_1)r_1^2 dr_1 \int Y_L^{M*}(\theta_1,\phi_1)Y_{l_i}^{m_i}(\theta_1,\phi_1)Y_{l_j}^{m_j *}(\theta_1,\phi_1)d\Omega_1 \\ &\qquad \times \int Y_L^M(\theta_2,\phi_2)Y_{l_i}^{m_i *}(\theta_2,\phi_2)Y_{l_j}^{m_j}(\theta_2,\phi_2)d\Omega_2\end{aligned}$$

then dropping the spin subscript and with Eq. (4.120) and exploiting Eq. (4.123)

$$\bar{v}_i = -\frac{1}{2}\sum_{n_j l_j} f_{n_j l_j}(2l_j+1)\sum_L \int G(n_i l_i, n_j l_j; r_1)r_1^2 dr \begin{pmatrix} L & l_i & l_j \\ 0 & 0 & 0 \end{pmatrix}^2 \tag{4.143}$$

The integral of the two spin densities can immediately separated into its radial and angular components such that

$$\frac{1}{f_{j\sigma}}\int \frac{n_{j\sigma}(\mathbf{r})n_{i\sigma}(\mathbf{r})}{n_\sigma(\mathbf{r})}d\mathbf{r} = \frac{1}{2}\int f_{n_i l_i}\frac{|R_{n_j l_j}(r)|^2 |R_{n_i l_i}(r)|^2}{\sum_{n_k l_k} f_{n_k l_k}\frac{2l_k+1}{8\pi}|R_{n_k l_k}(r)|^2}r^2 dr \int |Y_{l_j}^{m_j}(\theta,\phi)|^2 |Y_{l_i}^{m_i}(\theta,\phi)|^2 d\Omega \tag{4.144}$$

Using Eq. (4.128) and the orthogonality of the spherical harmonics, the angular integral can be written

$$\begin{aligned}\int |Y_{l_j}^{m_j}(\theta,\phi)|^2 |Y_{l_i}^{m_i}(\theta,\phi)|^2 d\Omega &= \int \sum_{LL'MM'} \frac{(2l_i+1)(2l_j+1)}{4\pi}\sqrt{(2L+1)(2L'+1)} \\ &\qquad \times \begin{pmatrix} l_i & l_j & L \\ m_i & m_j & M \end{pmatrix}\begin{pmatrix} l_i & l_j & L' \\ m_i & m_j & M' \end{pmatrix}\begin{pmatrix} l_i & l_j & L \\ 0 & 0 & 0 \end{pmatrix} \\ &\qquad \times \begin{pmatrix} l_i & l_j & L' \\ 0 & 0 & 0 \end{pmatrix} Y_L^{M*}(\theta,\phi)Y_{L'}^{M'}(\theta,\phi)d\Omega \\ &= \sum_{LM}\frac{(2l_i+1)(2l_j+1)(2L+1)}{4\pi}\begin{pmatrix} l_i & l_j & L \\ m_i & m_j & M \end{pmatrix}^2 \begin{pmatrix} l_i & l_j & L \\ 0 & 0 & 0 \end{pmatrix}^2\end{aligned} \tag{4.145}$$

Collating the above and initially making the assumption that $[\bar{v}_{xi}-\bar{v}_i]$ retains a dependence on m_i i.e. $[\bar{v}_{xn_il_im_i}-\bar{v}_{n_il_i}]$ as v_i has already be shown to be independent of m_i (Eq. (4.143)), then

$$\sum_{i=1}^{h-1}\left(\delta_{ij}-f_{j\sigma}^{-1}\int\frac{n_{j\sigma}(\mathbf{r})n_{i\sigma}(\mathbf{r})}{n_\sigma(\mathbf{r})}d\mathbf{r}\right)[\bar{v}_{x\sigma i}-\bar{v}_{i\sigma}]$$

$$=[\bar{v}_{xn_jl_jm_j}-\bar{v}_{n_jl_j}]-\sum_{n_il_i}^{h-1}\left(\frac{1}{2}\int f_{n_il_i}\frac{\left|R_{n_jl_j}(r)\right|^2\left|R_{n_il_i}(r)\right|^2}{\sum_{n_kl_k}f_{n_kl_k}\frac{2l_k+1}{8\pi}\left|R_{n_kl_k}(r)\right|^2}r^2dr\right.$$

$$\left.\times\sum_{m_i}\sum_{LM}[\bar{v}_{xn_il_im_i}-\bar{v}_{n_il_i}]\frac{(2l_i+1)(2l_j+1)(2L+1)}{4\pi}\begin{pmatrix}l_i & l_j & L\\ m_i & m_j & M\end{pmatrix}^2\begin{pmatrix}l_i & l_j & L\\ 0 & 0 & 0\end{pmatrix}^2\right)$$

$$=\bar{v}_{xn_jl_j}^{Sl}-\bar{v}_{n_jl_j}$$

implying that a function dependent on m_j is equal to a function which has been demonstrated to be independent of m_j. As this is incorrect, then the initial assumption that $[\bar{v}_{xi}-\bar{v}_i]$ is dependent on m_i must also be incorrect and therefore

$$\sum_{i=1}^{h-1}\left(\delta_{ij}-f_{j\sigma}^{-1}\int\frac{n_{j\sigma}(\mathbf{r})n_{i\sigma}(\mathbf{r})}{n_\sigma(\mathbf{r})}d\mathbf{r}\right)[\bar{v}_{x\sigma i}-\bar{v}_{i\sigma}]$$

$$=[\bar{v}_{xn_jl_j}-\bar{v}_{n_jl_j}]-\sum_{n_il_i}^{h-1}[\bar{v}_{xn_il_i}-\bar{v}_{n_il_i}]\left(\frac{1}{2}\int f_{n_il_i}\frac{\left|R_{n_jl_j}(r)\right|^2\left|R_{n_il_i}(r)\right|^2}{\sum_{n_kl_k}f_{n_kl_k}\frac{2l_k+1}{8\pi}\left|R_{n_kl_k}(r)\right|^2}r^2dr\right.$$

$$\left.\times\sum_{m_i}\sum_{LM}\frac{(2l_i+1)(2l_j+1)(2L+1)}{4\pi}\begin{pmatrix}L & l_i & l_j\\ M & m_i & m_j\end{pmatrix}^2\begin{pmatrix}l_i & l_j & L\\ 0 & 0 & 0\end{pmatrix}^2\right)$$

$$=[\bar{v}_{xn_jl_j}-\bar{v}_{n_jl_j}]-\sum_{n_il_i}^{h-1}[\bar{v}_{xn_il_i}-\bar{v}_{n_il_i}]\frac{(2l_i+1)}{8\pi}\int f_{n_il_i}\frac{\left|R_{n_jl_j}(r)\right|^2\left|R_{n_il_i}(r)\right|^2}{\sum_{n_kl_k}f_{n_kl_k}\frac{2l_k+1}{8\pi}\left|R_{n_kl_k}(r)\right|^2}r^2dr$$

$$=\bar{v}_{xn_jl_j}^{Sl}-\bar{v}_{n_jl_j} \qquad (4.146)$$

again employing Eq. (4.124). There is no inconsistency this time, and to clarify the glut of symbols and indices, $[\bar{v}_{xn_il_i}-\bar{v}_{n_il_i}]$, labelled $K_{n_il_i}$, follows from the solution of the m independent set of equations

$$\sum_{n_il_i}^{h-1}A_{n_jl_jn_il_i}K_{n_il_i}=J_{n_jl_j} \qquad (4.147)$$

where

$$A_{n_jl_jn_il_i} = \delta_{n_in_j}\delta_{l_il_j} - \sum_{n_il_i}^{h-1}\frac{(2l_i+1)}{8\pi}\int f_{n_il_i}\frac{\left|R_{n_jl_j}(r)\right|^2\left|R_{n_il_i}(r)\right|^2}{\sum_{n_kl_k} f_{n_kl_k}\dfrac{2l_k+1}{8\pi}\left|R_{n_kl_k}(r)\right|^2}r^2dr$$

$$J_{n_jl_j} = \bar{v}^{Sl}_{xn_jl_j} - \bar{v}_{n_jl_j} \tag{4.148}$$

Solving Eq. (4.147) is merely the task of solving the matrix-vector problem $AK = J$ (K and J vectors) – one of the standard tasks of linear algebra. As such, any linear algebra package, such as LAPACK [67] for example, should contain appropriate subroutines to solve for K, most achieving this without the explicit, and computationally expensive, formation of the inverse of A. A brief overview of two iterative methods that could, feasibly, also be employed is mentioned however in Appendix II, Introduction to Eigenvalues and Eigenvectors, or for a wider variety of techniques see Ref. [50] for example.

Continuing with the simplifications of v_x, then with $[\bar{v}_{xi} - \bar{v}_i]$ independent of m_i, the numerator of the second term in Eq. (4.136), called $T_{2\sigma}$, becomes

$$\begin{aligned}
T_{2\sigma}(\mathbf{r}) &= \sum_i^{h-1} n_{i\sigma}(\mathbf{r})[\bar{v}_{x\sigma i} - \bar{v}_{i\sigma}] \\
&= \frac{1}{2}\sum_{n_il_i}^{h-1}\left[\bar{v}_{xn_il_i} - \bar{v}_{n_il_i}\right] f_{n_il_i}\left|R_{n_il_i}(r)\right|^2 \sum_{m_i=-l_i}^{l_i}\left|Y_{l_i}^{m_i}(\theta,\phi)\right| \\
&= \frac{1}{2}\sum_{n_il_i}^{h-1}\left[\bar{v}_{xn_il_i} - \bar{v}_{n_il_i}\right] f_{n_il_i}\left|R_{n_il_i}(r)\right|^2 \frac{2l_i+1}{4\pi} \\
&= T_2(r)
\end{aligned} \tag{4.149}$$

independent of spin, m_i, θ and ϕ, and as such

$$v_x(r,\theta,\phi) = \frac{2}{n_e(r)}\left(T_1(r) + T_2(r)\right) = v_x(r) \tag{4.150}$$

indicating that as required, the KLI exchange potential is spherically symmetric, dependent as it is solely on the radial coordinate in these circumstances.

4.4.3 Implementing the Correlation Potential

Finally, for the treatment of the RPA correlation potentials, Eq. (4.112), of Perdew and Wang [22],

$$v^{hd}_{corr}([n_e];\mathbf{r}) = c_0(0)\left(\ln r_s(\mathbf{r}) - \frac{1}{3}\right) - c_1(0) + \frac{c_2(0)r_s(\mathbf{r})}{3}\left(2\ln r_s(\mathbf{r}) - 1\right) - \frac{2r_s(\mathbf{r})c_3(0)}{3}$$

$$v^{ld}_{corr}([n_e];\mathbf{r}) = \frac{-5d_0(0)}{4r_s(\mathbf{r})^{3/4}} + \frac{4d_1(0)}{3r_s(\mathbf{r})}$$

where

$$r_s(\mathbf{r}) = \left(\frac{3}{4\pi n_e(\mathbf{r})} \right)^{1/3} \tag{4.151}$$

and with the coefficients of the density expansions given in Table 4-1, then these are reducible simply with the substitution of Eq. (4.130) into the density parameter r_s :

$$r_s(\mathbf{r}) = \frac{3^{1/3}}{\left(4\pi \sum_{nl} f_{nl} \frac{2l+1}{4\pi} |R_{nl}(r)|^2 \right)^{1/3}} = r_s(r) \tag{4.152}$$

so that $v_{corr}(\mathbf{r}) \to v_{corr}(r)$ for both correlation potentials:

$$\begin{aligned} v_{corr}^{hd}(r) &= c_0(0)\left(\ln r_s(r) - \frac{1}{3} \right) - c_1(0) + \frac{c_2(0) r_s(r)}{3}\left(2 \ln r_s(r) - 1\right) - \frac{2 r_s(r) c_3(0)}{3} \\ v_{corr}^{ld}(r) &= \frac{-5 d_0(0)}{4 r_s(r)^{3/4}} + \frac{4 d_1(0)}{3 r_s(r)} \end{aligned} \tag{4.153}$$

4.4.4 Evaluation of the Numerical Procedures

Should a single electron, and a single electron only, reside within the semiconductor valence band then the Coulomb and exchange-correlation potentials should exactly cancel, as there are no electron-electron interactions since there is only one electron present. In this very particular example, the approximations used in this section are not strictly accurate as the complete cancellation of interaction effects is an effect due to the sole occupancy of one spin orbital only, not the partial occupancies of one spin up and one spin down orbital. For v_{corr}, the relative spin polarisation is no longer zero but ±1, dependent on whether the single electron present lies in a spin up or a spin down state, though from Eq. (4.107) it can be seen that the correlation energy per electron is the same regardless i.e. $\varepsilon_c(r_s,-1) = \varepsilon_c(r_s,1)$. As such, the spin polarised correlation potential expansions are simply

$$\begin{aligned} v_{corr}^{hd}(r,\pm 1) &= c_0(1)\left(\ln r_s(r) - \frac{1}{3} \right) - c_1(1) + \frac{c_2(1) r_s(r)}{3}\left(2 \ln r_s(r) - 1\right) - \frac{2 r_s(r) c_3(1)}{3} \\ v_{corr}^{ld}(r,\pm 1) &= \frac{-5 d_0(1)}{4 r_s(r)^{3/4}} + \frac{4 d_1(1)}{3 r_s(r)} \end{aligned} \tag{4.154}$$

replacing the $\zeta = 0$ expansion coefficients in Eq. (4.153) with those for $\zeta = 1$ from Table 4-1. Consequentially, although the local density based correlation potential will unphysically remain non zero, the KLI potential of Eq. (4.99) will reduce to exactly the negative of the Coulomb potential, spherical symmetry guaranteed as the spherical harmonic of the lowest energy

eigenstate is Y_0^0. However, as indicated in Figure 4-5 , even with the exchange approximated by Eq. (4.150), the exchange and Coulomb potentials cancel to the 11^{th} decimal place.

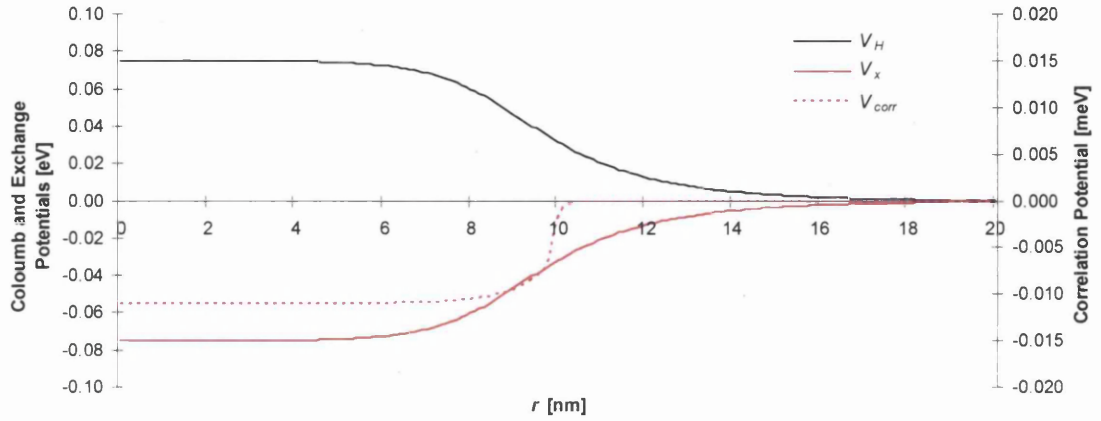


Figure 4-5 Plot of coulomb, exchange and correlation potentials for a single electron in the SnO_2 valence band for a 10nm radius grain. The correlation potential has been calculated using Eq. (4.154) and has been plotted in meV (right hand axis), while the other potentials are plotted in eV (left hand axis).

The material parameters used in Figure 4-5 are those of a sample 10nm radius SnO_2 grain, modifying the values of the atomic units to take into account the new permittivity and electron mass, and for the purposes of illustration, these values been taken to be uniform across the whole system (both inside and outside the grain). The work function was set at 4.53eV and the ionised donor density and surface barrier height were taken at $4.18\times10^{25}m^{-3}$ and 1.4eV respectively.

To demonstrate the effect of an increasing electron population, the electron-electron interaction potentials of integer electron quantities for the same 10nm SnO_2 system as above have been plotted in Figure 4-6 (Coulomb potential given by Eq. (4.132)), Figure 4-7 (KLI exchange potential calculated via Eq. (4.150)) and Figure 4-8 (correlation potential determined using Eq. (4.153)).

While the asymptotic behaviour of the local density correlation potential is incorrect, its magnitude is such that the effect of this will be negligible. For this particular system an electron population of ~70 represents the maximum self-consistent electron compliment.

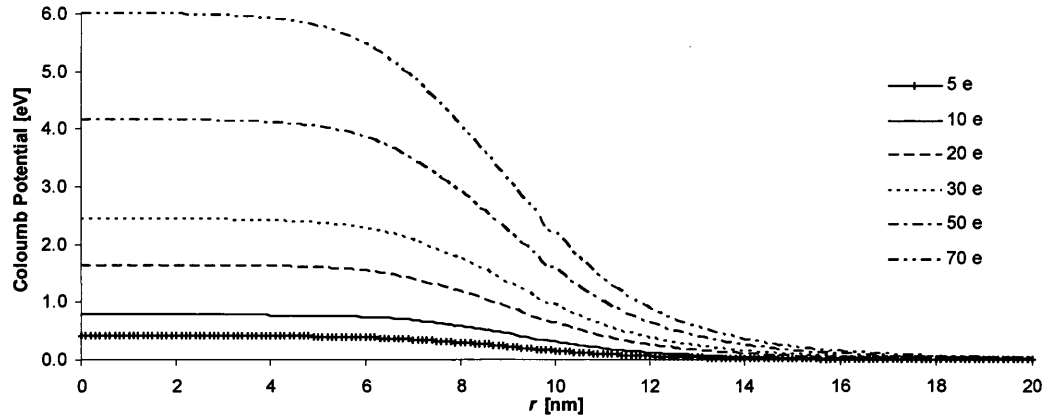


Figure 4-6 Plot of the Coulomb potential for a selection of integer electron populations for a 10nm SnO_2 grain.

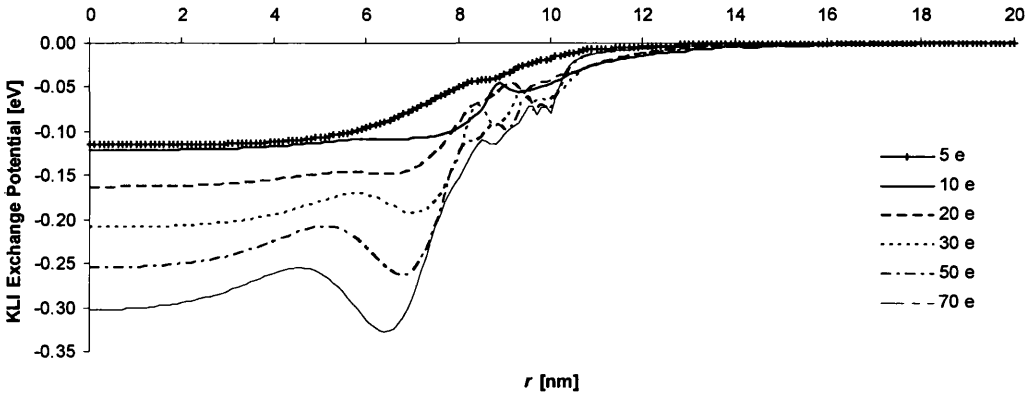


Figure 4-7 Plot of the exchange potential for a selection of integer electron populations for a 10nm SnO_2 grain.

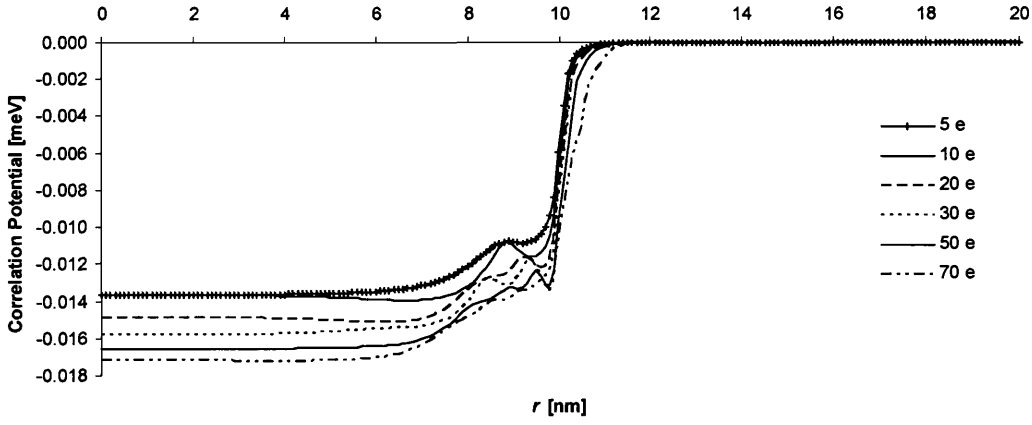


Figure 4-8 Plot of the correlation potential for a selection of integer electron populations for a 10nm SnO_2 grain.

4.5 Conclusions

In this chapter, the difficulties in calculating the electronic properties of a many particle system have been discussed, and via the Hartree and Hartree-Fock approximations, the Kohn-Sham approach has been introduced.

The Kohn-Sham method essentially maps the interacting many bodied problem onto an auxiliary single-particle problem, the ground state density of which is identical to that of the interacting system of electrons. The task of solving this non-interacting Schrödinger equation self-consistently then replaces the much harder problem of finding the many-bodied wave function that minimises the ground state total energy functional.

The particles in the single bodied auxiliary system only interact with the reference potential, not between themselves. This self-consistently determined reference potential includes the background potential from the ionised donor atoms as well as potentials allowing for electron-electron effects. The first of these interaction potentials, and largest in magnitude, is the Coulomb or Hartree potential, which allows for the classical electrostatic repulsion between the electrons and can be calculated exactly. The next potential in order of magnitude is the exchange potential, which accounts for the Pauli exchange principle prohibiting two identical electrons from occupying the same state. The smallest electron-electron effect is the correlation potential. This potential adjusts for the kinetic energy difference between the interacting and non-interacting systems and accounts for the energy reduction as the electrons coordinate their movements to minimise their energy due to their mutual electrostatic repulsion.

The exchange and correlation effects are the most difficult quantities to access, and after some discussion, the Krieger, Li and Iafrate approximation to the exchange potential is outlined and the correlation potential is accounted for via the homogeneous electron gas approximation of Perdew and Wang.

The numerical implementation of these three potentials is outlined, and employing the techniques developed in Chapter 3, some examples are given. These include atomic exchange only simulations, reproducing the results of literature with some considerable success, along with several sample potentials for an illustrative 10nm radius tin dioxide spherical nanocrystal system,

and which include a demonstration of the cancellation of the exchange and Coulomb potentials for one occupied spin orbital.

It can be taken that these methods and their application adequately enable a suffiently accurate simulation of the electron-electron interactions to be made that the way is now free to focus on some of the more subtle aspects of modelling these SnO_2 nanocrystals.

Chapter 5 Gauss' Theorem and the Importance of Boundary Values

This Chapter contains perhaps the most important of the theoretical developments of this work. It focuses on the issues involved in determining a unique potential, or rather unique for a given methodology, which satisfies the non-linear Poisson equation.

It proposes that for small radii, where the non-linear Poisson equation cannot be approximated by its linear cousin, it is not necessarily the case that the same values of the defining parameters R, n_d, E_f, T, Q, m^* and ε corresponding to the barrier height S_b also correspond to a value of zero for the second Poisson equation boundary value, the derivative of the potential at the grain centre. Through the consistency of all boundary values and material and operating parameters, it is suggested that the uniqueness of the Poisson potential for a given method, or form of electron density, can be ensured.

Through this premise, for the Kohn-Sham methodology, the potential dependent, variable, total electron population can be considered constant with respect to the self-consistent charge density. With the uniqueness of the self-consistent potential for the given values of the defining parameters secured by the values of the boundary conditions corresponding to those same defining parameter values, only then does the Kohn-Sham equations, as applied here, actually reflect the underlying principle of variational minimisation of the total energy with respect to a constant particle number constraint. As such, only then does the self-consistent density equal the ground state density of the interacting many-bodied system.

The treatment of the boundary values of the non-linear Poisson equation introduces significant questions regarding the value of some work of literature (principally Ref.'s [15,16,17,31]) in the calculation of such quantities as the band bending and density of occupied surface states in the nano-scale regime, where bulk semiconductor values and approximations are no longer appropriate, even outside of a DFT context.

The Chapter begins on a simple note however, and Section 5.1 gives some detail on the vector operator 'del' and discusses its action on scalar and vector fields with the concepts of gradient, divergence and curl. Also introduced is the theorem of Gauss equating the surface integral of a vector with the volume integral of its divergence.

Section 5.2 applies Gauss' theorem of the first section to the subject of electrostatics and develops Gauss' law, which relates the electric field on a hypothetical closed surface to the charge enclosed within this surface, and from which Poisson's equation is derived. The section then proceeds to apply these equations to a spherical chargeable semiconductor grain and discusses the density of occupied surface states, deriving the standard expression for the occupied surface acceptor density of literature [15,16,17], a function of the first derivative of the potential at the surface of the grain.

In Section 5.3 it is shown that should Poisson's non-linear equation be solved with the boundary requirements of Section 2.2: that the value of the potential on the grain surface is specified, and that this potential must have a stationary point at the grain centre, then unlike the linear case, these boundary conditions do not guarantee an unique solution to the non-linear Poisson equation. The effect of this is that within the previously developed Kohn-Sham methodology – with this self-consistently determined Poisson potential forming part of the single-particle reference potential - this issue of non-uniqueness implies that the total energy of the system can no longer be taken to be minimised with respect to the constraint of constant particle number, since the total particle number is not necessarily constant if there is ambiguity in the self-consistent potential, the total particle number being itself dependent upon this potential. Consequentially, it is feasible that there could be several self-consistent solutions minimised to differing total electron numbers for the system. As such, the self-consistent Kohn-Sham density is not, contrary to anticipation, necessarily the ground state density of the system.

Section 5.4 proves that regardless of this ambiguity in the potential, its first derivative on the grain boundary, $d\varphi/dr\big|_R$, and the potential dependant charge density, $\rho(\varphi, r)$, do uniquely determine

each other, and as such, the electric field at the grain boundary, and so the occupied surface state density, is uniquely set by a given electron density. Naturally, the boundary electric field itself is not uniquely specified unless the charge density, and consequentially, the potential, are themselves unique.

Section 5.5 discusses the requirements necessary to define an electric field uniquely and proves that the electric field does uniquely correspond to the potential, although as in the preceding arguments, if this potential is not unique then neither is the electric field. This is then contrasted with the actual 'physical' requirement that the system be in its ground state and as such, implying that the density, and thus the potential and electric field, are unique. To reconcile these two positions, its is proposed that the values of the defining parameters R, n_d, E_f, T, Q, m^* and ε corresponding to the barrier height S_b are not necessarily those which correspond to stationary point in the potential at the grain centre when the grain is sufficiently small that a region of zero potential does not develop. As such, it is suggested that for the non-linear Poisson equation, the boundary values used here do in fact satisfactorily define the potential if those particular boundary values are consistent with the values of the defining parameters. Then, and only then, will $R^2 \left.\frac{d\varphi}{dr}\right|_R$ equal $-\int_0^R \frac{\rho(\varphi,r)}{\varepsilon} r^2 dr$ and the standard expression for the occupied density of surface states, Eq. (5.43), be true.

In Section 5.6, validation is offered of the claims of the preceding sections, and some areas of error are highlighted in existing published work. It is suggested, that in a spherical geometry, if the surface barrier height is to be kept constant along with E_f, T, Q, m^*, and ε over a range of grain radii, then the ionised donor density cannot remain constant. This section then introduces a straightforward method that can be used to ascertain the true ground state density of the subject non-linear system, and demonstrates its use on a simplified SnO_2 system, modelling the movement of the Fermi level resulting from the injection of individual electrons, such as would be experienced during STM charge writing [6,7].

Section 5.7 provides a synopsis of the developments of the Chapter and their consequences.

Naturally, it should be stressed at all times that with sufficient, accurate data, Poisson's equation, linear or otherwise, can always be precisely and uniquely solved, and indeed, the essence of the techniques developed in this chapter is the divination of such data.

5.1 The Action of ∇ on a Field, Scalar or Vector, and Gauss's Theorem

The vector differential operator ∇, called del or nabla, can be written as

$$\nabla = \frac{\partial}{\partial x}\hat{\mathbf{x}} + \frac{\partial}{\partial y}\hat{\mathbf{y}} + \frac{\partial}{\partial z}\hat{\mathbf{z}} \tag{5.1}$$

in Cartesian coordinates.

When applied to a scalar field $\varphi(x, y, z)$ in a way such that

$$\nabla\varphi = \frac{\partial\varphi}{\partial x}\hat{\mathbf{x}} + \frac{\partial\varphi}{\partial y}\hat{\mathbf{y}} + \frac{\partial\varphi}{\partial z}\hat{\mathbf{z}} \tag{5.2}$$

the differential operator is said to return the gradient of the scalar φ. This is the vector field defined by the requirement that its dot product

$$d\varphi = \nabla\varphi \cdot d\mathbf{r} \tag{5.3}$$

is equal to $d\varphi$, the differential change in φ corresponding to the arbitrary space displacement $d\mathbf{r}$

$$d\mathbf{r} = dx\hat{\mathbf{x}} + dy\hat{\mathbf{y}} + dz\hat{\mathbf{z}} \tag{5.4}$$

From the definition of the dot product then

$$d\varphi = |\nabla\varphi||d\mathbf{r}|\cos\theta \tag{5.5}$$

where θ is the angle between the vector $\nabla\varphi$ and the displacement vector, and it is clear that the rate of change of φ is greatest if the differential displacement is in the direction of $\nabla\varphi$ i.e. $\theta = 0$ so cos 0 = 1. Consequentially, this defines the direction of the vector $\nabla\varphi$ as the direction of the maximum rate of change of φ.

To describe del's action on a scalar field succinctly, it is merely sufficient to state that the gradient of φ is the directional derivative in the direction of the maximum rate of change of φ. It plays many important roles in physics and in particular, it expresses the relation between a force field and a scalar potential field

$$\text{force} = -\nabla(\text{potential})$$

familiar for instance in electrostatics and Newtonian gravitation.

When del is applied to a vector field $\mathbf{V}$ such that

$$\nabla \cdot \mathbf{V} = \frac{\partial V_x}{\partial x} + \frac{\partial V_y}{\partial y} + \frac{\partial V_z}{\partial z} \tag{5.6}$$

a scalar field results that is known as the divergence of the vector field.

More difficult to envisage than the gradient of a scalar field, it can perhaps be better understood from a physical interpretation of its action [65].

Let the vector $\mathbf{V}(x, y, z)$ represent the momentum per unit volume of a liquid, then for a small volume *dxdydz*, see Figure 5-1, the rate of fluid flow into (direction of positive x) this volume per unit time through face EFGH is

$$\text{Rate}\big|_{EFGH} = V_x\big|_{x=0}\, dydz \tag{5.7}$$

the components of $\mathbf{V}$ perpendicular to V_x, V_y and V_z contributing nothing to the inward flow through this particular face.

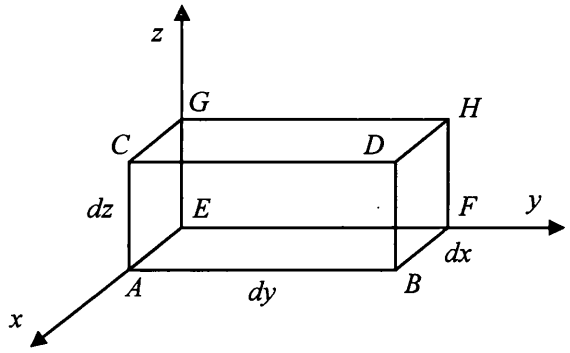


Figure 5-1 Diagram of a parallelepiped in Cartesian space

The rate of flow out through face ABCD is then

$$\begin{aligned}\text{Rate}\big|_{ABCD} &= V_x\big|_{x=dx}\, dydz \\ &\approx \left[V_x + \frac{\partial V_x}{\partial x}dx\right]_{x=0} dydz\end{aligned} \tag{5.8}$$

using a Taylor expansion about the origin (often called a Maclaurin series). The net rate of outward flow in the x direction is then

$$\begin{aligned}\text{Net Rate Out}\big|_x &= \text{Rate}\big|_{ABCD} - \text{Rate}\big|_{EFGH} \\ &= \left.\frac{\partial V_x}{\partial x}\right|_{x=0} dxdydz\end{aligned} \tag{5.9}$$

Naturally, these arguments hold for the other two axes, and so

$$\text{Net Rate Out} = \left(\left.\frac{\partial V_x}{\partial x}\right|_{x=0} + \left.\frac{\partial V_y}{\partial y}\right|_{y=0} + \left.\frac{\partial V_z}{\partial z}\right|_{z=0} \right) dxdydz$$
$$= \nabla \cdot \mathbf{V} d\tau \tag{5.10}$$

and therefore, the net rate of flow out of the volume element *dxdydz* (abbreviated to $d\tau$) per unit volume per unit time is $\nabla \cdot \mathbf{V}$.

However, what is the flow rate, or flux[i], through a surface? It is simply the surface integral

$$\int_S \mathbf{V} \cdot d\boldsymbol{\sigma} \tag{5.11}$$

where the element of area $d\boldsymbol{\sigma}$ can be written $\boldsymbol{n}dA$ where $\boldsymbol{n}$ is a normal unit vector indicating the positive direction. Conventionally, this positive direction is the outward normal if the surface is closed. Therefore, Eq. (5.10) can be written

$$\sum_{\text{6 surfaces}} \mathbf{V} \cdot d\boldsymbol{\sigma} = \nabla \cdot \mathbf{V} d\tau \tag{5.12}$$

For an arbitrary volume V defined by a boundary S, if the space is divided into an arbitrary large number of infinitesimally small parallelepipeds, then for the total volume V, the flux through S is just

$$\sum_{\text{exterior surfaces}} \mathbf{V} \cdot d\boldsymbol{\sigma} = \sum_{\text{volumes}} \nabla \cdot \mathbf{V} d\tau \tag{5.13}$$

the $\mathbf{V} \cdot d\boldsymbol{\sigma}$ terms of all interior faces cancelling, as illustrated in Figure 5-2

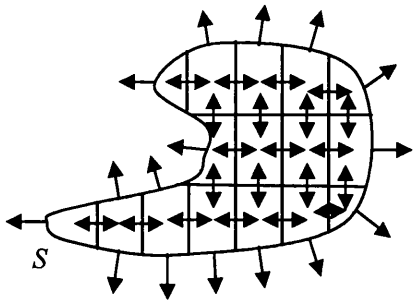


Figure 5-2 Representation of cancellation of $\mathbf{V}.d\boldsymbol{\sigma}$ on interior surfaces

[i] To be pedantic, flux is defined as the product of an area and the field across that area rather than the actual flow of something through an area. However, the later view is conceptually useful and for subtleties, the reader is referred to Haliday, Resnik and Walker [93] for example.

Taking the limit where the number of parallelepipeds tends to infinity and the volume of each tends to zero, then Eq. (5.13) becomes

$$\int_S \mathbf{V} \cdot d\boldsymbol{\sigma} = \int_V \nabla \cdot \mathbf{V} d\tau \tag{5.14}$$

commonly known as Gauss' theorem.

Physically, since $\nabla.\mathbf{V}$ is the net outflow per unit volume as discussed above, $\int_V \nabla \cdot \mathbf{V} d\tau$ is the total net outflow through the volume V, which through Gauss' theorem is equivalent to the surface integral $\int_S \mathbf{V} \cdot d\boldsymbol{\sigma}$ over the surface S defining V.

For completeness, it is sensible to conclude this section with the second application of del to a vector field - through the cross product, referred to as the curl of the field [65]:

$$\begin{aligned} \nabla \times \mathbf{V} &= \left(\frac{\partial V_z}{\partial y} - \frac{\partial V_y}{\partial z} \right) \hat{\mathbf{x}} + \left(\frac{\partial V_x}{\partial z} - \frac{\partial V_z}{\partial x} \right) \hat{\mathbf{y}} + \left(\frac{\partial V_y}{\partial x} - \frac{\partial V_x}{\partial y} \right) \hat{\mathbf{z}} \\ &= \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ V_x & V_y & V_z \end{vmatrix} \end{aligned} \tag{5.15}$$

the determinant expanded from the top down.

To visualise its action, consider the circulation of a fluid around a differential loop in the xy-plane, as seen in Figure 5-3.

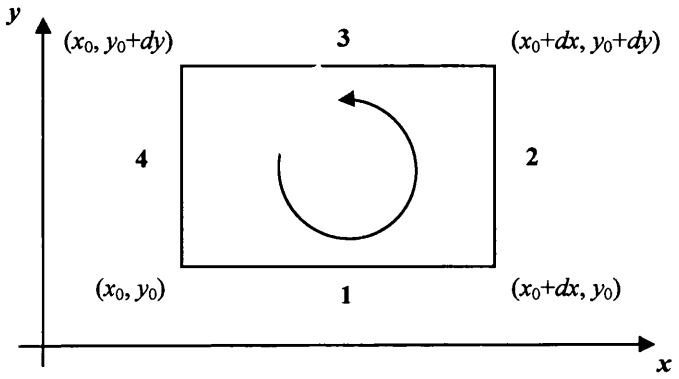


Figure 5-3 Representation of the circulation of a fluid around a differential loop in the xy-plane.

The circulation is given by the vector line integral

$$\int \mathbf{V} \cdot d\boldsymbol{\lambda} \tag{5.16}$$

over each line segment and thus the total circulation around the loop is given by

$$\text{circulation}_{1234} = \int_1 V_x(x,y)d\lambda_x + \int_2 V_y(x,y)d\lambda_y + \int_3 V_x(x,y)d\lambda_x + \int_4 V_y(x,y)d\lambda_y \tag{5.17}$$

From Figure 5-3 it can be seen that over the first line integral $d\lambda_x = dx$, while for the third line integral $d\lambda_x = -dx$. Similarly for $d\lambda_y$: $d\lambda_y = dy$ in the second and $d\lambda_y = -dy$ in the fourth integrals. Then using the Taylor expansions

$$\begin{aligned} V_y(x_0 + dx, y_0) &= V_y(x_0, y_0) + \left.\frac{\partial V_y}{\partial x}\right|_{x_0,y_0} dx + \ldots \\ V_x(x_0, y_0 + dy) &= V_x(x_0, y_0) + \left.\frac{\partial V_x}{\partial y}\right|_{x_0,y_0} dy + \ldots \end{aligned} \tag{5.18}$$

in the limits $dx \to 0$ and $dy \to 0$, the circulation can be expressed as

$$\begin{aligned} \text{circulation}_{1234} &= V_x(x_0,y_0)dx + \left(V_y(x_0,y_0) + \left.\frac{\partial V_y}{\partial x}\right|_{x_0,y_0} dx \right) dy \\ &\quad - \left(V_x(x_0,y_0) + \left.\frac{\partial V_x}{\partial y}\right|_{x_0,y_0} dy \right) dy - V_y(x_0,y_0)dy \\ &= \left(\frac{\partial V_y}{\partial x} - \frac{\partial V_x}{\partial y} \right) dxdy \end{aligned} \tag{5.19}$$

The circulation per unit area in the xy-plane is therefore

$$\text{circulation per unit area} = \nabla \times \mathbf{V}\big|_z \tag{5.20}$$

dividing Eq. (5.19) by $dxdy$.

As such, the curl of the vector **V** can be visualised in terms of its individual components, the axis of each curl component perpendicular to the plane formed by the remaining two axes, and given by the circulation per unit area of **V** on that plane.

5.2 Gauss' Law and Poisson's Equation

Of particular importance to this work are the consequences of Gauss' theorem, developed in the previous section, for electrostatics.

For a point electric charge, q, situated at the origin of a coordinate system, the electric field **E** produced by this charge is defined as the force per unit charge acting on a small test charge q_t

$$\mathbf{E} = \frac{\mathbf{F}}{q_t} \tag{5.21}$$

The force on q_t from q is then

$$\mathbf{F} = \frac{q_t q}{4\pi\varepsilon_r\varepsilon_0}\frac{\hat{\mathbf{r}}}{r^2} \tag{5.22}$$

from Coulomb's law, and so **E** is simply

$$\mathbf{E} = \frac{q}{4\pi\varepsilon_r\varepsilon_0}\frac{\hat{\mathbf{r}}}{r^2} \tag{5.23}$$

From Gauss' theorem then, if a closed surface S does not include the point electric charge q at the origin, as illustrated in Figure 5-4,

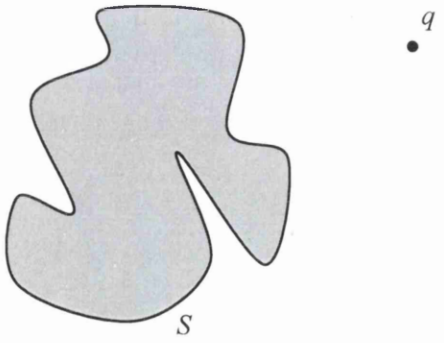


Figure 5-4 Representation of a closed surface S not encompassing point q at the origin of the coordinate system

then

$$\begin{aligned}\int_S \mathbf{E}\cdot d\boldsymbol{\sigma} &= \int_V \nabla\cdot\mathbf{E}\,d\tau \\ &= \frac{q}{4\pi\varepsilon_r\varepsilon_0}\int_V \nabla\cdot\left(\hat{\mathbf{r}}r^{-2}\right)dr \\ &= 0\end{aligned} \tag{5.24}$$

using the general relationship for a function $f(r)$ that

$$\begin{aligned}\nabla\cdot\left(\mathbf{r}f(r)\right) &= \left(\hat{\mathbf{x}}\frac{\partial}{\partial x}+\hat{\mathbf{y}}\frac{\partial}{\partial y}+\hat{\mathbf{z}}\frac{\partial}{\partial z}\right)\cdot\left(\hat{\mathbf{x}}xf(r)+\hat{\mathbf{y}}yf(r)+\hat{\mathbf{z}}zf(r)\right) \\ &= \frac{\partial}{\partial x}xf(r)+\frac{\partial}{\partial y}yf(r)+\frac{\partial}{\partial z}zf(r) \\ &= 3f(r)+\frac{x^2}{r}\frac{df(r)}{dr}+\frac{y^2}{r}\frac{df(r)}{dr}+\frac{z^2}{r}\frac{df(r)}{dr} \\ &= 3f(r)+r\frac{df(r)}{dr}\end{aligned} \tag{5.25}$$

since

$$\frac{\partial}{\partial x}f(r)=\frac{df(r)}{dr}\frac{\partial r}{\partial x}=\frac{df(r)}{dr}\frac{\partial}{\partial x}\left(x^2+y^2+z^2\right)^{1/2}=\frac{x}{r}\frac{df(r)}{dr} \tag{5.26}$$

and where in particular, for $f(r) = r^{n-1}$

$$\begin{aligned}\nabla\cdot\left(\mathbf{r}r^{n-1}\right) &= \nabla\cdot\left(\hat{\mathbf{r}}r^{n}\right) \\ &= (n+2)r^{n-1}\end{aligned} \tag{5.27}$$

which vanishes for $n = -2$ unless $r = 0$ - hence the simplification in Eq. (5.24).

Consider now the second choice of situation; if the surface S does enclose the origin, as seen in Figure 5-5(a).

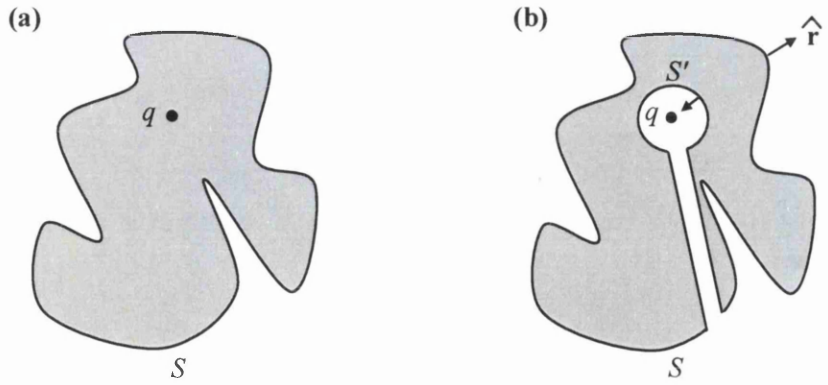


Figure 5-5 Representation of (a) a surface S enclosing the point q at the origin, and (b) one surface made up of the two surfaces S and S' connected by an infinetesimally small hole such that the origin is no longer enclosed

Place a second surface S' of radius δ lying within S and surrounding the origin, see Figure 5-5(b). If these two surfaces are connected by a small hole then both surfaces can be treated as one simply connected closed surface, and should the radius of the connecting hole be allowed to tend to zero, then its contribution to the surface integral vanishes and the total surface integral is just

$$\frac{q}{4\pi\varepsilon_r\varepsilon_0}\left[\int_S\left(\hat{\mathbf{r}}r^{-2}\right)\cdot d\boldsymbol{\sigma}+\int_{S'}\left(\hat{\mathbf{r}}\delta^{-2}\right)\cdot d\boldsymbol{\sigma}'\right]=0 \tag{5.28}$$

the volume defined by S and S' not now containing the origin.

If S' is chosen to be spherical then the element of area $d\boldsymbol{\sigma}'$ can be written

$$d\boldsymbol{\sigma}'=\hat{\mathbf{r}}'\delta^2 d\Omega=-\hat{\mathbf{r}}\delta^2 d\Omega \tag{5.29}$$

where $d\Omega$ is an element of solid angle, and the unit normal vector follows convention and faces outward from the volume i.e. $\hat{\mathbf{r}}'=-\hat{\mathbf{r}}$. Therefore,

$$\int_{S'}\frac{\hat{\mathbf{r}}}{\delta^2}\cdot\text{-}\hat{\mathbf{r}}\delta^2 d\Omega=-4\pi \tag{5.30}$$

which is, significantly, independent of the radius δ.

Thus, from Eq. (5.28) and Eq. (5.30), for the arbitrarily shaped surface S of Figure 5-5(a) enclosing q, the surface integral of the electric field

$$\int_S\mathbf{E}\cdot d\boldsymbol{\sigma}=\frac{q}{\varepsilon_r\varepsilon_0} \tag{5.31}$$

is true.

These two important results, Eq. (5.24) and Eq. (5.31) i.e.

$$\int_S\mathbf{E}\cdot d\boldsymbol{\sigma}=\begin{cases}0 & q \text{ not enclosed by } S\\ \dfrac{q}{\varepsilon_r\varepsilon_0} & q \text{ enclosed by } S\end{cases} \tag{5.32}$$

are collectively known as Gauss' law.

Now consider a charge distribution such that

$$q=\int_V\rho\, d\tau \tag{5.33}$$

Applying Gauss' law, q interpreted as the total distributed charge enclosed within S, then

$$\begin{aligned}\int_S \mathbf{E}\cdot d\boldsymbol{\sigma} &= \int_V \frac{\rho}{\varepsilon_r\varepsilon_0} d\tau \\ &= \int_V \nabla\cdot\mathbf{E} d\tau\end{aligned} \tag{5.34}$$

Since the volume V is arbitrary, as a consequence of the surface S being arbitrary, then the two volume integrands must be equal, and therefore

$$\nabla\cdot\mathbf{E} = \frac{\rho}{\varepsilon_r\varepsilon_0} \tag{5.35}$$

is true. Eq. (5.35) is familiar as one of Maxwell's famous equations of electromagnetism.

Letting the electric field be represented as the gradient of a scalar potential field φ

$$\mathbf{E} = -\nabla\varphi \tag{5.36}$$

then Eq. (5.35) becomes

$$\nabla\cdot(\nabla\varphi) = \nabla^2\varphi = -\frac{\rho}{\varepsilon_r\varepsilon_0} \tag{5.37}$$

the Poisson equation of Section 2.2

Applying Gauss' law to the spherical quantum dot system under study, place a spherical surface S infinitesimally within the boundary of the grain, i.e. at R_-, such that it encloses the charge distribution due to the ionised donors and the disassociated electrons but not the charge within any surface states. As the scalar field is radially dependent only, then

$$\begin{aligned}\mathbf{E} &= -\nabla\varphi(r) \\ &= -\left(\frac{\partial\varphi(r)}{\partial x}\hat{\mathbf{x}} + \frac{\partial\varphi(r)}{\partial y}\hat{\mathbf{y}} + \frac{\partial\varphi(r)}{\partial z}\hat{\mathbf{z}}\right) \\ &= -\left(x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}}\right)\frac{1}{r}\frac{d\varphi(r)}{dr} \\ &= -\hat{\mathbf{r}}\frac{d\varphi(r)}{dr}\end{aligned} \tag{5.38}$$

using Eq. (5.26). With the element of area

$$d\boldsymbol{\sigma} = \hat{\mathbf{r}}R_-^2 d\Omega \tag{5.39}$$

the surface integral of the electric field over S can be written

$$\begin{aligned}\int_S \mathbf{E}\cdot d\boldsymbol{\sigma} &= \int_S -\hat{\mathbf{r}}\left.\frac{d\varphi}{dr}\right|_{R_-}\cdot\hat{\mathbf{r}}R_-^2 d\Omega \\ &= -4\pi R_-^2\left.\frac{d\varphi}{dr}\right|_{R_-}\end{aligned} \tag{5.40}$$

This is equal to the volume integral of the contained charge distribution such that

$$\frac{q}{\varepsilon_r \varepsilon_0} = \int_V \frac{\rho}{\varepsilon_r \varepsilon_0} d\mathbf{r}$$
$$= \frac{4\pi}{\varepsilon_r \varepsilon_0} \int_0^{R_-} (n_d - n_e(r)) r^2 dr \tag{5.41}$$

Requiring the charge within the grain plus the charge contained within the occupied surface states to equal the charge of the grain, Q, then a 'charge balance' equation can be formed such that

$$Q = 4\pi e R^2 N_s + \frac{4\pi}{3} e R^3 n_d - 4\pi e \int_0^R n_e(r) r^2 dr \tag{5.42}$$

where N_s is the occupied surface state density. In the limit $R_- \rightarrow R$, N_s can be written

$$N_s = \frac{Q}{4\pi e R^2} + \frac{\varepsilon_r \varepsilon_0}{e} \left. \frac{d\varphi}{dr} \right|_R \tag{5.43}$$

using Eq. (5.40) and Eq. (5.41). This surface density is assumed to be uniform over the grain surface.

When the grain is uncharged, $Q = 0$, Eq. (5.43) is the charge neutrality condition of Malagú *et al.* [15] and their surface acceptor density, the negative of the occupied surface state density here, is simply Eq. (2.48)

$$N_t = -\frac{\varepsilon_r \varepsilon_0}{e} \left. \frac{d\varphi}{dr} \right|_R$$

a plot of which, as a function of R, can be found in Figure 5-7 as outlined in Section 2.4.

Through the charge balance equation, Eq. (5.42), to some degree the ionised donor density, occupied surface state density, electron density and the scalar potential can all be considered interdependent. As such, they are all rather complex functions, either directly or indirectly, of the temperature, Fermi level, dot radius, surface barrier height as well as the individual material parameters of the semiconductor, such as its effective electron mass and permittivity, and any initial doping of the semiconductor during fabrication. This charge balance equation, although trivial in itself, will in later sections prove to be quite useful.

The spherically symmetric scalar potential φ follows from solving Poisson's equation subject to two boundary conditions. The first requirement is that at the grain centre, the electric field is set to

zero; otherwise, the potential will not be smooth with regard to a transversal of the origin and the symmetry will be broken. The second boundary stipulation is that at the surface of the grain, the potential must be equal to the Schottky-like barrier height, which as discussed in detail in Section 2.2, is the surface barrier at the semiconductor interface. In the case of the tin dioxide grains under study here, this barrier arises through the formation of surface states on the semiconductor as chemisorbed oxygen species (O^-, O_2^-) from the surrounding air act as electron acceptors. The charge in these surface states creates the surface barrier.

Whether the electron density is expressed with a Fermi-Dirac integral or determined self-consistently, it retains a dependence on φ itself, and as such, Poisson's equation is considered non-linear. Only if the radius of the grain is such that it is greater than the electron depletion layer (formed by the previously mentioned chemisorbed oxygen species acting as electron acceptors) penetrating into the nanocrystal, is it possible to neglect the electron density in the total charge density and approximate the potential with a linear Poisson equation. In the situation where the electron density is not superfluous, the non-linearity poses no insurmountable barrier to a solution of the equation via the computational methods discussed in Section 2.3, although it does potentially have very important consequences.

5.3 On the Uniqueness of Poisson Equation Solutions, and the Implications of a Variable Particle Number for a Ground State Density

To begin this section, consider the situation where Poisson's equation is linear:

$$\nabla^2\varphi(r) = \frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d\varphi}{dr}\right) = -\frac{\rho(r)}{\varepsilon} \tag{5.44}$$

Multiplying throughout by r^2, and then on integrating over the radial coordinate

$$r^2\frac{d\varphi}{dr} + C = -\int\frac{\rho(r)}{\varepsilon}r^2dr = A(r) \tag{5.45}$$

and applying the requirement that

$$\left.\frac{d\varphi}{dr}\right|_{r=0} = 0 \tag{5.46}$$

the first derivative of φ can be expressed as

$$\frac{d\varphi}{dr} = \frac{A(r) - A(0)}{r^2} \tag{5.47}$$

Integrating once again,

$$\varphi(r) + k = \int\frac{A(r) - A(0)}{r^2}dr = B(r) \tag{5.48}$$

then with the boundary condition that $\varphi(R) = S_b$, the solution φ becomes

$$\varphi(r) = B(r) + \left(S_b - B(R)\right) \tag{5.49}$$

If a scalar function $e(r)$ is now added to $\varphi(r)$ such that

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d(\varphi + e)}{dr}\right) = -\frac{\rho(r)}{\varepsilon} \tag{5.50}$$

and the same boundary conditions are applied to the new function $\varphi + e$ as to φ, then

$$\frac{d}{dr}(\varphi + e) = \frac{A(r) - A(0)}{r^2} \quad \text{with} \quad \left.\frac{d\varphi}{dr}\right|_{r=0} = 0 \text{ and } \left.\frac{d(\varphi + e)}{dr}\right|_{r=0} = 0 \tag{5.51}$$

and so

$$\varphi(r) + e(r) = B(r) + \left(S_b - B(R)\right) \quad \text{with} \quad \varphi(R) = S_b \text{ and } \varphi(R) + e(R) = S_b \tag{5.52}$$

Therefore, on comparing Eq. (5.49) with Eq. (5.52)

$$\varphi(r)+e(r)=\varphi(r) \tag{5.53}$$

for all r, implying that

$$e(r)=0 \tag{5.54}$$

for all r, and it can be said that the boundary conditions have uniquely determined the solution of Poisson's equation.

Now consider the case of the non-linear Poisson equation:

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d\varphi}{dr}\right)=-\frac{\rho(\varphi,r)}{\varepsilon} \tag{5.55}$$

Proceeding as in the linear case, and with the same boundary requirements, then

$$\frac{d\varphi}{dr}=\frac{A(\varphi,r)-A(\varphi,0)}{r^2} \tag{5.56}$$

where

$$A(\varphi,r)=-\int\frac{\rho(\varphi,r)}{\varepsilon}r^2dr \tag{5.57}$$

and

$$\varphi(r)=B(\varphi,r)+\left(S_b-B(\varphi,R)\right) \tag{5.58}$$

where

$$B(\varphi,r)=\int\frac{A(\varphi,r)-A(\varphi,0)}{r^2}dr \tag{5.59}$$

If a scalar function $e(r)$ is now added to the solution φ however, then this will also alter $\rho(\varphi, r)$ to $\rho(\varphi+e, r)$ and so

$$\begin{aligned}&\frac{d}{dr}(\varphi+e)=\frac{A(\varphi+e,r)-A(\varphi+e,0)}{r^2}\quad\text{with}\quad\left.\frac{d\varphi}{dr}\right|_{r=0}=0\ \text{and}\ \left.\frac{d(\varphi+e)}{dr}\right|_{r=0}=0\\&\varphi(r)+e(r)=B(\varphi+e,r)+\left(S_b-B(\varphi+e,R)\right)\quad\text{with}\quad\varphi(R)=S_b\ \text{and}\ e(R)=0\end{aligned} \tag{5.60}$$

implying that

$$e(r)=\left(B(\varphi+e,r)-B(\varphi,r)\right)-\left(B(\varphi+e,R)-B(\varphi,R)\right) \tag{5.61}$$

with the boundary conditions

$$\begin{aligned}&e(R)=0\\&\left.\frac{de}{dr}\right|_{r=0}=0\end{aligned} \tag{5.62}$$

and whilst this is naturally satisfied by $e(r) = 0$, it is not necessarily the only possible option. As such, the solution of Poisson's equation for those particular boundary conditions is not nessesarily unique.

What bearing does this conclusion have on the self-consistent solution of the coupled Kohn-Sham – Poisson equations? Here the form of the electron density, usually Eq. (1.34) for the first iteration and then essentially

$$n_e(r) = \sum_i f(\varepsilon_i)|\psi_i(r)|^2 \tag{5.63}$$

for the remaining iterations, changes on each cycle until the two sets of equations are self-consistent – that is to say, the electron density that generates the effective potential, generates itself through the consequent wave functions. This self-consistent density is the density that minimises the total energy of the system inclusive of electron-electron interaction effects; in the Kohn-Sham density functional methodology, this is then the ground state density of the system.

The potential does change cycle to cycle, but more importantly so does the entire nature of the relationship generating the electron density. In this way, many potentials, from the many forms of the generating function, can satisfy any boundary requirements, but importantly, only one potential will be self-consistent – only one unique potential will generate, and be generated by, the ground state density. Ergo, for the case of the coupled equations, the charge density in Eq.'s (5.55) to (5.62) should be the final self-consistent density, these equations and arguments having no meaning in the course of the interim self-consistency iterations[i].

As a consequence of the above, a scalar potential e cannot be added to φ without violating this self-consistency and moving the system away from its energetic minimum, unless of course $e(r) = 0$ for all r, and as such, again it can be said that the boundary conditions have uniquely determined the solution of Poisson's equation.

Therefore, based on the above arguments, it appears that if the system is described via Poisson's non-linear equation alone, then the two boundary values do not adequately specify the system.

---

[i] This adoption of the final Kohn-Sham self-consistent density as the only relevant contributor to any 'unique' / 'non-unique' debate involving the KS equations is tacitly embraced throughout the remainder of this Chapter, as indeed is the convention of using 'unique' when, strictly, what is meant is 'unique' for a given method (e.g. Poisson or Poisson-Kohn-Sham).

However, if the system is described via the coupled Kohn-Sham – Poisson equations, then, as a consequence of the additional requirement of energy minimisation, it is apparently implied that the satisfying potential is unique, and therefore, that the two boundary values are adequate after all.

There is one very significant caveat however, which negates this apparent benefit of self-consistent determination of the density and potential. The underlying variational energy minimisation takes place with regard to a constant particle number constraint - see Chapter 4 and Kohnanoff [29]. By including a thermal population of electrons, which are dependent not only on the position of the Fermi level (which is always set at a constant level over each Kohn-Sham – Poisson set of cycles) but also on the position of the conduction band bottom (for these non-flat band scenarios, the conduction band bottom is not a constant, and consequentially, is replaced by its mean value), the total electron population is actually variable, and as such, itself a function of the potential being sought and no longer a true minimisation constraint. While the self-consistency iterations do, without doubt, generate a self-consistent potential consistent with regard to its relevant total complement of electrons, it is not necessarily certain that this density itself is unique. There may exist several, equally valid, self–consistent solutions with different total electron complements, each minimising the total energy with respect to their differing constraints, and through Eq.'s (5.55) to (5.62), they can all satisfy the Poisson equation boundary values.

It could be argued on energetic principles that only the solution with the lowest electron population would be valid; this corresponding to the lowest energy contained in thermally excited electrons and interaction forces. However, the lower the negative charge component of the total charge density, the deeper the potential well, and the deeper the well, the greater the gradient of the potential at the grain surface. This means a larger electric field, and correspondingly more energy 'stored' within this field. Unfortunately then, this minimum electron argument does not provided a simple way out of the non-uniqueness dilemma inherent with this variable particle number system.

Therefore, as matters stand, the solution of Poisson's equation for the particular boundary conditions found here cannot be taken as unique, whether Poisson's equation is solved in isolation or in conjunction with the Kohn-Sham equations.

As such, to elucidate perhaps its most important consequence within the context of this article: if the solution to Poisson's equation is not guaranteed to be exclusive, and this non-unique potential forms part of the reference potential in a Kohn-Sham methodology, then an ambiguity is introduced to the Kohn-Sham electronic density through the variational constraint of total particle number. In this context, this 'constraint' cannot function as a constraint, being itself a variable, since in the Fermi-Dirac fractional occupancy scheme used here, the total electron number is also a function of the mean position of the conduction band bottom (the Kohn-Sham reference potential). With this non-uniqueness issue introduced, the total energy of the system can no longer be taken to be minimised, and as such, the self-consistent Kohn-Sham density is not necessarily the ground state density of the system and consequentially, not nessesarily the density of the many-bodied interacting electron system.

5.4 On the Uniqueness of the Surface Electric Field for a Given Charge Density

Having postulated that the solution of the non-linear Poisson equation is not unique for the given boundary conditions, can the potential dependent electron density be taken as uniquely determining the electric field on the surface S i.e. at the grain boundary?

Suppose that it does not, suppose that the surface electric field corresponding to a potential φ can be both generated by a charge density $\rho(\varphi, r)$ and a charge density $\rho(\varphi+e, r)$. As such, let

$$R^2 \left.\frac{d\varphi}{dr}\right|_R = \int_0^R \left(\frac{d^2\varphi}{dr^2} + \frac{2}{r}\frac{d\varphi}{dr}\right) r^2 dr$$
$$= -\int_0^R \frac{\rho(\varphi+e,r)}{\varepsilon_r \varepsilon_0} r^2 dr \qquad (5.64)$$

where, as before, $e(r)$ is an arbitrary function, the gradient of the potential at the grain centre is always taken to be zero and $\varphi|_R$ is equal to $-S_b$.

For Eq. (5.64) to hold for all radii R then

$$\frac{d^2\varphi}{dr^2} + \frac{2}{r}\frac{d\varphi}{dr} = -\frac{\rho(\varphi+e,r)}{\varepsilon} \qquad (5.65)$$

which is of course the requirement that $\nabla^2\varphi = -\rho(\varphi+e,r)/\varepsilon$. It can be assumed that both

$$\frac{d^2\varphi}{dr^2} + \frac{2}{r}\frac{d\varphi}{dr} = -\frac{\rho(\varphi,r)}{\varepsilon} \qquad (5.66)$$

and

$$-\frac{\rho(\varphi+e,r)}{\varepsilon} = \frac{d^2(\varphi+e)}{dr^2} + \frac{2}{r}\frac{d(\varphi+e)}{dr} \qquad (5.67)$$

are also true, as they are simply statements of Poisson's equation, and as such, proved already via Gauss' law.

Therefore, comparing Eq. (5.65) with Eq.'s (5.66) and (5.67)

$$\rho(\varphi,r) = \rho(\varphi+e,r) \qquad (5.68)$$

or equivalently

$$\frac{d^2\varphi}{dr^2}+\frac{2}{r}\frac{d\varphi}{dr}=\frac{d^2(\varphi+e)}{dr^2}+\frac{2}{r}\frac{d(\varphi+e)}{dr} \tag{5.69}$$

This implies that

$$\left.\frac{d\varphi}{dr}\right|_R=\left.\frac{d(\varphi+e)}{dr}\right|_R \tag{5.70}$$

and thus Eq. (5.64) will hold if, and only if,

$$R^2\left.\frac{d\varphi}{dr}\right|_R=-\int_0^R\frac{\rho(\varphi,r)}{\varepsilon}r^2dr \quad \text{and} \quad R^2\left.\frac{d(\varphi+e)}{dr}\right|_R=-\int_0^R\frac{\rho(\varphi+e,r)}{\varepsilon}r^2dr \tag{5.71}$$

Ergo, with regard to the potential, $\left.\frac{d\varphi}{dr}\right|_R$ and $\rho(\varphi,r)$ uniquely determine each other, and as such, if the electric field at the grain boundary is uniquely determined by the potential, then it is uniquely set by the given charge density.

Interestingly, neither Eq. (5.68) nor Eq. (5.69) actually requires $\varphi + e$ to equal φ. Indeed

$$\frac{d^2(\varphi+e)}{dr^2}+\frac{2}{r}\frac{d(\varphi+e)}{dr}=\frac{d^2\varphi}{dr^2}+\frac{2}{r}\frac{d\varphi}{dr}+\frac{d^2e}{dr^2}+\frac{2}{r}\frac{de}{dr} \tag{5.72}$$

implying that

$$\frac{d^2e}{dr^2}+\frac{2}{r}\frac{de}{dr}=0 \tag{5.73}$$

which is simply a statement of Laplace's equation $\nabla^2 e=0$. Of course, e must also meet the boundary requirements, Eq. (5.62):

$$e(R)=0$$
$$\left.\frac{de}{dr}\right|_{r=0}=0$$

However, this is not an issue of practical importance as the computational method employed ensures by its nature that the scalar field employed on the right hand side and the left hand side of Poisson's equation is always the same.

To summarise this section briefly then, it has been shown that the electric field at the grain boundary corresponding to φ and the density $\rho(\varphi)$ do have a unique correspondence. However, as $\rho(\varphi)$ is not unique since φ is not unique in this non-linear scenario, see Section 5.3, then the value of the electric field at the grain boundary is itself not uniquely given.

Identifying the occupied surface state density with the surface electric field, it can be taken, reiterating the above, that there is a one to one correspondence between the charge density and the occupied surface state density, but this surface density cannot be uniquely known until the potential is uniquely determined.

5.5 On the Uniqueness of the Electric Field, and Satisfying the Charge Balance (CB) Equation

Having ascertained that the electric field at the grain surface and the electron density uniquely determine each other if the field is uniquely set by the potential, the next question is whether this can be applied to the electric field as a whole. Is the electric field for a given potential in this situation unique?

In general, there are many different potentials that can generate the same field –a concept referred to as gauge invariance. A gauge is a particular choice of scalar and vector potential with which to define a field, and a gauge function is a scalar function that can be used to change the gauge. For instance, in its most general form, the electric field can be expressed in terms of the scalar potential φ and the vector potential $\mathbf{A}$ such that

$$\mathbf{E} = -\nabla\varphi - \frac{\partial}{\partial t}\mathbf{A} \tag{5.74}$$

then any gauge transform of the form

$$\begin{aligned} \varphi &\rightarrow \varphi - \frac{\partial\gamma}{\partial t} \\ \mathbf{A} &\rightarrow \mathbf{A} + \nabla\gamma \end{aligned} \tag{5.75}$$

where the gauge function $\gamma(\mathbf{x}, t)$ is an arbitrary function, leaves the field $\mathbf{E}$ unchanged.

The most popular gauge of electromagnetism is the Coulomb gauge, where it is required that

$$\nabla \cdot \mathbf{A} = 0 \tag{5.76}$$

As such, substituting Eq. (5.74) into Eq. (5.35) yields

$$\nabla^2\varphi + \frac{\partial \nabla \cdot \mathbf{A}}{\partial t} = -\frac{\rho}{\varepsilon} \tag{5.77}$$

which can be reduced to the familiar Poisson equation

$$\nabla^2\varphi = -\frac{\rho}{\varepsilon}$$

on employing Eq. (5.76).

However, when the charge density itself is a function of the scalar potential, the gauge transform will also affect it, and while Gauss's law will of course still be obeyed, it is not necessarily the

case that the gauge transformed charge density will still generate the gauge transformed scalar potential meeting the original boundary conditions.

What if the boundary conditions of the system were such that the potential was defined absolutely? Given that gauge transformations would appear to be no longer appropriate, can it be assumed that the electric field is also defined absolutely? The real question then that must be answered is actually what does it take to define the electric field uniquely?

The vector field uniqueness theorem stated and proven in Arfken and Weber [65] specifies that a vector is only uniquely given if its divergence and curl are known within a simply connected region and its normal component is specified over the boundary of the region:

$$\begin{aligned} &\nabla \cdot \mathbf{V} = s \\ &\nabla \times \mathbf{V} = \mathbf{c} \\ &V_n \end{aligned} \tag{5.78}$$

The scalar s is referred to as the source density, and the vector $\mathbf{c}$ as the circulation density.

Initially, choose the Coulomb gauge for simplicity, then for the electric field $\mathbf{E} = -\nabla\varphi$, its curl can be written

$$\nabla \times (-\nabla\varphi) = -\begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ \dfrac{\partial\varphi}{\partial x} & \dfrac{\partial\varphi}{\partial y} & \dfrac{\partial\varphi}{\partial z} \end{vmatrix} \tag{5.79}$$

which, on expanding the determinant, is always equal to zero, and so $\nabla \times \mathbf{E}$ can be taken to be specified as zero, satisfying requirement (2) of Eq. (5.78).

The divergence of the electric field $\nabla \cdot \mathbf{E} = -\nabla \cdot (\nabla\varphi)$, requirement (1) of Eq. (5.78), is taken as being determined by the solution of Poisson's equation $\nabla^2\varphi = -\rho(\varphi, r)/\varepsilon_r\varepsilon_0$ meeting the usual boundary conditions

$$\begin{aligned} &\varphi(R) = S_b \\ &\left.\frac{d\varphi}{dr}\right|_{r=0} = 0 \end{aligned}$$

as discussed in Section 5.2. However, it was shown in Section 5.3 that φ, and so $\rho(\varphi)$, are not necessarily unique under these circumstances.

With regard to requirement (3) of Eq. (5.78), the value of the electric field on the boundary; from the work of Section 5.4, it can be seen that once again, there will be a one to one correspondence between the boundary value of **E** and the scalar potential φ /charge density $\rho(\varphi)$ (gauge temporarily fixed), although, the value of the electric field on the boundary will not itself be uniquely given, as again, φ is not necessarily unique.

It would appear then, that if φ were unique, then from Eq. (5.78), the electric field would be unique, but since the potential φ is not unique, neither is **E**. In this case (non-unique **E**) can it also be assumed that there is a one to one correspondence between the non-unique φ and the non-unique **E**?

Consider the reverse. Assume for a moment that more that one electric field can correspond to a given potential φ, and consequentially it is possible to add a vector function **e** to the field so that

$$\nabla \cdot (\mathbf{E} + \mathbf{e}) = \frac{\rho(\varphi, r)}{\varepsilon_r \varepsilon_0} \tag{5.80}$$

is also true in addition to

$$\nabla \cdot \mathbf{E} = -\nabla^2 \varphi = \frac{\rho(\varphi, r)}{\varepsilon_r \varepsilon_0}$$

This implies that

$$\nabla \cdot \mathbf{e} = 0 \tag{5.81}$$

and so

$$\begin{aligned} -4\pi R^2 \left.\frac{d\varphi}{dr}\right|_R + 4\pi \int_0^R \nabla \cdot \mathbf{e} r^2 dr &= 4\pi \int_0^R \frac{\rho(\varphi, r)}{\varepsilon_r \varepsilon_0} r^2 dr \\ &= -4\pi R^2 \left.\frac{d\varphi}{dr}\right|_R \end{aligned} \tag{5.82}$$

then

$$\begin{aligned} \int_S \mathbf{e} \cdot d\boldsymbol{\sigma} &= \int_S (\mathbf{E} + \mathbf{e}) \cdot d\boldsymbol{\sigma} - \int_S \mathbf{E} \cdot d\boldsymbol{\sigma} \\ &= -4\pi R^2 \left.\frac{d\varphi}{dr}\right|_R + 4\pi R^2 \left.\frac{d\varphi}{dr}\right|_R \\ &= 0 \end{aligned} \tag{5.83}$$

and thus the normal component of **e** on the grain boundary is

$$e_n = 0 \tag{5.84}$$

5.5 On the Uniqueness of the Electric Field, and Satisfying the Charge Balance (CB) Equation

Requiring the electric field to remain spherically symmetric then **e** can only be a function of the radial coordinate, and as such, its curl must be zero

$$\nabla \times \mathbf{e} = 0 \tag{5.85}$$

from the definition Eq. (5.15) recast in spherical polar coordinates

$$\nabla \times \mathbf{V} = \begin{vmatrix} \hat{\mathbf{r}}/(r^2 \sin\theta) & \hat{\boldsymbol{\theta}}/(r\sin\theta) & \hat{\boldsymbol{\phi}}/r \\ \dfrac{\partial}{\partial r} & \dfrac{\partial}{\partial \theta} & \dfrac{\partial}{\partial \phi} \\ V_r & rV_\theta & rV_\phi \sin\theta \end{vmatrix} \tag{5.86}$$

A vector with zero curl is called an irrotational vector, and as such, can be represented as the negative gradient of a scalar function

$$\mathbf{e} = -\nabla e \tag{5.87}$$

on consideration of Eq. (5.79).

From Green's theorem[i] and Eq. (5.84)

$$\int_V \nabla e \cdot \nabla e \, d\tau = \int_V \mathbf{e} \cdot \mathbf{e} \, d\tau = 0 \tag{5.88}$$

and since $\mathbf{e}\cdot\mathbf{e} = e^2$ is greater than or equal to zero, then

$$\mathbf{e} = 0 \tag{5.89}$$

With Eq.'s (5.81), (5.84) and (5.85), **e** is uniquely specified, and from Eq. (5.89), uniquely set at zero. Therefore, there can only be one electric field corresponding to φ.

The choice of gauge here is actually immaterial – the electric field must remain unchanged and so consequentially, the source and circulation densities must also be unaffected by any gauge transformation. Gauge is just a matter of convenience. As such, for a given potential and charge density there will be an unique electric field, specified by Eq.'s (5.79), (5.35) and (5.38) at *R*. Although, if the potential for a system is not unique, as in the case of the non-linear Poisson

---

[i] From the identity [65]

$$\nabla \cdot (u\nabla v) = u\nabla \cdot \nabla v + \nabla u \cdot \nabla v$$

linking the continuous scalar functions *u* and *v*, then for a volume *V* with a surface *S* Gauss' theorem implies that

$$\int_S u\nabla v \cdot d\boldsymbol{\sigma} = \int_V u\nabla \cdot \nabla v d\tau + \int_V \nabla u \cdot \nabla v d\tau$$

a result known as Green's theorem, or to be more precise, one of its forms.

equation with the boundary values used here, then the electric field of the system will not be unique overall, but will uniquely correspond to that particular potential.

Nevertheless, regardless of all this ambiguity, when the total electronic system is at its energetic minimum, the electron density is the ground state density, and since the ground state is taken to be unique, the potential generating it must also be unique. Ergo, the electric field is unique and the occupied surface state density, Eq. (5.43),

$$N_s = \frac{Q}{4\pi e R^2} + \frac{\varepsilon_r \varepsilon_0}{e} \left. \frac{d\varphi}{dr} \right|_R$$

appearing in the charge balance equation Eq. (5.42),

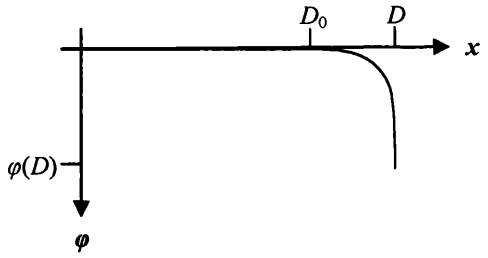
$$Q = 4\pi e R^2 N_s + \frac{4\pi}{3} e R^3 n_d - 4\pi e \int_0^R n_e(r) r^2 dr$$

is determined exactly, specified at the values of R, S_b, n_d, E_f, T, Q, m^* and ε which correspond to the ground state density.

How can the work of Section 5.3, that the density is not necessarily determined uniquely, be reconciled with this?

The key issue is the boundary values of the Poisson equation. The surface barrier height is naturally dependent on the parameters defining the system – for instance consider the standard planar geometry Schottky equation[ii]

[ii] In a planar geometry, for the illustrated system between D_0 and D



Poisson's equation is

$$\frac{d^2\varphi}{dx^2} = -\frac{en_d}{\varepsilon_0 \varepsilon_r}$$

and

$$V_b = \frac{eN_t^2}{2\varepsilon_0\varepsilon_r n_d} \tag{5.90}$$

of Ref. [94] for example, where the activation energy, or in the terminology used here, the built in potential V_b is S_b minus the energy difference between the conduction band bottom and the Fermi level, assuming bulk semiconductor behaviour. Change a parameter such as the permittivity or ionised donor density and unless the other contributing factors are adjusted to compensate, the barrier height will change.

What about the gradient of the potential at the grain centre? In the depletion approximation (DA), it is a pre-requisite that the grain radius is sufficient in extent that for the parameters defining the system, a field free, flat band region always exists at the grain centre. Therefore both Poisson boundary values are consistent with the defining quantities of (R,) n_d, E_f, T, Q, m^*, and ε, and in this approximation, are sufficient to set the potential uniquely.

$$\left.\frac{d\varphi}{dx}\right|_{x=D_0} = 0$$

$$V_b = \varphi(D) - \varphi(D_0)$$

Integrating Poisson's equation then

$$\frac{d\varphi}{dx} + C = -\frac{en_d}{\varepsilon_0\varepsilon_r}x$$

and the constant of integration C can be found by evaluating the above expression at D_0:

$$C = -\frac{en_d}{\varepsilon_0\varepsilon_r}D_0$$

Integrating again

$$\varphi + k = -\frac{en_d}{\varepsilon_0\varepsilon_r}\left(x^2 - D_0 x\right)$$

and again evaluating the expression at D_0 to find the constant of integration k:

$$k = -\frac{en_d}{\varepsilon_0\varepsilon_r}D_0^2 - \varphi(D_0)$$

The charge from the ionised donors in the depletion region D-D_0 must equal the charge from the surface acceptor density:

$$(D - D_0)n_d = N_t$$

and so finally, assessing the value of the potential D yields

$$\begin{aligned} V_b &= \varphi(D) - \varphi(D_0) \\ &= -\frac{en_d}{\varepsilon_0\varepsilon_r}\left(D^2 - 2DD_0 + D_0^2\right) \\ &= -\frac{eN_t^2}{\varepsilon_0\varepsilon_r n_d} \end{aligned}$$

When the band bending does not fully evolve, when the non-linear Poisson equation must be used, is it not an inherent assumption that the parameters which are consistent with S_b are also consistent with $d\varphi/dr = 0$ for $r = 0$?

Consider instead the possibility that this derivative is also a function of those same parameters defining S_b - R, n_d, E_f, T, Q, m^* and ε - and furthermore, that there exists only one unique potential, corresponding to one unique charge density, which satisfies both boundary conditions.

In this way, it is proposed that the potential satisfying R, S_b, $\left.\frac{d\varphi}{dr}\right|_{r=0} = 0$, n_d, E_f, T, Q, m^*, and ε consistently is unique, and so the non-linear Poisson equation is solved uniquely. As such, the electron density represents that of the actual 'physical' system for those values of parameter, which in the case of the coupled KS-Poisson equations, can be identified as the ground state density.

The concept of taking $d\varphi / dr|_{r=0}$ as dependent on the defining parameters is quite logical. At the start of the section, it was proven that the electric field does uniquely correspond to a given charge density, and, as that charge density is undoubtedly dependent on R, n_d, E_f, T, Q, m^* and ε, it follows that **E** must be dependent also. With the Coulomb gauge in this spherically symmetric geometry, what else is the electric field other than the first derivative of the potential along the radial axis? Indeed, this offers an explanation for the unique/non-unique potential divide over the switch in the Poisson equation from linear to non-linear; the charge density within the neutral region ($r < R_0$) of a large grain ($R > \Lambda$) is, by its very definition, zero i.e. constant, hence the independence of **E** at $r = 0$ (the boundary value $d\varphi / dr|_{r=0}$) to the defining parameters[iii], whilst for small grains ($R < \Lambda$), a neutral region is not present and the charge density, and consequently **E**, at $r = 0$ is undeniably dependent on the defining parameters. Ergo, small R implies dependent $d\varphi / dr|_{r=0}$.

The assumption that $d\varphi / dr|_{r=0} = 0$ and $\varphi(R) = -S_b$ may not be consistent for all sets of the defining parameters is also rather a logical conclusion. It would seem that something of this

[iii] although Λ itself would always be dependent on the defining parameters R, n_d, E_f, T, Q, m^* and ε.

nature is certainly implied when considering the limiting case where the grain radius R tends to zero.

Consider an uncharged grain. When $R \geq \Lambda$, see Figure 5-6(a), the band bending does fully develop and as $R \rightarrow \infty$, an extended region, where both the first and the second derivatives of the potential equal zero, can be seen to develop. However, when $R < \Lambda$, as discussed in Section 2.4, the potential difference between the surface and the centre of the grain is no longer the built in potential, V_b, as $\varphi(0)$ moves downwards, slowly approaching $-S_b$ on diminishing R, as seen in Figure 5-6 (b) and (c).

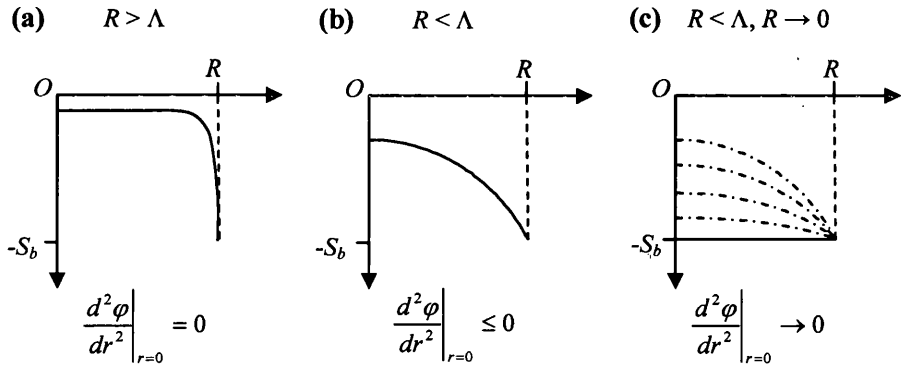


Figure 5-6 Representation of the potential φ for a spherical grain where (a) its radius is larger than its depletion width (b) its radius is less than the depletion width – the band bending does not fully evolve, and (c) the radius tends to zero.

Indeed, as $R \rightarrow 0$ both $\varphi(0) \rightarrow \varphi(-S_b)$ and $d\varphi/dr|_{r=R} \rightarrow d\varphi/dr|_{r=0}$, and therefore, from the usual definition of the second derivative, see for instance Ref. [50],

$$\left.\frac{d^2\varphi}{dr^2}\right|_{r=0} = \lim_{\delta r \to 0} \frac{d\varphi/dr|_{r=\delta r} - d\varphi/dr|_{r=0}}{\delta r}$$

when the radius tends to zero and δr must be less than or equal to R then $d^2\varphi/dr^2|_{r=0} \rightarrow 0$.

In addition, in the same limit $n_e(0) \rightarrow n_e(R)$, and using one of the bulk approximations of n_e, say Eq. (1.41) of Section 1.3,

$$n_e(r) = N_c e^{(E_f + e\varphi)/k_B T}$$

which in the usual reference frame (the Fermi level represents the zero of the energy), means that

$$n_e(0) \rightarrow N_c e^{-eS_b/k_B T} \tag{5.91}$$

then, unless $-S_b$ is equal to the value of $\varphi(0)$ for a $R > \Lambda$ grain

$$^{R>\Lambda}\varphi(0) = \frac{k_B T}{e} \ln\left(\frac{n_d}{N_c}\right) \tag{5.92}$$

then $\rho(0)$($= en_d - en_e(0)$) will not be equal to zero. In general, since $S_b > - {}^{R>\Lambda}\varphi(0)$, then as R tends to zero $\rho(0) \rightarrow \Sigma$ where Σ is some positive number.

As such, with a zero first derivative, Poisson's equation at $r = 0$ is given by

$$\left.\frac{d^2\varphi}{dr^2}\right|_{r=0} = -\frac{\rho(0)}{\varepsilon_r \varepsilon_0}$$

which implies that

$$0 = -\frac{\Sigma}{\varepsilon_r \varepsilon_0} \tag{5.93}$$

as $R \rightarrow 0$, clearly incorrect unless $\varepsilon \rightarrow \infty$ or Σ is, in fact, zero. So, save the defining parameters changing (for example the permittivity tending to infinity or $n_d \rightarrow 0$ together with $N_c \rightarrow 0$), then this inconsistency will only be avoided if S_b is equal to $^{R>\Lambda}\varphi(0)$. Ergo, it can be taken in this case the parameters consistent with $\left.d\varphi/dr\right|_{r=0} = 0$ are not consistent with general S_b, and are only consistent with one particular value of S_b.

While this limiting case is certainly not conclusive proof of the conjecture that the charge density satisfying both of the Poisson equation boundary values is unique[iv], it is unquestionably suggestive.

Accepting that this speculation concerning $\left.d\varphi/dr\right|_{r=0}$ and the uniqueness of the charge density is not necessarily unrealistic, what are the implications?

[iv] For instance, it could be argued that as $R \rightarrow 0$ Eq. (1.41) is no longer applicable and should be replaced by a quantised expression. While of course this is true (and indeed taken into account in later sections), what is important is the possibility of the inconsistency that has been suggested.

Although, in support of Eq. (5.93) in a quantised treatment, for a small enough radius there will be no energy levels present at all in the grain and as such, no electron density and thus only positive ρ possible, $\Sigma = en_d$. Interestingly, this will happen at a radius greater than zero and as such, Eq. (5.93) can't be so easily dismissed by asserting that as $R \rightarrow 0$ the actual number of ionised donors present in the grain will reach zero (although the density of these donors would not have to change), and hence confirming the apparent inconsistency of Eq. (5.93) for general boundary values. However, see also footnote [iv] of Section 6.1.

5.5 On the Uniqueness of the Electric Field, and Satisfying the Charge Balance (CB) Equation

Regardless of the radius of the grain, the zero potential derivative (vitally) forms one of the initial conditions in the shooting method solving the Poisson equation, incidentally handily providing a means to avoid the computational issues of a 1/0 factor, and so a zero potential is automatically assumed for all sets of operating parameters. If at $R < \Lambda$ the above assumption is correct however, then it would be expected that for all operating parameter sets other than the set best simulating the energetic minimum i.e. the set containing those values most suitable for approximating the actual physical system, then some manner of discontinuity will appear between this zero value of the derivative and the smoothly evolving derivatives of the following points on the actual mesh of computed values. Consequentially, for the true model of the physical system, while it would be expected that for a small displacement from the origin, say the inter mesh spacing Δ for a sufficiently closely spaced mesh, the first derivative at that point would deviate from zero, that deviation would be negligible; if it proved not to be, that its value was not insignificant, then the inequality

$$\begin{aligned}\int_0^R \left(\frac{d^2\varphi}{dr^2}+\frac{2}{r}\frac{d\varphi}{dr}\right)r^2dr &= \int_0^R \frac{d}{dr}\left(r^2\frac{d\varphi}{dr}\right)dr \\ &= R^2\frac{d\varphi}{dr}\bigg|_R - \lim_{\Delta\to 0}\Delta^2\frac{d\varphi}{dr}\bigg|_\Delta \\ &\neq R^2\frac{d\varphi}{dr}\bigg|_R \end{aligned} \tag{5.94}$$

will be evident, and the defining parameter set will not be that of the physical system. Gauss' equation will still hold of course for whatever potential satisfies Poisson's equation, but, as a consequence of Eq. (5.94), only for the true system, the one whose parameters minimise its energy, is the actual occupied density of surface states given by Eq. (5.43).

As a result, for practical computation, an additional δ term is incorporated into the charge balance equation Eq. (5.42) so that

$$Q = 4\pi eR^2N_s + \frac{4\pi}{3}eR^3n_d - 4\pi e\int_0^R n_e(r)r^2dr + \delta \tag{5.95}$$

is true for the general system, and where the deviation factor δ can be subdivided into its constituent components such that

$$\delta = \delta^{der} + \delta^{disc} + \delta^{KS} \tag{5.96}$$

The term δ^{der} is commonly the most important contribution to δ and follows from Eq. (5.94) and its accompanying discussion. Explicitly, it is written

$$\delta^{der} = 4\pi\varepsilon_0\varepsilon_r\Delta^2\left.\frac{d\varphi}{dr}\right|_{\Delta} \tag{5.97}$$

and estimates the charge present due to the departure of $d\varphi/dr|_{r=0}$ from zero. For an uncharged grain and the systems represented here, $d\varphi/dr$ at Δ is non-positive and as such, δ^{der} represents a zero or negative charge. On charging, this term is more flexible, representing charges of any value and polarity.

The second term on the RHS of Eq. (5.96), δ^{disc}, represents the $O(\Delta^3)$ error ($O(\Delta^4)/\Delta$) due to the estimation of the continuous derivatives $R^2\left.\frac{d\varphi}{dr}\right|_R$ and $\Delta^2\left.\frac{d\varphi}{dr}\right|_{\Delta}$ with their discretisations, the $O(\Delta^3)$ error from the numerical integration of the electron density, and finally, the error present as a consequence of approximating terms at the origin by their values at Δ, namely $d\varphi/dr|_{\Delta}$. As $\Delta \rightarrow 0$ so to will δ^{disc}; however, at practical inter mesh spacings, $\Delta \approx 0.01 - 0.1$nm, δ^{disc} may not be negligible for high electron densities (for instance those in the order 10^{25}m^{-3} - a consequence of high ionised donor densities) and appears to become dominated by the error from the numerical integration of the electron density. In general, δ^{disc} provides a zero or positive charge contribution.

The final term contributing to δ, δ^{KS}, represents the error that can occur in the coupled methodologies (Schrödinger-Poisson or Kohn-Sham-Poisson) given that the final electron density appearing in the charge balance equation is a consequence of the generating potential, not part of the charge density that generated the potential. As such, unless perfect self-consistency has been achieved, Gauss' law will not be exactly satisfied if this new electron density replaces the old, generating density, and therefore an inconsistency between the charge contained within the grain and the surface integral of the electric field, calculated from the existing potential, will then appear. This charge is accounted for via δ^{KS}. In general, this term is a complex object and can be non-negligible for high electron densities or densities where a significant percentage of the total electron population lies in thermally populated states above the Fermi level, and as such, are electron densities particularly susceptible to flux and thus difficult to make self-consistent. Of course, any acceptable solution to the coupled equations must be self-consistent to within a high degree of tolerance, regardless of the appropriateness of the defining parameters, and consequentially δ^{KS} should always tend to zero if a self-consistent solution is possible for the

given criteria. When Poisson's equation alone is being used to model the band structure, then δ^{KS} is naturally also zero.

To summarise then, it is proposed that both boundary values used here in solving the non-linear Poisson equation are dependent on the material and operating parameters, and not just S_b. As such, fully consistent R, S_b, $\left.\frac{d\varphi}{dr}\right|_{r=0} = 0$, n_d, E_f, T, Q, m^*, and ε, and fully consistent combinations only, will uniquely specify the potential, and thus the electron density. This unique electron density, for the specific values of R, n_d, E_f, T, Q, m^* and ε that are consistent with the two boundary values, corresponds to some energetic minimum of the total electronic system, be that the full DFT Kohn-Sham- Poisson scheme, where the density is the ground state density, or simply the Poisson equation alone with a bulk semiconductor model of the electron density. Only for these fully consistent boundary values and defining parameters is δ in Eq. (5.95) zero, and does Eq. (5.43) for the density of occupied surface states hold. Only then can the true potential, charge density and occupied surface state density of the given system be considered actually known.

Of course, for large grains of sufficient radius that the depletion approximation can be used and it is possible to neglect the effects of the mobile charge carriers, then Poisson's equation is linear. In this case δ is automatically zero, $\left.\frac{d\varphi}{dr}\right|_{r=0} = 0$ consistent with any S_b, provided S_b is consistent with the defining parameters (n_d, E_f, T, Q, m^*, and ε) of the system.

The charge error or deviation term, δ, when it is present, will contain the error of any deviation from zero of the derivative of the potential at the grain centre, as well as contributions from the consequences of the unavoidable deviation between a continuous space and its discretised approximation, and any residual error from the failure of coupled systems to achieve self-consistency. In actuality, the point of zero δ represents the most physical parameters of the discretised system, not necessarily the continuous one, although as $\Delta \rightarrow 0$ these should become identical to each other.

5.6 Evidence in Validation of the Claims Regarding the Boundary Values of the Non-Linear Poisson Equation

If the reasonings of the preceding sections are correct, then this question of uniqueness certainly does have implication for some of the existing publications within the literature, and in particular for the theoretical calculations of Malagú *et al.* in Ref.'s [15,16,17,31].

For instance, consider the relationship between surface acceptor density and grain radius for SnO_2 nanocrystals investigated in Ref. [15] using standard semi-classical models – as briefly introduced in Section 2.4. Specifically, take Figure 2-12, reprinted below as Figure 5-7 with the non-degenerate trend removed and two depletion approximation series added,

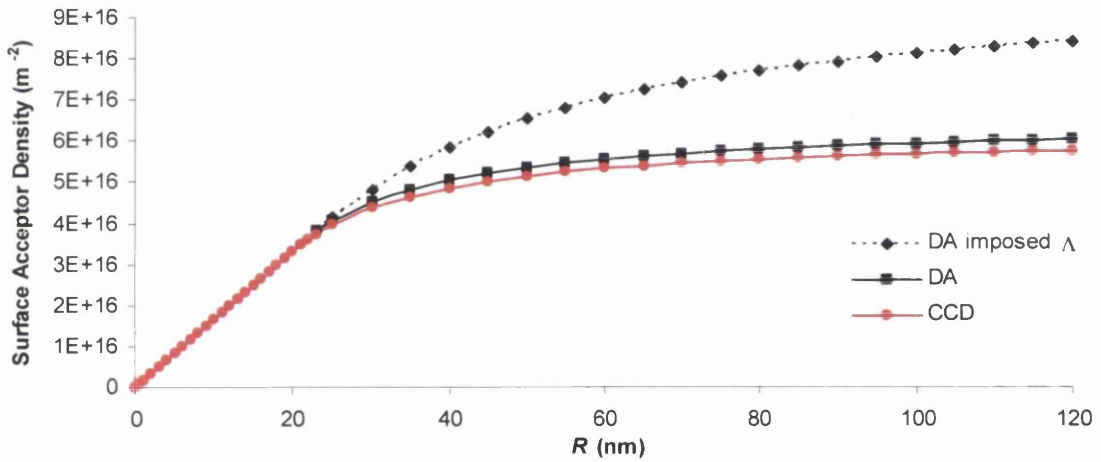


Figure 5-7 Plots of surface acceptor density against SnO_2 grain radius for the Malagú *et al.* [15] system. Both the complete charge density (CCD) and depletion approximation (DA) (when $R \geq 22$nm only) are plotted in addition to the depletion approximation imposing a constant Λ of 22nm (DA imposed Λ).

which is calculated as in Ref. [15], duplicating their Figure 6, with Eq.'s (2.45) to (2.48) i.e.

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d\varphi(r)}{dr}\right) = -\frac{en_d}{\varepsilon_r\varepsilon_0}\left(1-e^{e\varphi(r)/k_bT}\right)$$

$$-\left.\frac{d\varphi}{dr}\right|_{r=0} = 0$$

$$\varphi(R) = -V$$

$$-\left.\frac{d\varphi}{dr}\right|_{r=R^-} = \frac{eN_t}{\varepsilon_r\varepsilon_0}$$

and assuming that Eq. (5.42) holds:

$$4\pi eR^2 N_t = \frac{4\pi}{3} eR^3 n_d - 4\pi e \int_0^R n_d e^{e\varphi / k_b T} r^2 dr$$

The zero of the potential is set at the value of the potential at the centre of a sufficiently large grain that the band bending fully evolves, with V the value of the potential at the surface of the grain relative to that potential zero. If the band bending is fully evolved, then V is equal to V_b - the built in potential of the grain, defined as the difference between the potential at the centre of the grain and the potential at the grain boundary.

As mentioned in Section 2.2, this model is usually referred to as the complete charge density model (CCDM) since the charge density includes both the effects of the ionised donor atoms as well as the thermal electron population. This contrasts with the depletion approximation (DA), where only the influence of the ionised donor atoms is included in the charge density.

For the experimentally measured values of a sample SnO_2 system in air ($T = 673.15$K, $n_d = 5\times10^{24}$ m^{-3}, $V = 0.68$V and $\varepsilon_r\varepsilon_0 = 10^{-10}$ Fm^{-1}), the authors of Ref [15] report a constant depleted region, Λ, of ~22nm (provided of course that $R > \Lambda$). They conclude that the DA is a good approximation in this range as it correctly models the surface state density to within less than 5% of the CCD values.

Representing the charge deviation δ through the percentage difference R_{ch}

$$\begin{aligned} R_{ch} &= 100 \times \frac{q_1 - q_2}{q_1} \\ q_1 &= 4\pi e \int_0^R n_d e^{e\varphi / k_b T} r^2 dr \\ q_2 &= \frac{4\pi}{3} eR^3 n_d + 4\pi R^2 \varepsilon_r \varepsilon_0 \left.\frac{d\varphi}{dr}\right|_{r=R} \end{aligned} \tag{5.98}$$

where R_{ch} expresses δ (= q_1 - q_2) as a percentage of the electron density, then Figure 5-8 illustrates the behaviour of this percentage difference against grain radius for the tin dioxide system described above.

When $R \geq 22$nm and the DA is a good approximation i.e. when the effect of the thermal electrons is negligible and Poisson's equation is effectively linear, then R_{ch} initially lies at ~0.02%, tending to ~ -0.0006% as $R \rightarrow \infty$. Then as discussed in the preceding sections, in this situation the

boundary conditions do uniquely determine the potential, $\left.\frac{d\varphi}{dr}\right|_{r=0} = 0$ independent of the defining parameters, and Eq. (5.42) is automatically satisfied, at least to within the bounds of the error inherent within the experimental quantities and the computational process itself.

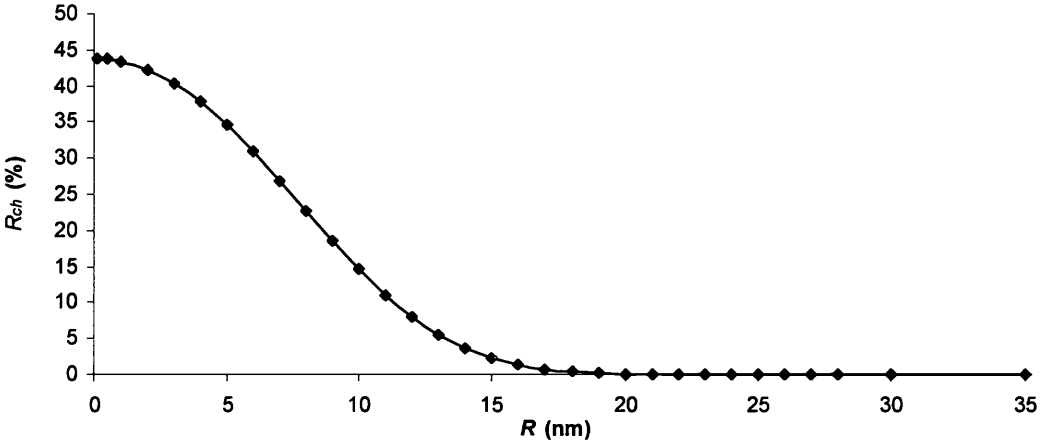


Figure 5-8 Plot of the percentage difference, R_{ch}, given by Eq. (5.98) against SnO_2 grain radius for the Malagú *et al.*[15] CCD system.

However, in the range, $R < 22$nm, where a region of zero potential does not form, then a clear disparity exists between $R^2\left.\frac{d\varphi}{dr}\right|_R$ and $-\int_0^R \frac{\rho(\varphi,r)}{\varepsilon} r^2 dr$, reaching a substantial 43.9% as $R \rightarrow 0$. Separate from the need to allow for the effects of a discrete energy spectrum, this discrepancy can be understood through the existence of the charge error term δ (the component δ^{KS} set at zero); the given parameters of T, n_d, V and $\varepsilon_r\varepsilon_0$ not consistent with $\left.\frac{d\varphi}{dr}\right|_{r=0} = 0$ for these smaller radii.

Consequently, in this particular scenario, with these particular parameters, below 22nm the surface state acceptor density of Figure 5-7 cannot be accepted as accurate.

Should T and $\varepsilon_r\varepsilon_0$ be taken as exact for all radii, and with the assumption that V remains unchanged over the various sizes of grain, then, by the theories proposed here, if the Fermi level is not to move or the grain to become spontaneously charged, then it is implied that the ionised donor density must be variable with respect to the grain radius. When the grains are of such an extent that the band bending can fully develop and a region of zero potential can form at the grain centre, this ionised donor density will tend to a constant bulk value.

If this is so, then for the range $R < 22$nm, it should be possible to find the point where δ is zero by varying n_d, maintaining the values of the other parameters. Indeed, for a 10nm radius grain with the usual $T = 673.15$K, $V = 0.68$V and $\varepsilon_r\varepsilon_0 = 10^{-10}$ Fm^{-1}, the charge error δ, its component parts, and the corresponding value R_{ch} (δ as percentage of the charge of the thermal electrons), can be found in Figure 5-9, where n_d is varied from 1×10^{24} m^{-3} to 1×10^{26} m^{-3}.

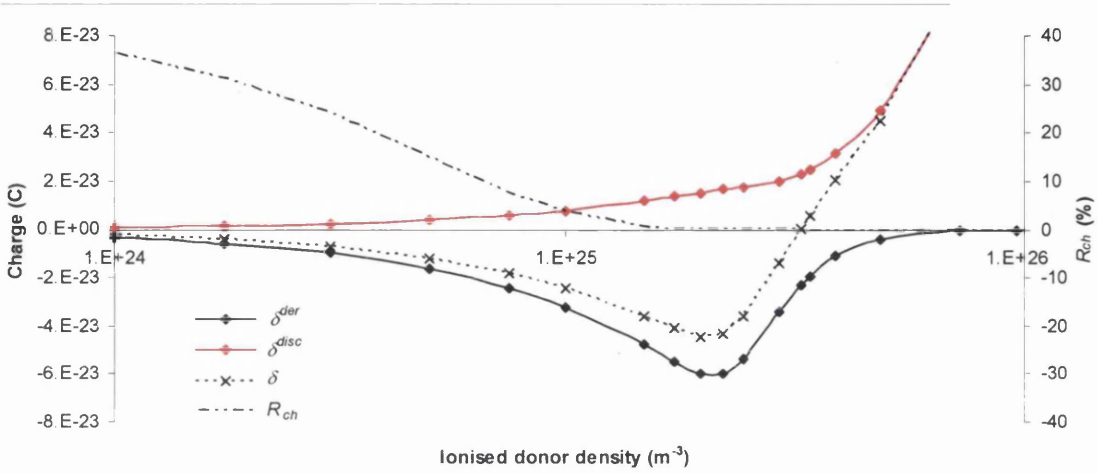


Figure 5-9 Plots of the charge error δ and its components, along with the percentage difference, R_{ch}, given by Eq. (5.98) against the ionised donor density, n_d, for a 10nm radius SnO_2 grain based on the Malagú *et al.*[15] CCD system

The most important contribution, δ^{der} - from the deviation of $d\varphi/dr$ at $r = 0$ from its required value of zero, can be seen to smoothly increase in magnitude from ~ -2×10^{-24} C at $n_d = 1\times10^{24}$m^{-3} to ~ -6×10^{-23}C at around $n_d = 2.25\times10^{25}$m^{-3}, before dwindling away to essentially nothing (~ -6×10^{-26} C) by $n_d = 1\times10^{26}$m^{-3}. The positive charge component δ^{disc} lies at 8×10^{-25} C at $n_d = 1\times10^{24}$m^{-3} and follows the exponential-like increase of q_1, the total charge of the thermal electrons, over the range of ionised donor densities, dominating δ after 3.25×10^{25}m^{-3}, the point at which R_{ch} crosses the x-axis. As such, the zero value of R_{ch} (and therefore δ), the point at which the model actually reflects the physical grain, is taken to lie at $n_d = 3.25\times10^{25}$m^{-3}. R_{ch} is increasingly negative after this point, although this is hard to see from the graph due to the damping effect of the rapidly increasing electron charge q_1.

At $n_d = 5\times10^{24}$m^{-3} the density of occupied surface acceptor states for a 10nm grain is 1.67×10^{16}m^{-2}, as seen in Figure 5-7. For the newly proposed ionisation donor density of $n_d = 3.25\times10^{25}$m^{-3}, the corresponding occupied surface acceptor density rises to 1.05×10^{17}m^{-2} – a six fold increase on the old N_t.

Interestingly, this increase in ionised donor density for decreasing radius is hinted at even from the depletion approximation. To demonstrate this in a way pertinent to the models of Malagú *et al.* [15], a slightly circuitous route is followed, the reasoning behind which should shortly become evident.

Ref. [15], working in the depletion approximation, develops the analytical expression for the potential of the spherical system

$$\varphi(r) = \frac{en_d}{\varepsilon_r \varepsilon_0}\left(-\frac{r^2}{6} - \frac{R_0^3}{3r} + \frac{R_0^2}{2}\right) \tag{5.99}$$

and the relation

$$N_t = n_d\left(\frac{R}{3} - \frac{R_0^3}{3R^2}\right) \tag{5.100}$$

for the surface acceptor density (see Section 6.1 for more details), where R_0, as always, represents the radius of the flat band region. For the usual values of T = 673.15K, $n_d = 5\times10^{24}\text{m}^{-3}$, and V_b = 0.68V, the authors of Ref. [15] conclude that the neutral region grows linearly with the radius, generating a constant depletion width, Λ, where $\Lambda = R - R_0$, of ~20nm – consult their Figure 2.

However, Eq. (5.99) can of course be written as a third order polynomial in R_0,

$$R_0^3 - \frac{3R}{2}R_0^2 + \left(\frac{R^3}{2} - \frac{3V_b\varepsilon_r\varepsilon_0 R}{en_d}\right) = 0 \tag{5.101}$$

the roots[i] of which are the acceptable values of R_0 for the particular R dependant coefficients used. In this case, R_0 must be both real positive or zero, and less than R, and as such, only one root is

---

[i] Formulae for the solutions of the general third order polynomial

$$ax^3 + bx^2 + cx + d = 0$$

can be found in any good mathematical reference book, see for instance Woan [73] or for an historical account, Gullberg [95]. Let

$$p = \frac{1}{3}\left(\frac{3c}{a} - \frac{b^2}{a^2}\right)$$

$$q = \frac{1}{27}\left(\frac{2b^3}{a^3} - \frac{9bc}{a^2} + \frac{27d}{a}\right)$$

and

$$D = \frac{p^3}{27} + \frac{q^2}{4}$$

Then if $D \geq 0$

physically acceptable. The depletion width for this root is plotted against grain radius in Figure 5-10.

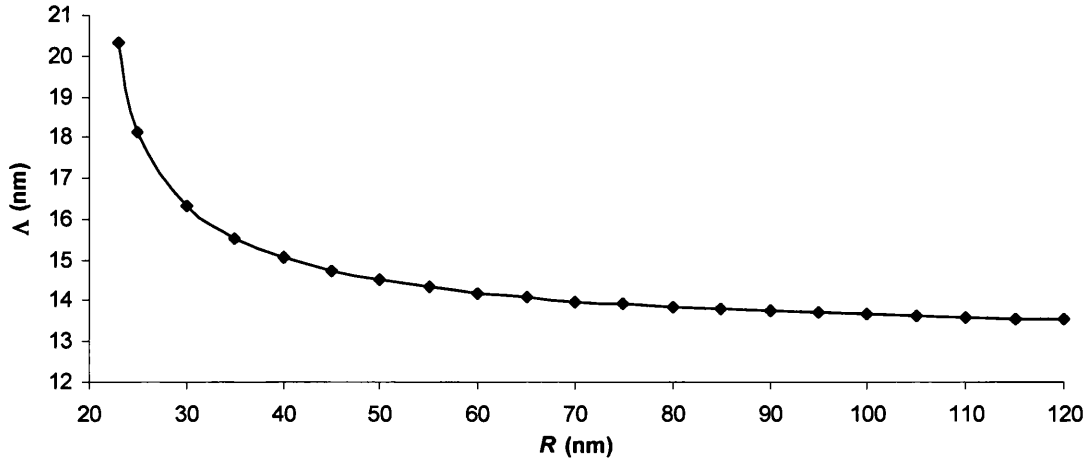


Figure 5-10 Plot of depletion width, Λ, against SnO_2 grain radius for the Malagú *et al* [15] DA system. The grain radii are taken to be sufficiently large that it can be assumed that the band bending fully forms and as such, that the depletion approximation (DA) is appropriate.

Evidentially, Λ tends to a constant value of around ~13nm, and does not remain a steady 20nm as reported. The ionised donor density used in generating this data series is kept constant throughout.

$$u = \left(-\frac{q}{2} + \sqrt{D}\right)^{1/3}$$

$$v = \left(-\frac{q}{2} - \sqrt{D}\right)^{1/3}$$

$$x_1 = u + v - \frac{b}{3a}$$

$$x_{2,3} = -\frac{u+v}{2} - \frac{b}{3a} \pm i\sqrt{3}\frac{u-v}{2}$$

If D is greater than zero then there are two complex and one real root, otherwise if D is equal to zero, then the three roots are real roots and at least two of them will be equal.

While if $D < 0$ then

$$\theta = \arccos\left[-\frac{q}{2}\left(\frac{3}{|p|}\right)^{3/2}\right]$$

$$x_1 = 2\sqrt{\frac{|p|}{3}}\cos\frac{\theta}{3} - \frac{b}{3a}$$

$$x_{2,3} = -2\sqrt{\frac{|p|}{3}}\cos\frac{\theta \pm \pi}{3} - \frac{b}{3a}$$

and there are three distinct real roots.

This trend in R_0 can be seen in the profile of the DA surface acceptor density in Figure 5-7. At a grain radius of ~22nm, the N_t's of both depletion approximation series' and the complete charge model are very close, and all have similar depletion widths, ~20nm, ~20nm and ~22nm respectively, implying that the effects of the thermal electrons are minimal at this radius.

However, as the radius increases, the volume of the shell depleted of charge carriers also increases, in proportion to R^3-R_0^3, and consequentially the charge on the surface of the grain must also increase in line with this. The effect then of neglecting the charge of the mobile charge carriers within this shell will also be greater since the shell volume is greater, hence the slow but increasing separation of the true DA and the complete charge carrier profiles. This is assuming that the mobile charge carrier concentration over the depletion width is independent of the actual radius of the grain; that is to say, that the charge per unit volume due to the thermal electrons in the depletion width is independent of grain size. Of course, the increase of the N_t overestimation will not continue indefinitely as the second term in Eq. (5.100) will tend to zero as $R \rightarrow \infty$ and the surface density will become linearly dependent on R, independent of R_0, and the separation between the profiles constant.

The effect on the surface state density for a constant depletion width of 20nm is plotted in Figure 5-7 as the trend 'DA imposed Λ'. The observant reader will notice the difference between this and Figure 3 of Ref. [15] where the authors state that they plot the same quantity from the same equations, presumably with their claim of constant ~20nm depletion width. Their figure bears a closer resemblance to the correct DA surface density calculated with a variable depletion width tending to ~13nm. The artificial maintenance of this extended depletion width naturally results in an overestimation of the contribution from the ionised donor atoms, over and above the positive charge overestimation effect due to the neglect of the negative thermal electrons, and as such, causes a corresponding overestimation of the negative charge within the surface states.

A cursory glance at Figure 2.11's (c) and (d), identical to Ref. [15]'s Figure 4a and Figure 4b for a 30nm and a 100nm grain respectively, does seem to counter the above discussion and indicate that their depletion widths are constant. However, a closer look at the region of interest of these figures, say the last 25nm as shown in Figure 5-11, indicates otherwise, and clearly exhibits the expected diminution of the depletion width. Although, despite an actual increase in the average thermal electron presence per unit volume on the increase of the grain radius (augmenting the effects of discounting these electrons in the DA further), it is the complete charge density profile

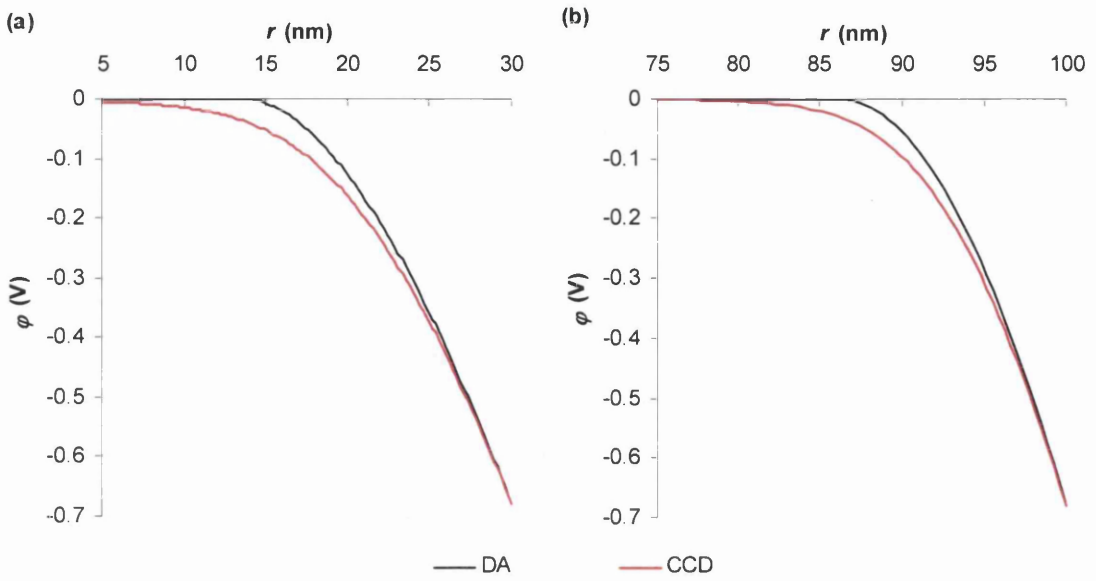


Figure 5-11 Plots of the complete charge density (CCD) and depletion approximation (DA) potentials for the last 25nm of (a) a R = 30nm grain, and (b) a R = 100nm grain, of the Malagú *et al* [15] system.

of the 30nm grain that differs the most from its complementary DA profile. This can perhaps be explained by the increased role of the first derivative in Eq. (2.45) (through its effect on the values of φ and their effect in the electron density) at smaller radii.

The complete charge density potential does not exactly reach zero, but if R_0 is taken at the point which the potential passes -0.01V, then for 30nm, Λ = 21.33nm (DA Λ = 16.32nm) and for 100nm, Λ = 16.29nm (DA Λ = 13.67nm).

Having now clearly demonstrated that if the ionised donor density is constant then Λ is variable, what of the case where the depletion width is kept constant?

In this situation, it is the ionised donor density that again must be allowed to vary if the other parameters of Eq. (5.99) - the permittivity of the space and the barrier height - are maintained at a constant level. For the two depletion widths, Λ = 13nm and Malagú *et al*'s Λ = 20nm, Eq. (5.99) yields the ionised donor density against grain radius trend of Figure 5-12. As already stated, only a depletion width of ~13nm is consistent with a bulk ionised donor density of $5\times10^{24}m^{-3}$.

The increase in the required ionised donor density to maintain the values of the other parameters on the decrease of radius is very clear, and is certainly supportive of the claims of rising n_d on

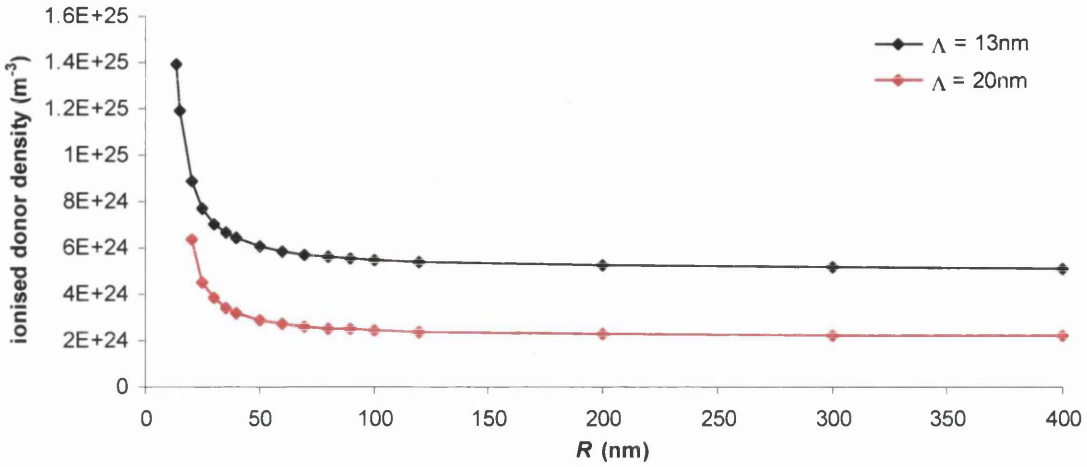


Figure 5-12 Plot of ionised donor density for the two constant depletion widths of 13nm and 20nm, against SnO_2 grain radius for the Malagú *et al* [15] DA system.

decreasing R for the CCD system. In the latter, there is no 'choice' whereby another quantity, such as the depletion width (there is no Λ), can be varied to maintain the consistency of the boundary values $d\varphi/dr\big|_{r=0}$ and S_b, and the remaining, 'fixed', defining parameters of R, E_f, m^* and ε.

Indeed, this ionised density movement , along with the effects of energy quantisation, may go some way to explain the discrepancy between the measured ratio of occupied surface state densities between 4nm and 15nm grains, 5.90, and the theoretical ratio of ~3 calculated using Malagú *et al.*'s method [31].

For the experimental parameters pertinent to the MNC Charge Writing Group, the effects of the charge discrepancy δ and its implications for the ionised donor density are discussed and treated in some detail in Chapter 6.

Accepting that the quantity δ must be reduced to zero for the true ground state density (KS methodology), or just some energetic minimum density (Schrödinger-Poisson or Poisson alone), to be reached and for the resultant electronic properties to be any reflection of the physical system being modelled, then provided that sufficient data can be experimentally gathered or calculated that all but one of the independent variables determining δ is known, the missing value, pertinent to the ground state, can be found by varying the unknown quantity until δ is caused to vanish, and

$R^2 \left.\frac{d\varphi}{dr}\right|_R$ equals $-\int_0^R \frac{\rho(\varphi,r)}{\varepsilon} r^2 dr$. This was in essence the procedure of extracting the 'correct' n_d from Figure 5-9. The bisection method, already discussed in regard to the numerical solution of Poisson's equation in Section 2.3, is particularly useful in achieving this, as its robust nature ensures the best possible chance of determining a convergent solution of what, at times, can be a quite sensitive and ill-conditioned problem.

For the purpose of illustrating this outer self-consistency iterative cycle to satisfy the charge balance equation (adopting 'inner' for use with the self-consistency cycles used to simultaneously solve the Kohn-Sham and Poisson equations where this is relevant), consider a 5nm radius SnO_2 grain at 250K. If the surface barrier height, S_b, relative to the Fermi level is 1eV, then neglecting electron-electron effects and only using s orbitals, then, for a neutral grain, self consistently reconciling Poisson's equation, Schrödinger's equation and the charge balance equation, and thereby requiring δ to equal zero, an ionised donor density of 5.75×10^{25} m^{-3} is necessary.

Using that ionised donor density, it is a simple matter to repeat the calculations, now varying the charge on the grain, Q, to simulate charge injection, and solving for the Fermi level at each step. The Fermi level cannot of course remain constant as this situation is effectively simulating the addition of electrons, to the discrete, unoccupied (or possibly partially occupied) energy levels, lying (by definition) above the Fermi level of the grain of no, or lower, charge.

With the previous constraints of allowing only $l = 0$ orbitals and neglecting electron-electron interactions, then Figure 5-13 demonstrates the Fermi level movement of the grain as it undergoes charging. A clear stepped pattern emerges, with two electrons stored at each plateau. This corresponds to the double occupancy (two possible spin states) of each s-orbital, although since the temperature of the system is non-zero, a thermal population of electrons exits above the Fermi level, and so the Fermi level does not simply equal the energy of each eigenstate in turn. Each added electron from the STM tip should be seen as more of an addition of a unit charge to the overall electron sea present within the semiconductor conduction band, as opposed to the addition of an electron to any particular orbital.

The increased energy spacing between each plateau reflects the increasing difficultly in moving the system further from its natural, uncharged, state of energetic equilibrium, and is consistent

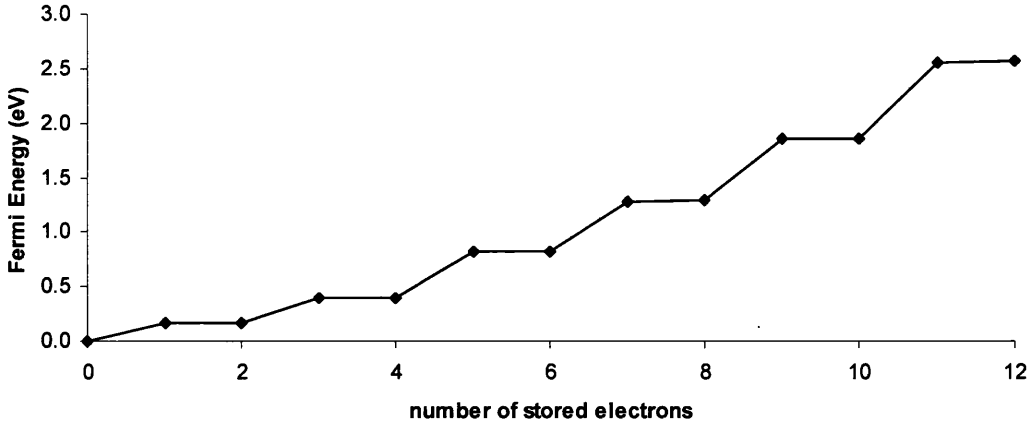


Figure 5-13 Plot of Fermi level movement against number of stored electrons for a 5nm radii SnO_2 grain at 250K with S_b = 1eVand $n_d = 5.75\times10^{25}m^{-3}$. Self-consistent calculations, neglecting all electron-electron interaction effects and working solely with *s* orbitals ($l = 0$).

with the increasing energetic separation of the discrete eigenstates.

Conceptually at least, this model of charging presents no difficulties, and could quite feasibly produce the characteristic Coulomb staircase I-V plot of single-electron charging [6,11,12]. Again, this topic of charge storage and Fermi level movement will be revisited in more detail in the following Chapters and validated with experimental data.

In summary, this section has provided validation of the claim of the preceding section that the derivative of the potential at the centre of a grain is not always zero for all choices of radius and material parameters corresponding to the specified barrier height.

In this way, support has been given for the assumption that this dependence of $d\varphi/dr$ at $r = 0$ enables the non-linear Poisson equation to be solved uniquely. This once again ensures that the resulting self-consistent density of the coupled Kohn-Sham-Poisson equations is the true ground state density of the interacting many bodied system and thereby endows the results of the numerical simulation with some physical meaning.

In providing this validation, some possible flaws in the work of Malagú *et al.* have been highlighted, with both the CCD and DA models, and the idea of a variable ionised donor density with respect to grain radii has been introduced.

Finally, a simple and useful method has been demonstrated whereby the charge balance equation can be satisfied in conjunction with the Poisson and, if required, Schrödinger or Kohn-Sham equations, provided that all but one of the defining parameters R, n_d, E_f, T, Q, m^*, ε and S_b are determined. As an illustrative example, the charging of a simplified 5nm radius SnO_2 grain at 250K was modelled, its ionised donor density first determined for the neutral case, then on the addition of electrons, its Fermi level was solved for, with a clear stepped pattern evident in its movement.

5.7 Conclusions

This Chapter discusses the nature of scalar and vector fields in some detail, covering some aspects of the work of Gauss in this area and deriving Poisson's equation. More importantly, it also investigates the importance of uniqueness with relation to the potential, electric field and charge density of the spherical system under study.

It is suggested that, when the spatial dimensions are such that a flat band region does not develop at the grain centre and the non-linear Poisson equation must be solved to ascertain the potential, then it is no longer the case that the Poisson equation boundary value $d\varphi/dr\big|_{r=0}=0$ is automatically consistent with all values of the defining parameters R, n_d, E_f, T, Q, m^* and ε. As such, it is proposed that the potential and therefore the electron density satisfying S_b, $d\varphi/dr\big|_{r=0}=0$, R, n_d, E_f, T, Q, m^* and ε consistently is unique, and represents the 'physical' system for those values of parameter.

Consequentially, it is argued that the charge balance equation Eq. (5.42),

$$Q=4\pi eR^2N_s+\frac{4\pi}{3}eR^3n_d-4\pi e\int_0^R n_e(r)r^2dr$$

with the occupied surface state density of Eq. (5.43),

$$N_s=\frac{Q}{4\pi eR^2}+\frac{\varepsilon_r\varepsilon_0}{e}\frac{d\varphi}{dr}\bigg|_R$$

is only satisfied for these sets of consistent values.

Evidence validating these premises is then presented from the consideration of the bandbending in the limit $R\rightarrow 0$, the behaviour of the depletion approximation in order to maintain a constant depletion width, and from an analysis of the work of Malagú *et al.* [15,31] with particular attention to Ref. [15]. Along with some error pertaining to their depletion approximation section, the main area of fault appears to lie in Malagú *et al.*'s calculation of the (occupied) surface acceptor density from the potential of the non-linear Poisson equation where the grain radius is less that ~22nm, as demonstrated in Figure 5-8. For the case of a 10nm grain, maintaining the surface potential at -0.68 V and varying the ionised donor density, the point at which the charge balance equation is actually satisfied rests at an ionised donor density 6.7 times greater than that

used in Ref [15], and consequently there is a corresponding increase in the occupied surface acceptor density by a factor of 6.3. It is proposed that this ionised donor density increase, in conjunction with the effects of quantisation, may account for the considerable discrepancy between experiment and theory when this model is put into practise in Ref. [31].

The final development of this Chapter is a more qualitative validation of its assertions: a simple method is introduced which can be used to ascertain the true density satisfying the charge balance equation for a non-linear system, and this applied to a simplified SnO_2 system modelling the movement of the Fermi level on the injection of electrons into the grain (see Figure 5-13). A clear stepped pattern is in evidence, two electrons stored at each plateau corresponding to the double occupancy of each *s*-orbital, and which could very feasibly produce the characteristic Coulomb staircase I-V plot of single-electron charging.

The dependence of the non-linear Poisson equation boundary value $d\varphi / dr\big|_{r=0}$ on the values of the defining parameters R, n_d, E_f, T, Q, m^* and ε, is of particular consequence when the Poisson equation is not solved in isolation, but coupled with the Kohn-Sham equations. It assures, that even with a variable electron population (with a dependence on the reference potential), that the self-consistent density is the unique ground state density. The electron population is then constant with respect to the self-consistent density, and as the boundary conditions satisfied by the self-consistent potential ensure that the self-consistent potential is unique for the given values of the defining parameters, only then does the Kohn-Sham equations as applied here actually reflect the underlying variational principle, i.e. the minimisation of the total energy with respect to a constant particle number constraint, and only then does the self-consistent density equal the ground state density of the interacting many bodied system.

The way is now paved for Chapter 6 where all the developed techniques of this work will be applied to modelling true physical systems and the theoretical results compared with those of experiment.

Chapter 6 Application of the Theoretical Models to Experimental Systems

This Chapter presents results based on the concatenation of the models developed over the previous Chapters. These preceding sections have introduced and created the theory and approximations necessary to provide an accurate simulation of the electronic properties of nano-scale quantum dots, and the following portion of this treatise applies the culmination of these reasonings to modelling nanocrystalline SnO_2 grains in a spherical geometry in order to achieve a fuller understanding of the experimental results collated for these systems.

Section 6.1 is devoted mainly to assessing the surface state densities of the 4nm and 15nm radii nanocrystals of Maffeis *et al.* [31]. Divided into five sub-sections, the first, 6.1.1, estimates the likely depletion width of the semiconductor using the non-linear Poisson's equation, with the traditional expression for the electron density, in conjunction with the charge balance equation. The second sub-section, 6.1.2, uses this depletion width within the analytical depletion approximation to determine the surface state density and the ionised donor density of the bulk semiconductor. This enables the limit on the maximum possible physical ionised donor density to be set. The third sub-section, 6.1.3, is independent of the two preceding sub-sections, other than in the sense that they place its results within some larger context, and applies the full Poisson-Kohn-Sham-Charge Balance scheme to ascertain the ionised donor densities of small radii grains ($R \leq 20$nm). In this way the 4-15nm surface state density ratio is determined. Sub-section 6.1.4 introduces a few additional thoughts and further, though tentative, evidence in validation of the variable ionised density approach through a study of existing SnO_2 gas sensing data. The final sub-section presents a summary of the section and a discussion the results of the three very different models.

Section 6.2 is an investigation into the charging of 4nm grains as reported in Maffeis *et al.* [7] and, in particular, Wilks *et al.* [6]. Separated into two sub-sections, the first, 6.2.1, models the movement of the Fermi level, using the Poisson-Kohn-Sham-Charge Balance method, with the incremental increase of charge stored within the grain, and offers an estimate of the total electron complement for comparison with the experimentally inferred limit. This data is then used in sub-section 6.2.2 in a simple tunnelling model to simulate the response of the current through the structure to variations in tip-substrate potential difference. This is then evaluated against Wilks *et al.*'s measured tunnelling current. This elementary tunnelling model also allows the voltage interval between charge storage events to be calculated and compared with that inferred from experiment by Ref.[6].

In the usual manner, the final section, Section 6.3, concludes the Chapter with a synopsis of the work presented within it. It summaries the results found herein and their likely value from evaluation against the relevant experimental data published in literature.

6.1 Determination of Ionised Donor Density and Surface State Density in Neutral SnO_2 Grains

Experimental work has been performed on SnO_2 grains in air at 296K by Maffeis *et al.* [31] in order to assess their occupied surface state densities; or to be more exact, the ratio of occupied surface state densities between two sizes of grain, namely R = 4nm and R = 15nm. As reported in Section 5.6, the existing theoretical model unsatisfactorily predicts a ratio nearly half that actually measured. The nanocrystals of the experimental group were manufactured via an evaporation-condensation method using a Differential Mobility Analyser (DMA) [42], sintering taking place during the particle formation process at 923K. STM measurements confirm their radii at 4nm and 15nm, and STS data places their conduction band minimums at the grain surfaces, relative to the Fermi level, at 1.3(±0.05)eV and 1.4(±0.1)eV respectively[i]. Utilizing further STM-STS measurements, normalized conductivity spectra integrated the over the surface band gap have been used to give a representation of the surface state density. Over an average of 6400 *I-V* curves, the mean values of the conductivities were 0.5V for the 4nm, and 2.95V for the 15nm nanoparticles – implying a surface density ratio of 5.90 between the two grain sizes.

In order to offer a model that calculates the surface state ratios between the two samples accurately, the full charge neutrality scheme developed in Chapter 5 must be applied to determine the ionised donor density, n_d, corresponding to the measured barrier heights for the required radii.

6.1.1 The Non-Linear Poisson Equation and the Determination of the Approximate Depletion Width of the Bulk Semiconductor

Neglecting the effects of a discrete energy spectrum, and utilising the traditional representation of the electron density in the bulk, Eq. (1.34),

$$n_e(r) = N_c \mathscr{F}_{1/2}\left[\frac{E_f - v_p(r)}{k_B T}\right]$$

[i] Interestingly, whilst the 0.1V difference between the surface band edges measured by Maffeis *et al.* is within their bounds of experimental error, this shift could also be direct evidence of Fermi level unpinning, see Section 2.1 and Section 2.4.

the first assessment of the problem can be made, coupling the non-linear Poisson equation with the charge balance equation (P-G approximation] to ensure consistent $d\varphi/dr|_{r=0} = 0$ and $\varphi(R) = -S_b$ as outlined in Chapter 5. In this way, the surface barrier heights of 1.3 eV and 1.4 eV form the upper and lower limits of each data point, the ionised donor density, n_d, being varied to ensure the overall neutrality of the grain, as can be seen in Figure 6-1.

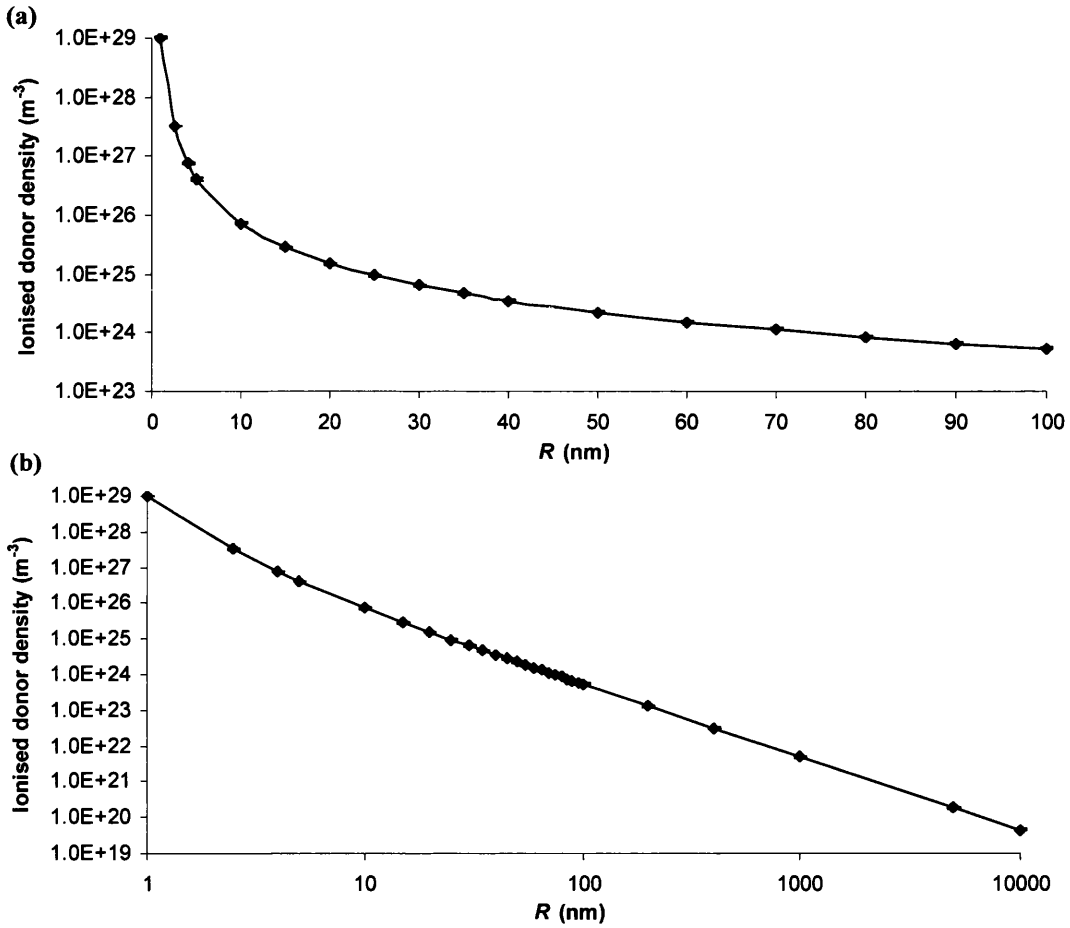


Figure 6-1 Plots of ionised donor density against SnO_2 grain radius for the P-G approximation, in the ranges (a) 1-100nm (b) 1-10,000nm. Upper and lower limits of error bars correspond to the surface barrier heights, S_b, 1.3eV and 1.4eV respectively.

The density of donors, N_d, must be constant, and as such, independent of grain size (being oxygen vacancies introduced through the annealing of the sample in an oxygen atmosphere at an elevated temperature during particle formation). This then implies that the ionised density would be expected to tend to a constant level, the ionised donor density seen in the bulk semiconductor, as the radius of the particle tends to infinity. It also implies that the ionised density must have an upper limit, the donor vacancies being finite in number (and therefore a seemingly ever-increasing density on diminution of grain size not physically possible).

It is clear then, that the methods involved in producing Figure 6-1 are not valid approximations either for the very small or for the very large. Indeed, the accompanying surface state acceptor densities, see Figure 6-2, are troubled with similar deficiencies. The 4nm to 15nm surface density ratio lies at 0.26, so presumably at least one of these two radii falls outside of this method's valid range.

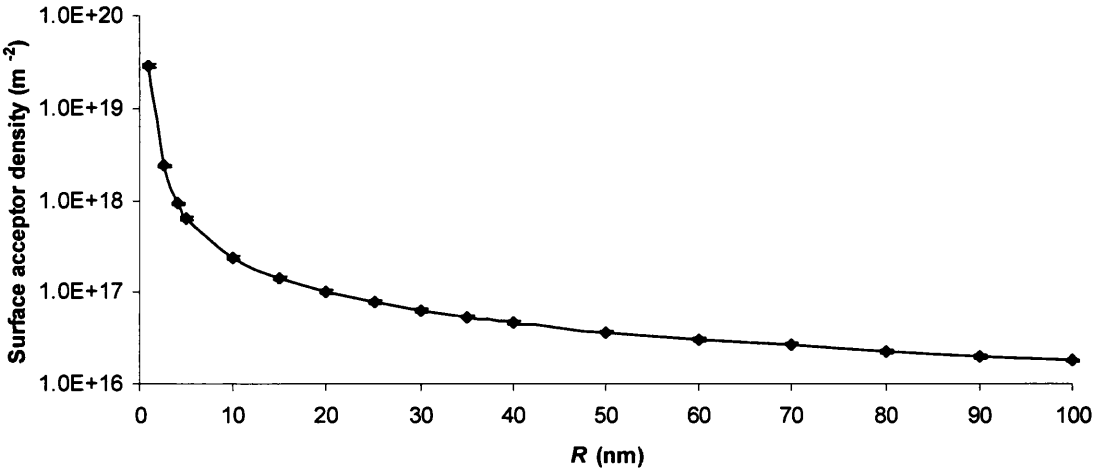


Figure 6-2 Plot of surface state acceptor density against SnO_2 grain radius for the P-CB approximation with a variable ionised donor density. Upper and lower limits of error bars correspond to the surface barrier heights, S_b, 1.3eV and 1.4eV respectively

What would happen if the density of the surface states were required to remain constant over the changes in radii? The charge from the ionised donor atoms and mobile electrons is proportional to the volume of the grain, whilst the charge contained within the surface states is proportional to the surface area of the grain. The consequence of this is that on the decrease of radius, the surface state charge will play an increased role, the trend in its behaviour proportional to the $1/R$ behaviour of the surface area to volume ratio of a sphere, as plotted in Figure 6-3. Even allowing for a surface acceptor density variation similar to that predicted by Malagú *et al.* [15], see Figure 2.12, (although this pattern is itself a consequence of Figure 6-3 for constant n_d) this steady region would still be expected, albeit reached at larger radii.

It is probable then that this lack of a bulk constant region in the P-G data is due to a problem in computation. As the radius increases, the determination of N_s from the gradient of the potential at the grain boundary is likely to become less accurate with the unavoidable increase in mesh spacing (although this can be partially compensated for with a variable mesh). As such, δ in Eq. (5.95) becomes dominated by its δ^{disc} component as R increases, rather than the preferred δ^{der}. In this way, the charge balance process of Chapter 5 is not so much determining the conditions for

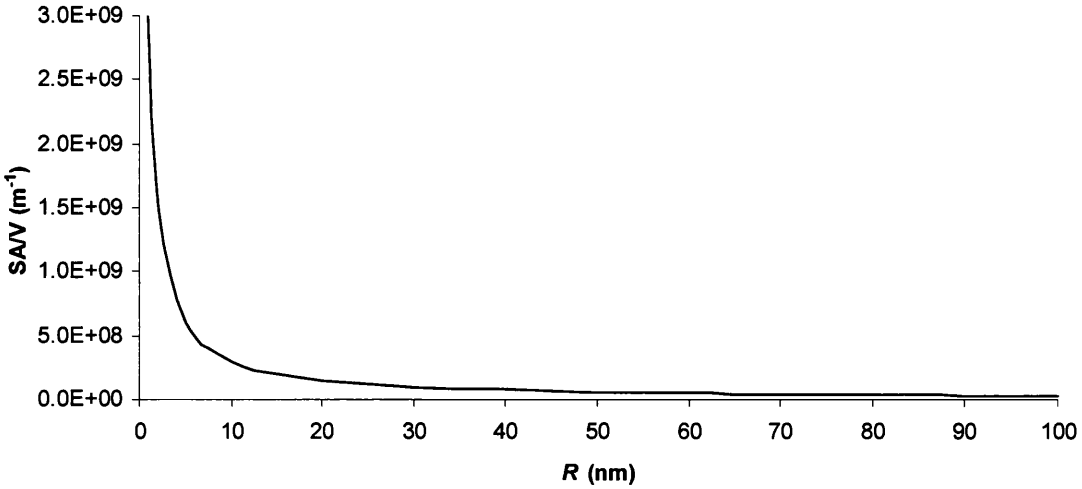


Figure 6-3 Plot of the surface area (SA) to volume (V) ratio of a sphere against radius. SA/V behaviour is proportional to $1/R$.

which both $d\varphi/dr\big|_{r=0} = 0$ and $\varphi(R) = -S_b$ are consistent, but rather tends towards detecting the parameters that minimise δ^{disc}.

An additional factor to consider is of course that at large radii, there might be significant effects to the form of the potential from the accumulation of errors that is inherent in the Taylor expansion method of solving Poisson's equation; either through an increase in the number of mesh points or through the aforementioned increase in mesh spacing. Close inspection of Figure 6-1(a) indicates that the likely region of large R failure in the model is after ~70nm, with a slide from ~$1/R$ to quasi-linear behaviour. This has been highlighted in Figure 6-4, where trendlines

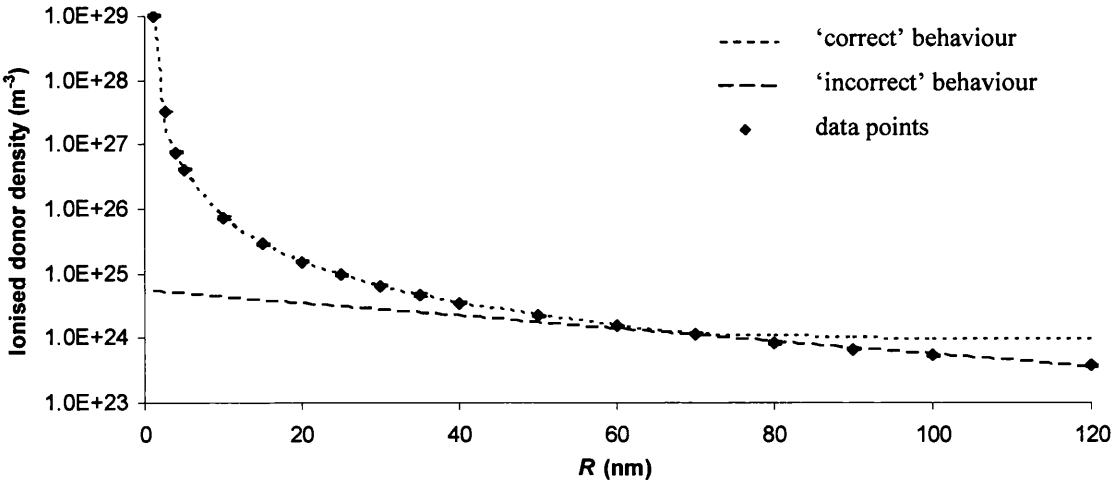


Figure 6-4 Reprint of Figure 6-1(a) of the plot of ionised donor density against SnO_2 grain radius. The range has now been extended to 120nm and trendlines added to indicate the possible switch in behaviour and the probable point of failure of the model at ~70nm.

have been added to Figure 6-1(a) to indicate the possible switch in behaviour and the likely point of breakdown.

This observation and its conclusion for the upper limit of the validity of this model, is supported by Figure 6-5, where the depletion width, Λ, has been plotted against radius for R = 40nm to 100nm.

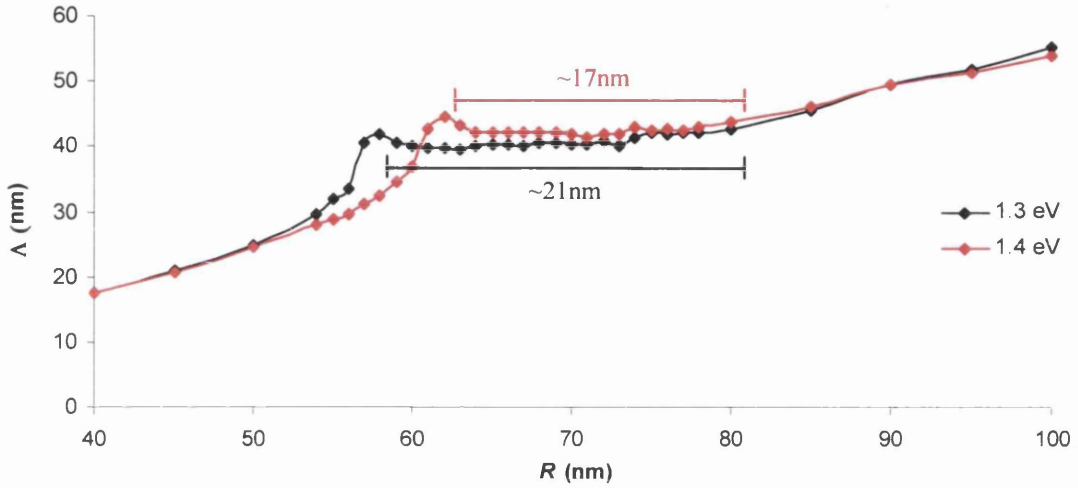


Figure 6-5 Plot of the depletion width, Λ, against SnO_2 grain radius for the P-CB approximation with a variable ionised donor density.

As would be expected, a reasonably constant Λ develops once the grains reach such a size that a field free region can be stably sustained at the grain centre; in this case after small peaks at ~58nm (S_b = 1.3eV) and ~62nm (S_b = 1.4eV), a result of the grains becoming momentarily degenerate. Unfortunately, this region of constant Λ is only sustained for an R span of ~21nm at a mean Λ of 40.6nm for S_b = 1.3eV and ~17nm at a mean Λ of 42.5nm for S_b = 1.4eV, before beginning to drift larger. This departure from the desired (and expected) behaviour, constant Λ in this case, after approximately 80nm is more apparent here than in say Figure 6-1, and with a fair degree of certainty draws the upper limit on the applicability of the non-linear Poisson equation applied with the charge balance equation for this system.

Determining the exact value of Λ is a little open to interpretation. For Figure 6-5 the procedure used defined Λ as the point at which the potential deviated more than a small given percentage of the value of the potential at the grain centre. This was found to be more suitable than relating Λ to any exact value or a percentage of S_b, as in practice, there are variations in the value of φ at r =

0 with respect to n_d, and as such R. In this way the most universally consistent definition of Λ is maintained over all values of R within this coupled Poisson – Charge Balance methodology.

On a positive note, this reasonably constant depletion width, 41.55±0.95nm, lies within the range 1nm to 100nm reported by McAleer *et al.* [43] for 'typical' SnO_2 crystals ($n_d \approx 10^{24}$-10^{26} m^{-3}, $V_b \approx$ 1 eV).

6.1.2 The Analytical Depletion Approximation and the Determination of the Surface State Density and the Ionised Donor Density of the Bulk Semiconductor

A simpler alternative to the P-CB procedure is available when a grain is large enough that a field free region is assumed to exist in its interior $r < R_0$ and the charge from the ionised donors is exactly balanced by the disassociated electrons present. The region of grain outside of this neutral zone is considered to be completely depleted of these charge carriers by the surface states, thus forming a shell around the neutral region that extends to the grain edge, see Figure 6-6. This assumption is known as the depletion approximation (DA), and, as discussed in Chapter 2 and again briefly in Chapter 5, allows an analytic solution.

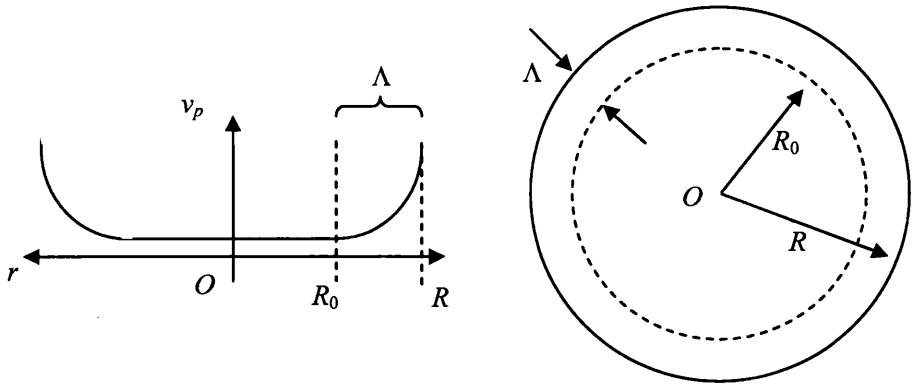


Figure 6-6 Schematic of SnO_2 grain and conduction band bottom, v_{sp}, in the depletion approximation, where the grain is assumed to be large enough that a field free region exists where $r > R_0$, and that a depletion width, Λ, can be defined such that $\Lambda = R - R_0$.

Dividing the spatial regions as in Figure 6-6, then for the depleted shell i.e. the radial interval $R_0 \leq r \leq R$, the linear Poisson's equation in spherical polar coordinates can be expressed as

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dv_p(r)}{dr}\right)=\frac{e^2 n_d}{\varepsilon} \tag{6.1}$$

on recalling Eq. (2.9) and Eq. (2.10) and discarding the electron density. Then on integrating twice, the bottom of the conduction band is described by

$$v_p(r) = \frac{r^2 e^2 n_d}{6\varepsilon} - \frac{c_1}{r} + c_2 \tag{6.2}$$

c_1 and c_2 being the constants of integration. From the behaviour of v_p and its derivative at the boundary of the neutral zone then

$$\left.\frac{dv_p(r)}{dr}\right|_{r=R_0} = 0 \;\Rightarrow\; c_1 = -\frac{e^2 R_0^3 n_d}{3\varepsilon}$$

$$v_p(R_0) = v_p(0) \;\Rightarrow\; c_2 = v_p(0) - \frac{e^2 R_0^2 n_d}{2\varepsilon}$$

allowing the equation governing the v_p to be expressed as:

$$v_p(r) = \frac{e^2 n_d}{\varepsilon}\left(\frac{r^2}{6} + \frac{R_0^3}{3r} - \frac{R_0^2}{2}\right) + v_p(0) \tag{6.3}$$

For the regions of constant Λ, $v_p(R) - v_p(0) = S_b - 0.03\ \text{eV} \approx S_b$, and therefore rearranging Eq. (6.3) yields

$$n_d = \frac{\varepsilon\ S_b}{e\Lambda^2\left(\frac{\Lambda}{3R} - \frac{1}{2}\right)} \tag{6.4}$$

and then, from the requirement of charge equality, balancing the charge within the shell R_0 to R with the charge on the surface of the grain

$$N_t = n_d\left(\frac{R}{3} - \frac{R_0^3}{3R^2}\right) \tag{6.5}$$

follows, finally giving analytic expressions for the ionised donor density and corresponding surface state density at given radius. From Eq.(6.4), maintaining a constant depletion width (and as such, only for radii larger than the specified Λ), Figure 6-7 follows.

It can be seen from this figure that as $R \rightarrow \infty$, n_d tends to a constant level, ${}^{bulk}n_d$, as required by the earlier discussions. For Λ = 40.6nm and S_b = 1.3eV , ${}^{bulk}n_d$ equals $9.85\times10^{23}\text{m}^{-3}$, and for Λ = 42.5nm, S_b=1.4eV, ${}^{bulk}n_d = 9.68\times10^{23}\text{m}^{-3}$. This is an identical trend to that seen in Figure 5-11, which was obtained with the same procedure, see Section 5.6, but for different defining parameters.

To offer a brief summary, essentially the mean of the two ${}^{bulk}n_d$ values, $(9.76\pm0.08)\times10^{23}\ \text{m}^{-3}$, is

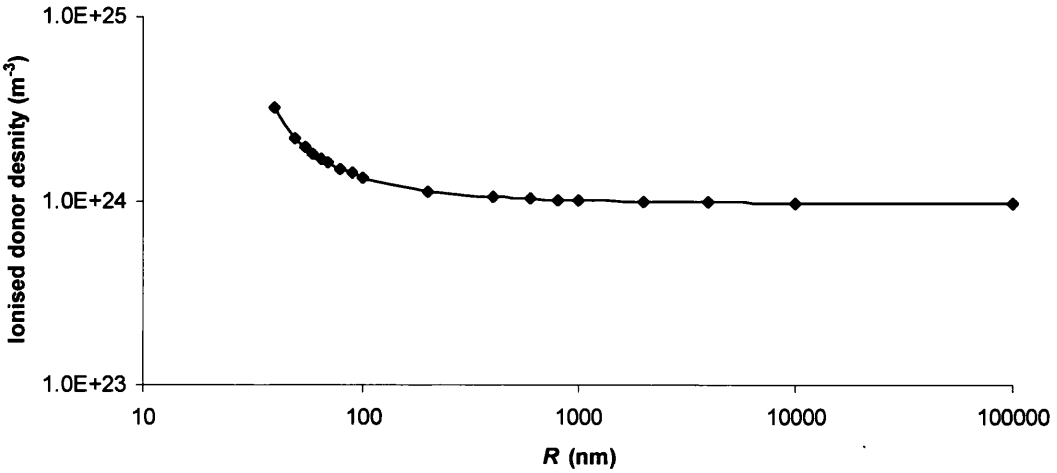


Figure 6-7 Plot of ionised donor density against SnO_2 grain radius for the depletion approximation analytic solution. Upper and lower limits of error bars correspond to the surface barrier heights, S_b, 1.3eV and 1.4eV with their corresponding depletion widths, Λ, of 40.6nm and 42.5nm, respectively (following from Figure 6-5).

taken as the true value of the ionised donor density of the 'bulk' semiconductor. The constant depletion width follows from the complete charge density model (CCDM) (non-linear Poisson equation) coupled with the charge balance equation, and the final bulk ionised donor density figure is the application of this constant Λ within the analytic expressions of the depletion approximation, and as such, can be treated as exact as the grain radius tends to infinity.

It has been experimentally determined [53,96] that for bulk SnO_2, the doubly ionisable oxygen vacancies introduced through annealing obey the relationship

$$ {}^{bulk}n_d = \frac{N_d}{1+g_1 e^{\frac{(E_d^1-E_f)}{k_bT}}} + \frac{N_d}{1+g_2 e^{\frac{(E_d^2-E_f)}{k_bT}}} \qquad (6.6) $$

where $g_1 = 2$, $g_2 = 0.5$ and the energies of the first and second ionised levels of the oxygen vacancy (E_d^1 and E_d^2) taken relative to the conduction band bottom are 0.34eV and 0.145eV respectively [97[ii]]. Ergo, at room temperature for ${}^{bulk}n_d = (9.76\pm0.08)\times10^{23}m^{-3}$, the total

[ii] As an interesting aside, should any reader wish to refer to this paper of Carotta, Dallara, Martinelli and Passari [97] a few points are worthy of note. From their surface electron density of $1.5\times10^{21}m^{-3}$ (printed as $1.5\times10^{15}cm^{-3}$) the Hall constant is in fact $4.16\times10^{-3}m^3/C$ (not their stated $4.16\times10^{-3}cm^3/C$), and that using the correct expression for the electron density (see for example Eq. (1.34) or Eq. (1.41), not their printed expression), then the Fermi level and bulk electron density are given by E_f = -0.0671eVand n_b = $3.10\times10^{24}m^{-30}$ - which are the two values reported (or very close to those values at least) by the authors. However, the donor density N_d is then $2.881\times10^{25}m^{-3}$, not their reported value of $4.11\times10^{24}m^{-3}$, and the

concentration of donors must be $N_d = (7.91\pm0.07)\times10^{24}m^{-3}$. Each oxygen vacancy is doubly ionisable, and as such, the ionised donor density must always, when bulk statistics are no longer applicable, be less than or equal to twice the total donor density i.e. $n_d \leq (1.58\pm0.01)\times10^{25}m^{-3}$.

Consequentially, the maximum value of n_d calculated from the depletion approximation as applied here, $n_d = 3.18\times10^{24}m^{-3}$, is consistent with this maximum limit. However, as already qualitatively stated (and not withstanding the significant error in the 4nm-15nm N_s ratio), the behaviour of n_d in Figure 6-1 – generated using the non-linear Poisson equation and the charge balance equation only – on diminishing R is clearly unphysical with the upper limit on n_d being exceeded at around 24.5nm. Clearly then, if the outer iteration charge balance method and the idea of variable ionisation density are to stand up to scrutiny, then not only must the effects of quantisation applied within the charge balance scheme correct the 4nm-15nm ratio, they must also prohibit n_d exceeding $(1.58\pm0.01)\times10^{25}m^{-3}$.

The corresponding graph to Figure 6-7 for surface acceptor density, N_t, is Figure 6-8 generated

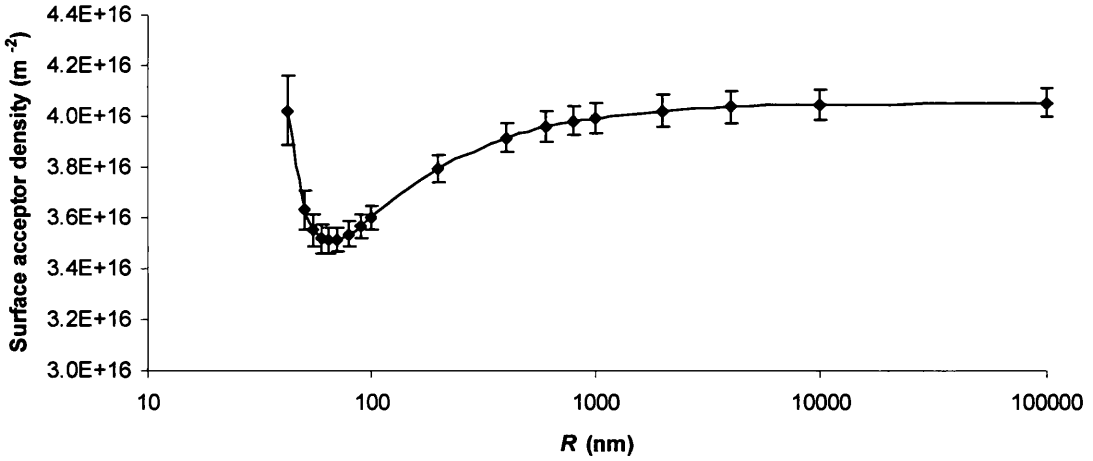


Figure 6-8 Plot of surface acceptor density against SnO_2 grain radius for the depletion approximation analytic solution with a variable ionised donor density. Upper and lower limits of error bars correspond to the surface barrier heights, S_b, 1.3eV and 1.4eV with their corresponding depletion widths.

---

value of the electron mobility, μ, is likely to lie in the range 6.70×10^{-4} to $6.70\times10^{-5}m^2/Vs$, not their stated $1.6\times10^{-5}m^2/Vs$. The value of μ at 350°C cannot be exactly calculated for this footnote, as only data for G_0 at 200, 300, 400 and 480°C is tabulated. Some of the authors of this paper [97] contributed to Malagú *et al.* [15] (and so to Malagú *et al.* [16,17]), and the referenced source of the N_d used there ($5\times10^{24}m^{-3}$) was unfortunately Ref. [97].

from Eq. (6.5). Again, this is applicable only to grains sufficient in extent to possess the given depletion width. It is found that as $R \rightarrow \infty$, N_t tends to $(4.05\pm0.06)\times10^{16}$ m^{-2} in qualitative agreement with that inferred by McAleer *et al.* [43] from experimental work ($N_t \approx 10^{17}m^{-2}$ for ${}^{bulk}n_d \approx 10^{24}$-$10^{26}m^{-3}$, $V_b \approx$ 1eV).

Perhaps the most interesting feature of Figure 6-8 (in contrast to Figure 5-7 and more specifically the following Figure 6-9) is the influence of the rising ionised donor density on diminishing R in reversing the downward trend in the density of occupied surface acceptor states. N_t rises to $(4.02\pm0.14)\times10^{16}m^{-2}$ by R = 42.5nm (smallest radius possible within the upper limit on the depletion width), approximately equal to its value in the bulk. This increase of N_t below ~ 65nm (N_t at its minimum of $(3.51\pm0.06)\times10^{16}m^{-2}$) is unlikely to be found to continue beyond R = 42.5nm, and in all probability, even N_t at this point will be found to be overestimated (lying at the extreme of the applicable range of the DA), else this model will contradict the experimental results of Williams and Coles [41]. This paper will be discussed in more detail at the end of this section, but essentially, they report a decrease in the sensitivity of SnO_2 polycrystalline gas sensing films from 20nm diameter to coarser, micron sized grains, implying an increase from the occupied surface state density of the 20nm to the 'bulk' grains.

For the purposes of illustration, should the ionised donor density be maintained at a constant level through a range of grain sizes, neglecting the implications of an inconsistent $d\varphi/dr\big|_{r=0} = 0$ and $\varphi(R) = -S_b$ other than to set the constant n_d at the calculated bulk semiconductor value, ${}^{bulk}n_d$, then a pattern similar to that predicted by Malagú *et al.* [15] develops, see Figure 6-9. The bulk electron density approximation, non-linear Poisson equation (complete charge density), has a surface density ratio for the 4nm to 15nm grains of 3.75 (1.3-1.3eV and 1.4-1.4eV ratios equal to seven decimal places), but fails to converge after ~220nm. The analytic DA (using Eq. (6.5) only, with both constant n_d and constant Λ as specified in Malagú *et al.*'s Ref. [15]) on the other hand, is not appropriate much less than 50nm, although without upper limit. The actual value of N_t for the range ~70nm to ~220nm is likely to lie between the two approximations, given the aforementioned difficulties in assessing N_t for large meshes (P) and the impact of the simplifications involved in the DA approach, especially at the smaller end of its effective range. For this region of overlap, some discussion is necessary.

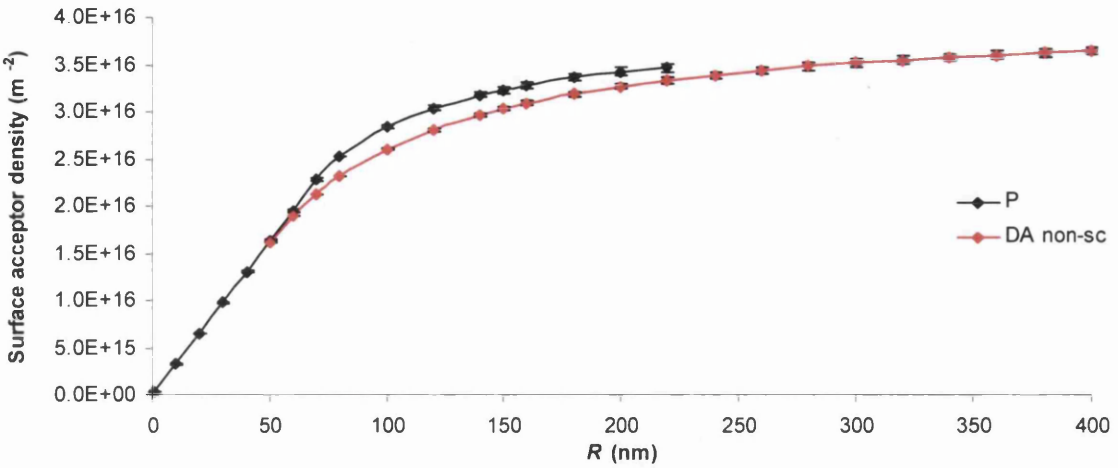


Figure 6-9 Plots of surface acceptor density against SnO_2 grain radius for both the non-self consistent bulk electron density approximation (P) and the analytic depletion approximation (DA). Upper and lower limits of error bars correspond to the surface barrier heights, S_b, 1.3eV and 1.4eV with their corresponding bulk ionised donor densities of $9.85\times10^{23}m^{-3}$ and $9.68\times10^{23}m^{-3}$, respectively. The analytic DA also uses the accompanying depletion widths, Λ, of 40.6nm and 42.5nm, respectively (following from Figure 6-5).

At grain radii of <60nm, the N_t's of both the depletion approximation series and the complete charge model are very close, implying that the effects of the thermal electrons are minimal on the surface state density at these radii.

As the radius increases, the volume of the shell depleted of charge carriers also increases, in proportion to R^3-R_0^3, and consequentially the charge on the surface of the grain must also increase in line with this. The effect then of neglecting the charge of the mobile charge carriers within this shell will also be greater since the shell volume is greater. This acts to increase the magnitude of the N_t from the DA over what it would otherwise be inclusive of the mobile electrons.

Running concurrent to this is another mechanism due to the fact that ${}^{bulk}n_d$ is an under-estimation of the 'equilibrium' P-CB ionisation density at radii less than ~100nm. As a consequence, to maintain the surface barrier height, the electron density is lower on the grain boundary in this CCD model than it would be for higher n_d. Indeed, the whole electron density is lower over all, both in absolute magnitude and in relation to the relative amount of positive charge present due to the ionised donors. This affects N_t, increasing its magnitude in this CCD model over the corresponding 'equilibrium' P-CB values. The play off between these two factors is the likely cause of the variation between the DA and the bulk approximation CCD series'.

Nevertheless, returning to the main point, despite the improved estimate of the 4nm-15nm occupied surface density ratio of 3.75 (experimental value is 5.90 [31]), and its consistency with the inferred surface density increase between 20nm diameter and coarse grains of Williams and Coles [41], this illustrative Malagú *et al.*-like example of constant ionised donor density can only be just that – illustrative. Neither the DA, nor the non-linear Poisson equation model, as applied in Figure 6-9, are truly consistent within their own respective schemes.

Only Figure 6-2 for the non-linear Poisson equation, and Figure 6-8 for the DA, are consistent methodologies, although each is only accurate within a specific range of radii for these defining parameters (DA: $R \geq 42.5$nm; P-CB: ~24.5nm $\leq R \leq$ ~80nm – lower limit likely to actually be much higher, 24.5nm being the point at which n_d exceeds its maximum allowed value).

6.1.3 Full Non-Linear Poisson - Kohn-Sham - Charge Balance (P-KS-CB) Treatment and the Determination of the Surface State Density Ratio of 4nm to 15nm Radii Grains

To move forward in the simulation of these grains, the full Poisson-Kohn-Sham scheme needs to be brought to bear, modelling the discrete energy spectrum with the complete charge density, and including the electron-electron interaction effects through KLI exchange and LDA correlation (as described in Chapter 4).

Accepting, firstly, that the non-linear Poisson equation in conjunction with the charge balance equation is likely to prove a good model for radii less than ~80nm down to some lower limit, the very lowest possible value of which would be ~25nm, after which quantisation effects are dominant. Secondly, that the analytical depletion approximation, using the constant depletion width of 41.55±0.95nm inferred from the P-CB scheme, is a good model for radii greater than ~43nm, increasing in accuracy as $R \rightarrow \infty$. Then the computationally expensive Poisson-Kohn-Sham method need only be applied to model small radii grains, where the effects of quantisation must be included.

In order to do this, Poisson's equation and the Kohn-Sham equations must be self-consistently solved, in conjunction with the charge balance equation, equalising the charge in the surface states with that contained within the grain. By cycling through these three parts, a stable, consistent set of values can be achieved.

However, this is not a straightforward process. In addition to the influence of δ^{KS} now also contributing to δ in Eq. (5.95) (due to failings within the inner self-consistency cycles (Poisson-Kohn-Sham), see Section 5.5), the movement of the conduction band bottom (and to a lesser extent, the variation of the grains' work function, see Section 2.4) can cause the number of permissible energy levels to suddenly vary for an infinitesimal change in ionised donor density. This can result in discontinuities in the variation of the convergence factor, R_{ch}; examples of which are found in Figure 6-10 for the sample radii 4nm, 8nm, 10nm and 15nm.

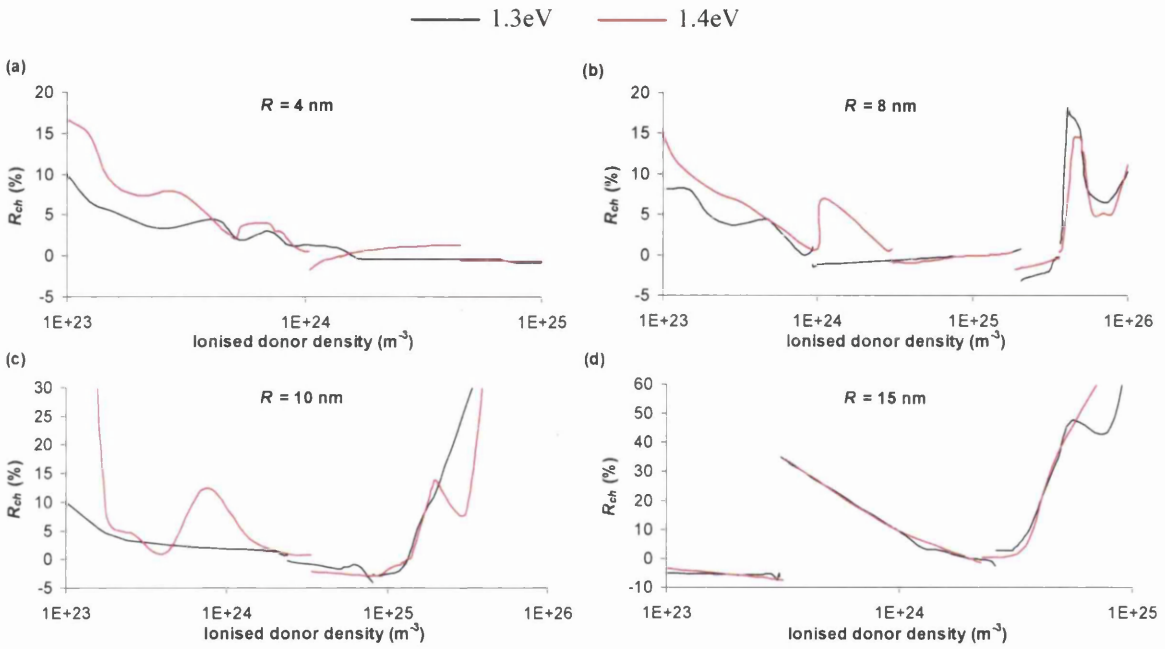


Figure 6-10 Plots of the convergence factor, R_{ch}, against the ionised donor density for SnO_2 grains of radii (a) 4nm (b) 8nm (c) 10nm and (d) 15nm, for both minimum and maximum surface barrier heights, S_b.

R_{ch} expresses the charge deviation $\delta (= q_1 - q_2)$ as a percentage of the electron density, see Section 5.6, and is defined by Eq. (5.98)

$$R_{ch} = 100 \times \frac{q_1 - q_2}{q_1}$$

where

$$q_1 = 4\pi e \int_0^R n_d e^{e\varphi / k_b T} r^2 dr$$

$$q_2 = \frac{4\pi}{3} e R^3 n_d - 4\pi R^2 \frac{\varepsilon_r \varepsilon_0}{e} \left. \frac{dv_{sp}}{dr} \right|_{r=R}$$

Both series in each plot of Figure 6-10 are typically made up of between 80 to 120 points, non-uniformly spaced, increasing in concentration as $R_{ch} = 0$ is approached (due to the bisection

method used in satisfying the charge balance equation, see Section 5.6). In this way, although discontinuities about this value are readily detected, their presence at higher R_{ch} may go unobserved. Figure 6-10 is illustrative of the difficulties encountered in obtaining the true consistent solution sets.

Simulating anything other than small systems can be difficult. The larger the grain radii, the greater the number of energy levels and so modelling systems much larger than ~20nm with these material and operating parameters becomes computationally prohibitive.

However, within this 20nm limit, as desired and plotted in Figure 6-11, the ionised donor density increases from $(1.554\pm0.004)\times10^{24}m^{-3}$ at 20nm, as the radius decreases, up to a maximum of $\sim(1.29\pm0.03)\times10^{25}m^{-3}$ at 10nm, driven by the increasing surface area to volume ratio (the increased effect of the density of charge in the surface states requiring a corresponding increase in n_d to maintain neutrality), and then decreases after this as quantum effects completely dominate, the form of the spatially localised electrons' probability distributions profoundly affecting the potential at the grain surface. In general, for $R > \sim2$nm, the discrete energy spectrum maintains a higher electron density at equal donor ionisation than the bulk continuum (P-CB) model.

The fact that the peak n_d lies below its theoretical maximum of $(1.58\pm0.01)\times10^{25}m^{-3}$, arrived at by what is essentially an independent calculation, see preceding sub-sections, is good evidence in favour of the validity of Figure 6-11.

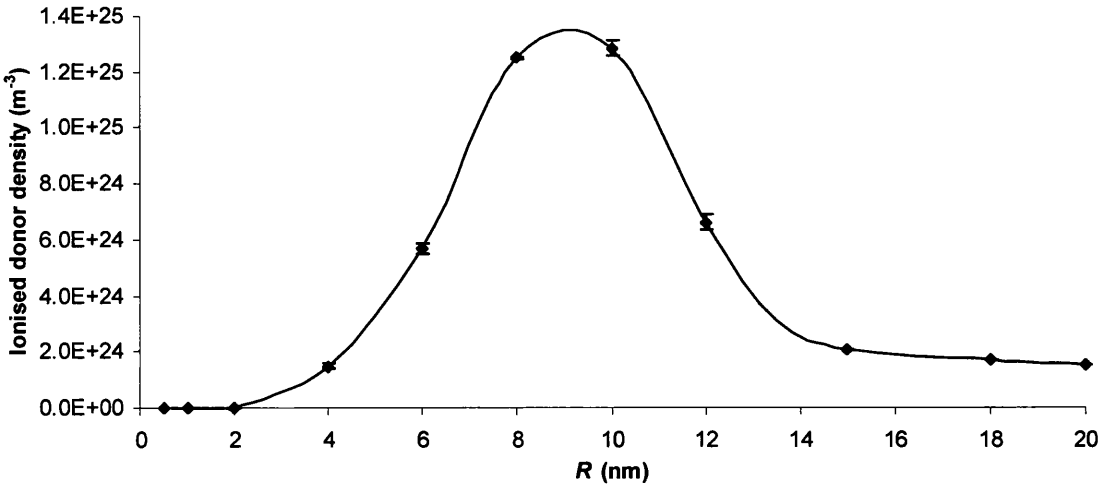


Figure 6-11 Plot of ionised donor density against SnO_2 grain radius. Upper and lower limits of error bars correspond to the surface barrier heights, S_b, 1.3eV and 1.4eV respectively

Below a radius of ~0.5nm the quantum well is too narrow to possess any discrete energy levels between the conduction band bottom and the value of the effective work function. As such, although this model then ceases to be applicable, without an electron population contributing to the charge density, an analytical solution is once again possible[iii], the apparent inherent contradiction of Eq. (5.93) aside.

As in the depletion approximation calculations, integrating Eq. (6.1) twice, but now applying the boundary requirement that at $r = 0$ $dv_p / dr\big|_{r=0} = 0$ as well as the usual $v_p(R) = S_b$, then

$$v_p(r) = \frac{en_d}{6\varepsilon_r\varepsilon_0}\left(r^2 - R^2\right) + S_b \tag{6.7}$$

and as such

$$N_t = \frac{Rn_d}{\varepsilon_r\varepsilon_0} \tag{6.8}$$

Indeed, without allowable eigenstates within the grain, then as extensively discussed in Chapter 5, the solution of the linear Poisson equation is uniquely specified with these boundary values, the charge balance equation automatically satisfied for all values of n_d. The disassociated electrons from the ionised donors' all required to reside within the surface acceptor states. Needless to say, the proposal of this and the preceding Chapter is that n_d is not constant at these small radii, and this hypothesis does certainly apply to Eq. (6.7). Therefore, without the means of determining n_d, unless it has been measured experimentally at the relevant radii, Eq. (6.7) and Eq. (6.8) are unfortunately of little practical use[iv]. See also the analytical $R \to 0$ limiting case discussed in Section 5.5.

Plotted in Figure 6-12 is the corresponding movement of the surface acceptor density for the ionised donor densities of Figure 6-11. From ~0.5nm the surface density increases with increasing radius to its maximum at ~10nm where $N_t = (3.95\pm0.08)\times10^{16}\text{m}^{-2}$, falling after this

[iii] Arguments regarding the applicability of such things as an energy independent effective mass to solids of such a small extent aside.

[iv] Although, interestingly, it could be argued that without any energy levels and therefore with no 'conduction band', unless a surface state was close enough in physical proximity to a donor flaw that an ionised electron could directly 'move' into it from the donor, then the donor would be unable to ionise at all i.e. $n_d \to 0$ when $R < 0.5$nm for these parameters. In this case, Eq.'s (6.7) and (6.8) are applicable, and v_p will tend to S_b for all r and N_t will tend to zero. Ergo, the analytical model would support the behaviour of the full quantum mechanical P-KS-CB model.

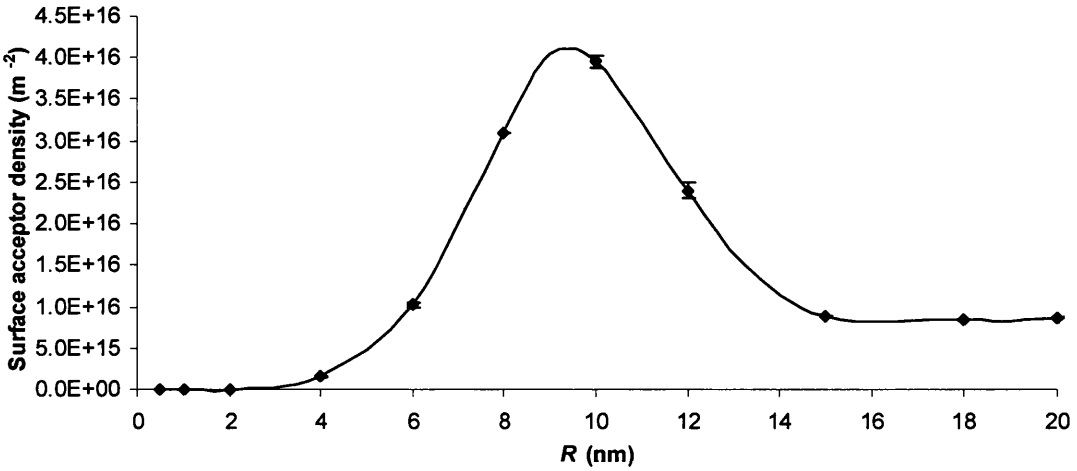


Figure 6-12 Plot of occupied surface acceptor density against SnO_2 grain radius. Upper and lower limits of error bars correspond to the surface barrier heights, S_b, 1.3eV and 1.4eV respectively.

until it drops to $(8.46\pm0.06)\times10^{15}m^{-2}$ at 18nm, before slowing beginning to rise once more. The peak surface acceptor density at ~10nm lies very close to that calculated for the bulk semiconductor, $(4.05\pm0.06)\times10^{16}m^{-2}$. Very importantly, this precise P-KS-CB data places the 4nm to 15nm ratio between 5.20 (S_b's at 1.3eV) and 5.93 (S_b's at 1.4eV) – in excellent agreement with experiment (5.90).

In Figure 6-13, both the ionised donor density, Figure 6-13(a), and the surface acceptor density, Figure 6-13(b), are plotted. Grain radii spanning 1nm to 100 000nm (0.1mm) are covered for the three different (consistent) modelling schemes: Poisson – Kohn-Sham-Charge balance (P-KS-CB) (see also Figure 6-11); Poisson-Charge balance (P-CB) (see also Figure 6-1); and the depletion approximation (DA) (see also Figure 6-7). The depletion approximation is plotted twice; once with the depletion width maintained at 41.55±0.95nm varying n_d (DA [n_d]), and also with constant n_d (set at ${}^{bulk}n_d$) varying Λ (DA [Λ]). As usual, the upper and lower limits of the error bars correspond to the surface barrier heights of 1.4eV and 1.3eV, with the accompanying bulk ionised donor densities of $9.68\times10^{23}m^{-3}$ and $9.85\times10^{23}m^{-3}$ respectively for the DA series'.

Quantitatively, the three models line up well at the edges of their respective applicable ranges, although there is a step when switching from P-KS-CB to P-CB after a radius of ~20nm. The difference between the two DA methodologies is only apparent for radii under 1 000nm, and most interestingly, the limit of each is found on, or very near, the point of their respective intersections with the series P-CB - ~43nm for DA [n_d] and ~75nm for DA [Λ] (this can be seen with more

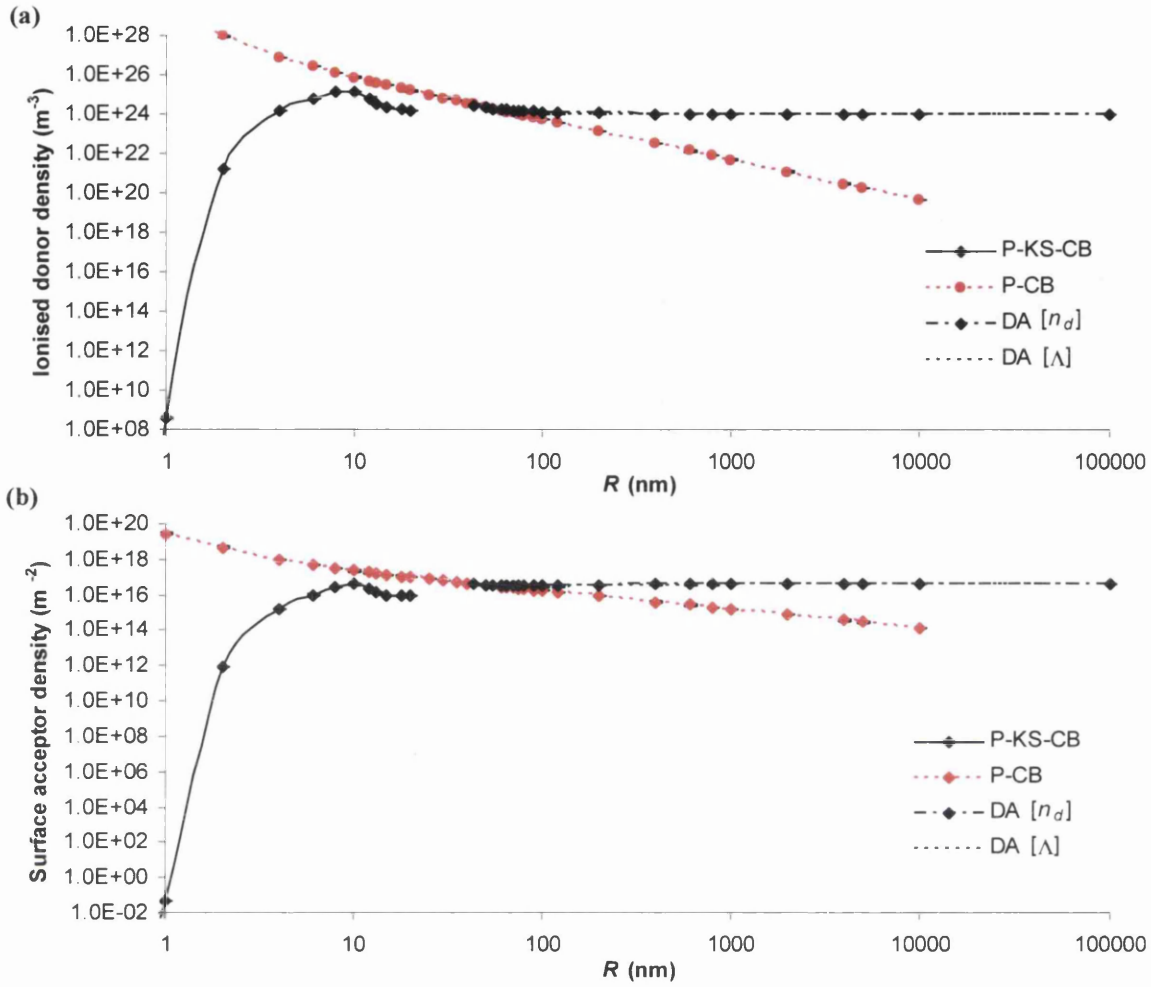


Figure 6-13 Plot of (a) ionised donor density, and (b) surface acceptor density against SnO_2 grain radius, in the range 1-100,000nm. Upper and lower limits of error bars correspond to the surface barrier heights, S_b, 1.3eV and 1.4eV respectively. Both plots are composed of data from the three consistent schemes: Poisson – Kohn-Sham - Charge Balance (P-KS-CB), Poisson - Charge Balance (P-CB) and the Depletion Approximation (DA). DA [n_d] maintains a constant Λ and varies n_d, and vice versa for DA [Λ].

clarity in Figure 6-14). If it were not for the region of quasi-constant Λ for P-CB (see Figure 6-5), it could be suggested that the behaviour of P-CB drifts from variable n_d, constant Λ at ~43nm to constant n_d, variable Λ at ~75nm (and onwards to computational failure – see Sub-Section 6.11). Indeed with this hypothesis, for radii < 75nm variation in the n_d of P-CB would be apparent, and at 75nm itself will equal $(9.75\pm0.34)\times10^{23}$ m^{-3} – the DA [Λ]'s n_d or DA [n_d]'s $^{bulk}n_d$ (= $(9.76\pm0.08)\times10^{23}m^{-3}$). In this way, the action of the P-CB electron density (absent in the DA) seems to provide the appearance of this drift in behaviour for the P-CB series with its constant Λ for this radii range.

The P-CB methodology is an approximation flawed both by its computational implementation in its upper limit (~70 / 80nm), and the inadequacy of the form of its electron density toward its lower limit (which is clearly much greater than 24.5nm – see Sub-Section 6.1.2). The above interpretation of the ~43 to ~75nm behaviour of its electron density (facilitating the appearance of a shift in n_d behaviour (when compared to the DA trend) from variable to constant whilst maintaining a constant Λ throughout) can be seen, if nothing else, as a positive contribution to the overall argument for variable n_d. It is not, perhaps, unjust to say that the values of the P-CB model are conceivably more qualitative than exact in its applicable range. These issues of accuracy are in no doubt due to the aforementioned problems, although it seems that the quantisation of the energy spectrum is significant for the entire range of radii where the charge due to the electron presence cannot be neglected. It is possible that computational inaccuracies are also considerable before this point of negligible electron density is reached.

This issue of the P-CB model's accuracy in mind, if the trend in n_d for DA [n_d] is taken as exact as $R \rightarrow \infty$, and that the trend in n_d for P-KS-CB is also exact, then it is likely that the behaviour of n_d in region after 20nm (computational limit of P-KS-CB) but before the presence of the electrons in the depletion width can be totally neglected, will favour a smooth transition from ${}^{P-KS-CB}n_d\big|_{R=20nm}$ to a value equal to it on the DA [n_d] profile, rather than perhaps the path evident from the P-CB profile. An even more likely alternative is that n_d would tend to a value less than the P-KS-CB's n_d at R = 20nm on the DA [n_d] profile, following the trend of P-KS-CB greater than 10nm. The ionised donor density of the P-KS-CB series at 20nm ($1.55\times10^{24}m^{-3}$) is reached at a DA [n_d] radius of 75nm. However, following the trend of the P-KS-CB profile and smoothly melding n_d into the DA [n_d] trend, (exact equality reached at R = 200nm, $n_d = 1.09\times10^{24}m^{-3}$) Figure 6-14 is the outcome.

The variation in n_d is represented in Figure 6-14(a) and the corresponding movement of N_t is plotted in Figure 6-14(b). Unfortunately, the comparatively small step in the ionised donor density between P-KS-CB and DA schemes is translated at this radius into a much larger step in the occupied surface acceptor density. However, the inferred movement of n_d produces a smooth trend in N_t bridging the gap, reminiscent of the N_t movement of constant n_d – for example see Figure 6-9. Although it should always be emphasised that when considering the data of these three schemes and the transitions between models, it is important to remember that while both the P-CB and DA models are complementary classical schemes (hence the smooth 'passover' of

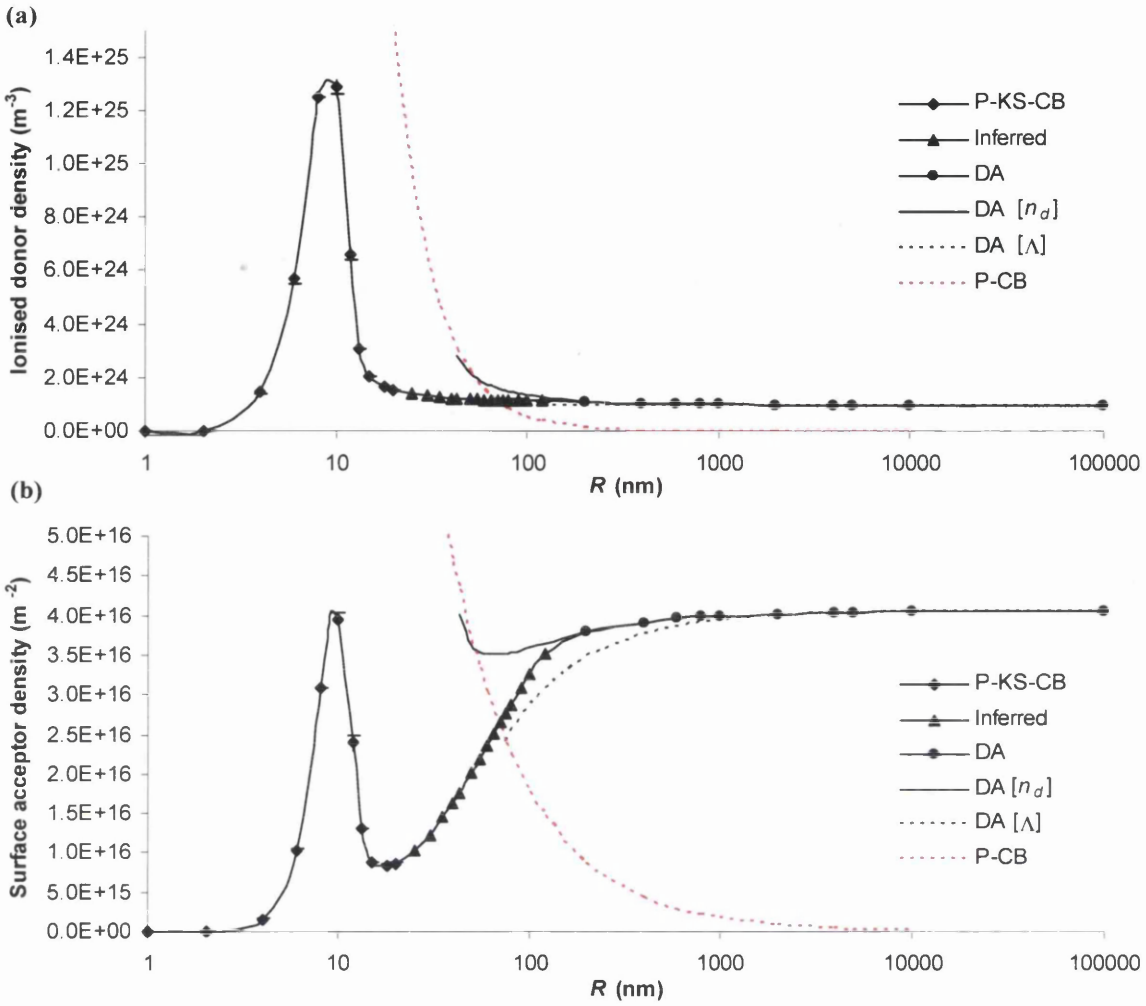


Figure 6-14 Plot of (a) ionised donor density, and (b) surface acceptor density against SnO_2 grain radius, in the range 1-100,000nm. Upper and lower limits of error bars correspond to the surface barrier heights, S_b, 1.3eV and 1.4eV respectively. Both plots are composed of data from the three consistent schemes (P-KS-CB, P-CB and DA (both DA [n_d] and DA [Λ] – constant Λ and n_d respectively)) plus an inferred movement of n_d with its corresponding N_t trend – see accompanying text.

P-CB to DA[n_d]), the P-KS-CB model is not. Instead it is a quantum mechanical treatment of the system, and as such, the reader should not dwell overly on the P-KS-CB / P-CB step, the emergence of classicality from quantum theory one of the longest standing conundrums in quantum mechanics [98].

In Figure 6-14, the mean values only of the P-CB, DA [n_d] and DA [Λ] trends have been plotted, with only the numerical values proposed as significant in modelling the actual densities denoted with a marker and its accompanying error bars.

6.1.4 Further Thoughts

The main supporting evidence up to this point for the theoretical results produced from this concatenation of modelling schemes is three-fold. The surface state density and depletion width for the bulk grains are comparable with that inferred from experiment, see McAleer *et al.* [43], and most importantly, the surface state density ratio between 4nm and 15nm radii grains of 5.57±0.37 is in excellent agreement with the measured ratio of 5.90 [31].

Some mention has been made of the work of Williams and Coles [41]. These authors measured the sensitivity of three sizes of SnO_2 particle over a range of temperatures to three test gases (H_2, CO and CH_4). Two sizes of spherical nanoparticle – one where the median diameter was 20nm and the other 8nm – were generated by vaporising tin dioxide with a laser in either air (20nm) or Argon (8nm) and sintering (heating in air) at 673K. The third sample was prepared using conventional chemical techniques. It consisted of micron (1000nm) size grains, although each grain was a conglomerate of smaller 20 – 100nm diameter crystals. On comparing their resistance with and without the presences of the test gases, they reported:

> "It appears that both nanocrystalline SnO_2 samples are significantly more sensitive than the conventional powder to each of the three test gases. In addition, by decreasing the particle size of the nanocrystalline sample, a marked increase in sensitivity can be achieved, especially in the case of hydrogen."

Unlike the work of Maffeis *et al.* [31] where the grains of different sizes were prepared in identical conditions, and enough of their defining parameters measured to simulate their behaviour with some accuracy, the barrier heights and work functions of the various sized William and Coles grains are likely to vary (see Section 2.4), and enough information about their final electronic properties is not known to construct a precise model. However, assuming that their behaviour is at least qualitatively comparable with the results of this Chapter, further support of the theoretical models can be offered.

The grains with a median radius of 10nm will lie around the 10nm surface density peak of Figure's 12, 13(b) and 14(b), and as such will have an average surface state density slightly below the value of this local maxima. Consequentially, the surface state density of the coarser sample will be at least 2.5 % greater than that of the 10nm radii grains (endowing these micron aggregate grains with bulk characteristics and not treating them as individual 10nm – 50nm radii grains, whose mean surface density (assuming an even distribution of sizes) would be less than the 10nm radii grain alone). These coarse grains will also have a surface state density approximately ~25.5 times that of the 4nm radii grains.

Equating a decrease in the occupied surface state density with an increase in gas sensing sensitivity, see Section 2.4, this implies that as experimentally observed, not only will the sensitivity of the grains increase on decreasing size, but the 4nm radii grains will be considerably more sensitive than even the 10nm radii ones.

In contrast, for a constant ionised donor density model, whilst the surface state density still decreases on decreasing radii, referenced to the bulk value of N_t, the 4nm and 10nm radii grains have surface state density ratios of ~12.5 and ~31 respectively. This means that the expected increase in sensitivity of the 4nm over the 10nm radii grains would only be a factor of 0.5 greater than the bulk – 10nm sensitivity increase.

Ergo, though not exact and certainly not conclusively, it would appear that the results of Williams and Coles [41] (see their Figure 3, plots (a) to (c)), would in general favour the variable ionised donor density methodology over the constant one, based on the behaviour of the surface density alone. It is possible, however, to offer a more quantitative comparison of the constant versus variable ionised donor density models for these gas sensing film sensitivity results by extending the work of Section 2.4.

Recalling Eq. (2.53)

$$S = \frac{n_{gas} e^{eV_s^{air}/k_B T}}{n_{air} e^{eV_s^{gas}/k_B T}}$$

based on a Schottky model of the granular conductance, then the premise of Section 2.4 was that the maximum possible sensitivity of the gas sensor, independent of the exact reducing gas, concentration and all other factors, would be given by a total return of all the charge carriers trapped within the surface states to the conduction band, with an implied disappearance of V_s^{gas}. This change in the conduction band electron density implies a change in the position of the Fermi level. While this was largely immaterial with Malagú *et al.* [15]'s simulation in Section 2.4 where the donor vacancies were assumed to be totally ionised, here, whether 'constant' or not, this Fermi level change will cause a change in the number of ionised donors from Eq. (6.6).

Take the 'constant' ionised donor case first. Working in the middle of William and Coles [41]'s temperature range, say at 673K, then from Eq. (6.6) with S_b = 1.35eV and $N_d = 7.91\times10^{24}\text{m}^{-3}$, the ionised donor density of the bulk in a clean air environment will be $2.84\times10^{24}\text{m}^{-3}$. From the

complete charge density model the electron density n_{air} can be determined, then inverting Eq. (1.34) with the Nilsson equation [57], accurate to within 0.5%, the Fermi level can be seen to drop to -0.08eV when all the electrons are to be found in the conduction band. From Eq. (6.6) this implies that the ionised donor density will be reduced to $8.34\times10^{23}m^{-3}$, constant for all radii. Using this, Figure 6-15 shows that the diminution of the carrier concentration in air below Λ_D ($\approx$ 42nm) causes a steep rise in the ratio n_{gas}/n_{air} but this does not totally follow through to S_{max}, see Figure 6-16 , the maximum sensitivity rising to 40nm on diminution of R, but then falling almost as sharply as it rose as $R \rightarrow 0$.

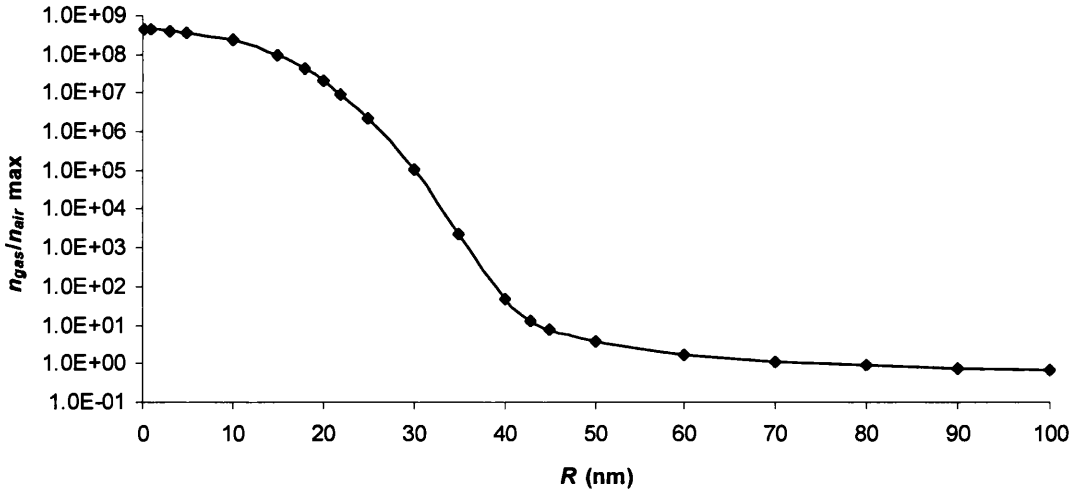


Figure 6-15 Plot of the electron density ratio n_{gas}/n_{air} against SnO_2 grain radius for a 'constant' ionised donor density of $2.84\times10^{24}m^{-3}$at 673K, S_b = 1.35eV.

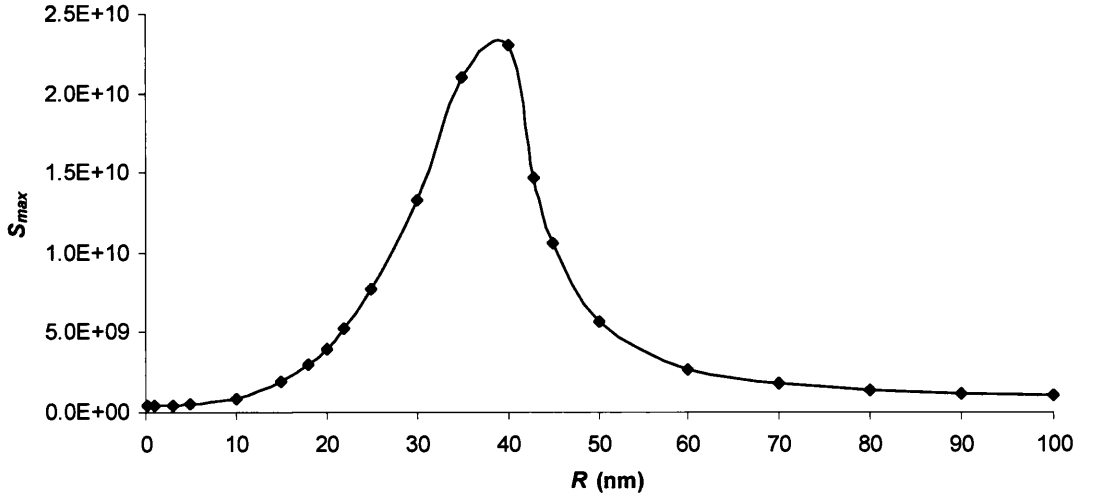


Figure 6-16 Plot of the theoretical maximum sensitivity S_{max} against SnO_2 grain radius for a 'constant' ionised donor density of $2.84\times10^{24}m^{-3}$at 673K, S_b = 1.35eV.

Clearly, the actual sensitivity would be vastly lower than this, and gas, concentration, exposure time etc. dependent, but the general trend should still be relevant (again neglecting any R dependency on permeability and the like).

For the variable ionised donor density, matters are more complex. For this brief diversion, the P-KS-CB scheme only will be applied. Working on a small mesh and relaxing slightly the convergence criterion, the preliminary results for the $T = 673$K, $S_b = 1.35$eV system are plotted in Figure 6-17 .

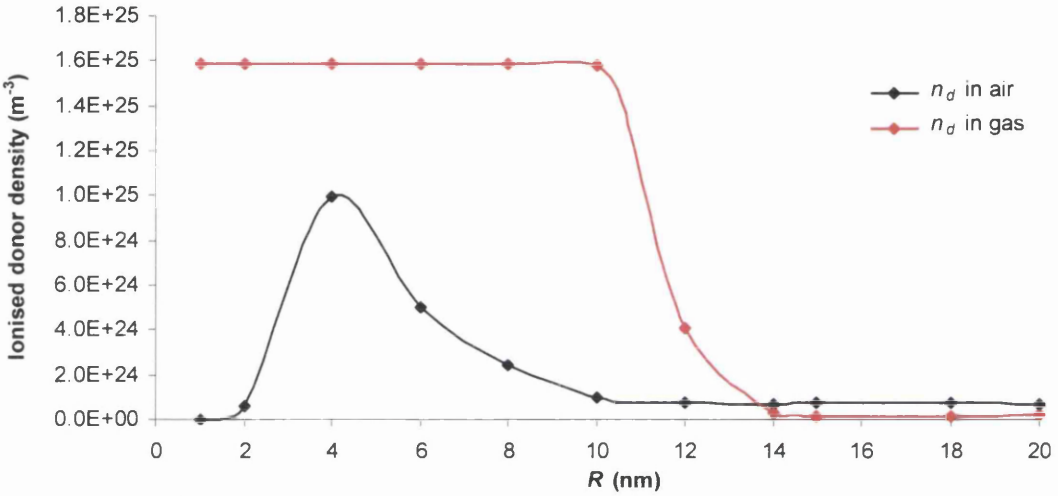


Figure 6-17 Plot of the ionised donor density n_d against SnO_2 grain radius at 673K, $S_b = 1.35$eV. n_d in air is determined from the usual P-KS-CB method. n_d in gas is estimated from a compensated version of Eq. (6.6) incorporated into the P-KS-CB iterations, solving for the Fermi level on the addition of the surface state electrons of the clean air scenario assuming no new surface states.

The peak at R = 10nm of Figure 6-11 has now moved to R = 4nm and is of a slightly lower magnitude but the general trend remains the same despite the temperature difference. If the reducing target gas returns all the surface state electrons to the conduction band, then the Fermi level will change and so to must the ionised donor density as previously discussed. This n_d shift is difficult to assess and is a problem encountered again in Section 6.2.1, but an estimate can be made from Eq. (6.6) as a simple approximation (see Sub-Section 6.2.1 for the procedure). With no surface states, this new n_d trend is also plotted in Figure 6-17, labelled 'n_d in gas'. It can be seen that a transition occurs as R becomes greater than 10nm, the donor vacancies no longer totally ionised, falling to 14nm, where the ionised donor density lies beneath its full surface state counterpart. This movement corresponds to the action of the Fermi level. Below $R \approx$ 14nm, on the addition of the extra electrons to the conduction band the Fermi level moves up into the quantised energy levels to match this increased carrier concentration, the ionised donor

concentration following. Above ~14nm, the charge balance equation is equalised by the Fermi level moving downwards, as in the constant n_d case, and the ionised donor density follows. Figure 6-18 plots the resulting n_{gas}/n_{air} ratio, which very interestingly, firstly rises to R = 10nm and then again sharply below a radius of 4nm.

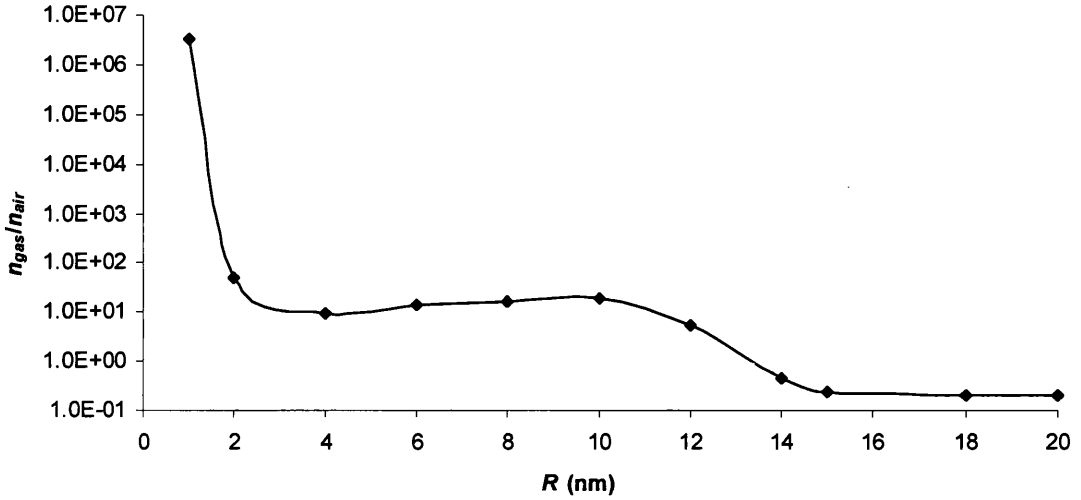


Figure 6-18 Plot of the electron density ratio n_{gas}/n_{air} against SnO_2 grain radius for the variable ionised donor density approach at 673K, S_b = 1.35eV.

This equates to Figure 6-19, where the maximum sensitivity is plotted against grain radius. As would be hoped to explain the experimental sensitivity increases of the sensing film on the decrease of grain size, particularly below the ~20nm diameter (R = 10nm) [41,42], S_{max} increases to R = 10nm and significantly, again below R = 4nm.

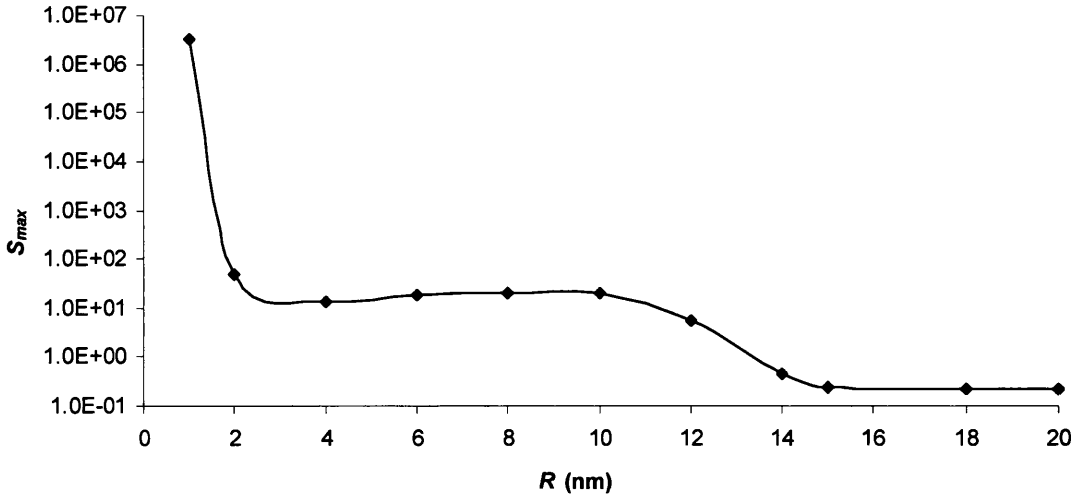


Figure 6-19 Plot of the theoretical maximum sensitivity S_{max} against SnO_2 grain radius for the variable ionised donor density approach at 673K, S_b = 1.35eV.

Crucially, this second sharp rise has been experimentally observed by Kennedy *et al.* [42], and see also Ref.'s [45,47], although attributed to R being less than the space-charged (depletion) width, estimated at 3nm in all three papers by these authors. However, in Ref. [42], λ_D alone was measured at ~ 7 to 8nm, and so it is certainly possible that another explanation is more likely for the observed further rise in sensitivity.

The very slight decrease in S_{max} between 10nm and 4nm in Figure 6-19 is possibly related to the preliminary nature of this data (the use of large meshes and relaxed convergence criterion).

To make one last observation on this little digression; the smaller the grain, the larger the sinter neck in relation to its size, all other things being equal (see Section 2.1). Dependant on the absolute size of the neck of course, but the larger the neck the more likely one of the other conduction mechanisms described in Section 2.1 is to dominate the film conductivity, rather than Schottky barrier transport. In this variable n_d simulation, unlike the other 'constant' case, as the surface barrier height tends to zero and the Fermi level significantly rises up the well above the conduction band bottom, the actual situation described appears to be that of open neck conductance, see Figure 2-4(a), the electron population of the grains overlapping. The determining factor now in the inter-granular conductance is the effective area of the channel between the grains (and of course the density of charge carriers), and with no depletion width, this is essentially the area of the sinter neck. While experimentally this could be measured, from Eq. (2.7), and assuming uniform sintering parameters over the range of radii, a comparison of relative sensitivities can still be estimated, although the exact magnitude is rather arbitrary.

It is not possible at this early stage in this model's development to compare the open neck and Schottky conductance's across the 14nm divide directly, not enough of the other factors (e.g. electron mobility) controlling the magnitude of the conductance known. However, for the sake of illustration only, in Figure 6-20 the dashed lines indicate a switch in conduction mechanism for the $R \geq$ 14nm grains also. Note that the slight decrease in S_{max} between R = 10nm and R = 4nm of Figure 6-19 appears to have been replaced by a slight increase as would be desired.

Essentially, this change of mechanism, if it occurs, would not seem to negate the effects tentatively put forward as a consequence of the variability of the ionised donor density. It would seem from these preliminary results that not only does the 'constant' ionised donor density model fail to successfully simulate the experimentally observed trend, but that the ionised donor density

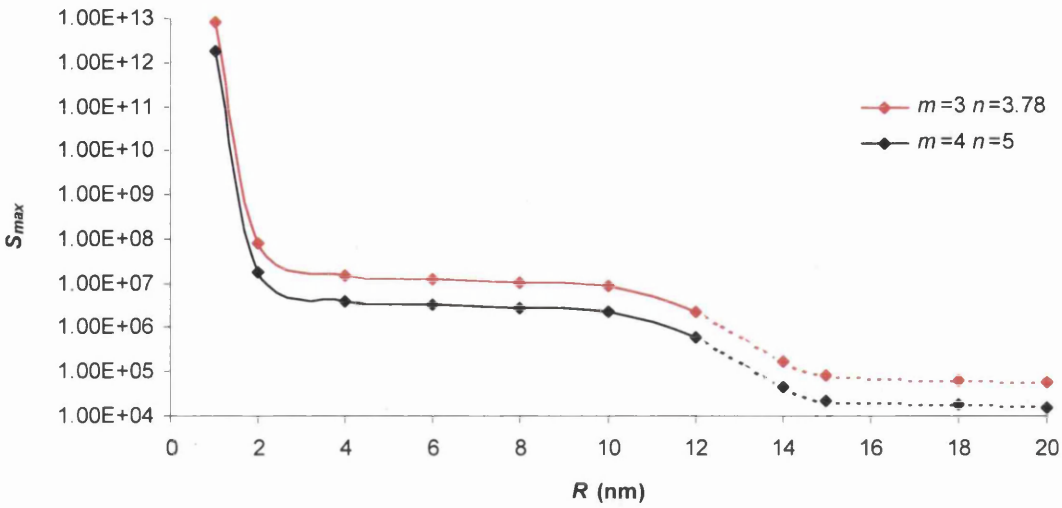


Figure 6-20 Plot of the theoretical maximum sensitivity S_{max} against SnO_2 grain radius for the variable ionised donor density approach at 673K, S_b = 1.35eV, allowing a switch of conduction mechanism from Schottky to open neck for the 'gas' exposed states, to estimate the effects of the Fermi level in the conduction band and the disappearance of the Schottky barrier. Two series are plotted; one where the sinter necks have formed from lattice diffusion (m = 3, n = 3.78) and one from surface diffusion (m = 4, n = 5) the R dependence of each different. As the model stands, the mechanism switch is only likely to occur for grains where R < 14nm; however, at this early stage the relative magnitudes of the two modes (Open/Schottky and Schottky/Schottky) cannot be compared, not enough data being known. Consequentially, all points are treated as if the mechanism switch occurs, the dashed lines indicting the region where this approach is unlikely to be true and likely to be an overestimation of S_{max}.

model does; both with the significant sensitivity increase of R < 10nm grains over those of a larger radius [41,42] and with the sharp sensitivity rise observed below 4nm [42,45,47]. Indeed, this mechanism shift may explain why these increases are seen at the same dimensions in both thick films (e.g. Williams and Coles [41]) and thin films (e.g. Kennedy [42]), where in the latter, only open and closed neck (not Schottky, see McAleer *et al.* [43]) conductance is probable. This topic is not the main focus of this work, but would provide interesting future investigation.

6.1.5 Section Summary

The importance and complexity of this section is such that a formal summary is necessary at this point. The preceding three sub-sections have seen the application of three very different independent models (although the depletion approximation is connected to the P-CB model via some input parameters) to the SnO_2 spherical grain systems in order to span the range of radii over which compatible experimental evidence is available.

With reservations, for >25nm to ~ 70 / 80nm the non-linear Poisson equation has been solved in

conjunction with the charge balance equation (P-CB) for the experimentally measured parameters of the Maffeis *et al.* [31] system. The predicted trend in ionised donor density and corresponding movement of occupied surface acceptor density can be found in Figures 6-1, 6-2, 6-4, 6-13, and 6-14. For grain radii between ~60nm and ~80nm, a relatively constant depletion width of 41.55±0.95nm can be seen to develop. This Λ is within the 1nm to 100nm range reported in Ref. [43] for typical tin dioxide grains.

For radii greater than ~ 43nm, the depletion approximation (DA) can be applied and an analytical solution to the linear Poisson equation achieved. In this way, using the same defining parameters as above and the P-CB calculated value of Λ, the trends in the ionised donor density and occupied surface acceptor density can again be plotted, and the bulk ($R \rightarrow \infty$) values for both ascertained; ${}^{bulk}n_d = (9.76\pm0.08)\times10^{23}\text{m}^{-3}$ and ${}^{bulk}N_t = (4.05\pm0.06)\times10^{16}\text{m}^{-2}$. The calculated bulk surface acceptor density is in qualitative agreement with the bulk $N_t \approx 10^{17}\text{m}^{-2}$ reported in Ref. [43] for typical tin dioxide grains.

From the calculated ${}^{bulk}n_d$ and the experimentally determined position of the donor levels [53,96], the total density of the doubly ionisable oxygen vacancies can be determined as $N_d = (7.91\pm0.07)\times10^{24}\text{m}^{-3}$. As such, an upper limit can be set on the maximum allowed ionised donor density in any of the three schemes, $n_d \leq (1.58\pm0.01)\times10^{25}\text{m}^{-3}$, completely ruling out the behaviour of the P-CB model for radii under 24.5nm.

For small radii grains, $R \leq 20$nm, the effects of quantisation are more and more important until they completely dominate any electronic behaviour. In order to model grains of these dimensions, the non-linear Poisson equation is self-consistently solved with both the Kohn-Sham equations and the charge balance equation (P-KS-CB). In this way, the problem of $n_d \rightarrow \infty$ as $R \rightarrow 0$ is averted (P-CB), and the quantisation effects / surface area to volume ratio play-off results in a peak ionised donor density of $(1.29\pm0.03)\times10^{25}\text{m}^{-3}$ at 10nm. This maximum P-KS-CB n_d does not exceed the maximum allowed n_d calculated independently, and interestingly, the peak in N_t at 10nm does not exceed its independently calculated bulk value either. The relationship between the occupied surface densities of the 4nm, 10nm and bulk semiconductor grains are in agreement (with some reservations) with the experimental work [41] on the sensitivity of gas sensing films using nanometric grains of diameter 8nm and 20nm, as well as coarser grained films.

Most significantly, the ratio between the 4nm and the 15nm radii grains, 5.57±0.37, is in very good agreement with that calculated experimentally, 5.90, from integrated normalized conductivity spectra reported by Maffeis *et al.* [31].

Taken together, the three models are all in quantitative agreement within their applicable ranges as illustrated in Figure 6-13. However, some issues regarding the accuracy of the P-CB data are present, attributable in the main to the likelihood that at these dimensions ($R < 200$nm), where the influence of electron population is perhaps not always negligible, that the energetic spectrum must be treated discretely, rather than as a continuum. Figure 6-14 presents a slightly speculative look at n_d and N_t variation, inferring the movement of n_d, and consequently N_t, from the last point of the P-KS-CB series into the DA [n_d] series i.e. from $n_d = 1.55\times10^{24}\text{m}^{-3}$ to $n_d = 1.09\times10^{24}\text{m}^{-3}$.

The prediction of variable ionised donor density, influential at small and medium radii, adds more complexity to the behaviour of N_t than would otherwise be found for constant n_d. Evidence to validate this hypothesis has been presented both in Chapter 5, and more pertinently, here. However, to prove/disprove this premise conclusively, further experimental data is needed. For the trend in the surface state density of system described in the preceding sub-sections, the peak at ~10nm and the trough at ~18nm, relative to the uniform n_d trend (see Figure 6-9), must be proven. As such, merely extending the work of Maffeis *et al.* to included further grain sizes should be sufficient. To minimise the experimental work, and so utilising the data already collated, adding a point at ~10nm would be adequate for this purpose (the theorised 15nm N_t value very close to that of a 18nm grain). Additional experimentally determined values would, of course, be preferred, and would provide further insight into the problem. For instance, data that would be particularly beneficial would be at the P-KS-CB limit of 20nm and at large R values to test the validity of the Λ calculation (as well as the obvious values of the bulk quantities themselves). Ideally, several points between 20nm and say, 200nm would also be recorded to evaluate the validity of the P-CB model compared to the inferred behaviour. Naturally, the new grains would have to be prepared in the same manner as the existing 4nm and 15nm particles and have equivalent material properties. In particular, the Schottky-like barrier heights should be approximately equal across all samples.

This in mind, Tables 6-1 to 6-4 presents the occupied surface acceptor density ratios for the four most probable concatenations of schemes for the Maffeis *et al.* system, whereby, given relevant experimental measurements, the argument of variable as opposed to constant ionised donor

density could be decisively settled and the validity of all the methodologies in this Chapter tested further. The mean values of N_t for the surface barrier heights of 1.3eV and 1.4eV have been used only for clarity, and the chosen value of the constant ionised donor density, where appropriate, is that of the calculated ${}^{bulk}n_d$.

The N_t ratios of Table 6-1 are based on the constant n_d scheme of Figure 6-9. The non-linear Poisson equation is solved with the electron density of Eq. (1.34) for $R \leq 200$nm, and the depletion approximation's Eq. (6.5) for $R \rightarrow \infty$ (the bulk semiconductor). In practice, any grain with a radius greater than ~ 1 000nm (1μm) can be used to approximate the bulk with a reasonable degree of accuracy.

Table 6-2 is again based on the constant n_d scheme as above, but self-consistent P-KS values replace the surface acceptor densities in the tabulated ratios for $R \leq 15$nm and introduce quantum effects (*Nota bene* P-KS is not the same as P-KS-CB!).

Table 6-3 represents the occupied surface density ratios of the variable n_d schemes of Figure 6-13. As such, the P-KS-CB method calculates the surface state densities for $R \leq 15$nm, the P-CB method covers the radii range $15\text{nm} < R < 43\text{nm}$, and the DA [$n_d$] determines N_t for radii greater than, or equal to, this.

The final table, Table 6-4, represents the surface density ratios of the variable n_d schemes of Figure 6-14, and as such, uses the P-KS-CB model for $R \leq 15$nm, the inferred values of n_d for $15\text{nm} < R < 200\text{nm}$, and DA [$n_d$] data for $R \geq 200$nm.

Table 6-1 Table of occupied surface acceptor density ratios for the constant n_d models, n_d set at ${}^{bulk}n_d$. The non-linear Poisson equation solved with the electron density of Eq. (1.34) for $R \leq 200$nm, and Eq. (6.5) of the depletion approximation is used for the bulk semiconductor value.

	4nm	10nm	15nm	30nm	70nm	200nm	Bulk
4nm	1	2.49	3.75	7.48	17.43	26.28	31.07
10nm		1	1.50	3.00	6.99	10.54	12.46
15nm			1	1.99	4.65	7.01	8.29
30nm				1	2.33	3.51	4.15
70nm					1	1.51	1.78
200nm						1	1.18
Bulk							1

Table 6-2 Table of occupied surface acceptor density ratios for the constant n_d models, n_d set at ${}^{bulk}n_d$. The non-linear Poisson equation solved with the electron density of Eq. (1.34) for 15nm < $R \leq$ 200nm, and self-consistently with the Kohn-Sham equations for $R \leq$ 15nm. The bulk semiconductor surface acceptor density is given by Eq. (6.5) of the depletion approximation.

	4nm	10nm	15nm	30nm	70nm	200nm	Bulk
4nm	1	2.51	3.93	10.96	25.53	38.48	45.50
10nm		1	1.57	4.37	10.17	15.33	18.13
15nm			1	2.79	6.50	9.79	11.58
30nm				1	2.33	3.51	4.15
70nm					1	1.51	1.78
200nm						1	1.18
Bulk							1

Table 6-3 Table of occupied surface acceptor density ratios for the variable n_d models. The non-linear Poisson equation is solved self-consistently with the Kohn-Sham equations and the charge balance equation for $R \leq$ 15nm, and just with the charge balance equation using the electron density of Eq. (1.34) for 15nm < $R <$ 43nm. The surface state densities of grains with radii larger than or equal to 43nm are given by Eq.'s (6.4) and (6.5) of the depletion approximation.

	4nm	10nm	15nm	30nm	70nm	200nm	Bulk
4nm	1	24.91	5.54	40.25	22.14	23.90	25.54
10nm		1	0.22	1.62	0.89	0.96	1.03
15nm			1	7.26	3.99	4.31	4.61
30nm				1	0.55	0.59	0.63
70nm					1	1.08	1.15
200nm						1	1.07
Bulk							1

Table 6-4 Table of occupied surface acceptor density ratios for the variable n_d models of Figure 6-14. The non-linear Poisson equation is solved self-consistently with the Kohn-Sham equations and the charge balance equation for $R \leq$ 15nm, the inferred values of the ionised donor density are used in the range 15nm < $R <$ 200nm, and finally, Eq.'s (6.4) and (6.5) of the depletion approximation are used for $R \geq$ 200nm.

	4nm	10nm	15nm	30nm	70nm	200nm	Bulk
4nm	1	24.91	5.54	7.68	16.70	23.90	25.54
10nm		1	0.22	0.31	0.67	0.96	1.03
15nm			1	1.38	3.01	4.31	4.61
30nm				1	2.18	3.11	3.33
70nm					1	1.43	1.53
200nm						1	1.07
Bulk							1

Based on the data at hand, the variability of the ionised donor density with respect to radius at small dimensions is not in any way disproved, and indeed, within the P-KS-CB model, offers a considerably better simulation of the Maffeis *et al.* system than other comparable hypotheses i.e. those with a constant n_d with respect to radius. Consequentially, in the next section, where charge storage is modelled in a 4nm SnO_2 grain, the P-KS-CB model will be used along with any other required variable-n_d data, without recourse to any alternative, constant n_d, model.

6.2 Charge Writing in 4nm SnO_2 Grains

This section theoretically represents the charge writing observed on 4nm SnO_2 nanocrystals by the same experimental group as measured the occupied surface state density ratio between 4nm and 15nm radii SnO_2 nanocrystals. This group (see Maffeis *et al.* [7] and in particular Wilks *et al.* [6]) has had considerable success in injecting electrons into the quantised energy levels of spherical grains, manufactured through vaporising tin dioxide with a laser in a mixture of Argon and Hydrogen [99] and then deposited on a Si substrate. These particles were then annealed at 400°C for 20 min in air, then at 300°C for 1hr in a vacuum before being charge injected and scanned at room temperature. Transmission Electron Microscopy (TEM) and X-Ray Diffraction (XRD) analysis indicated that the particles were roughly spherical with a median diameter of 8nm. From their results, they infer that for these 4nm radii SnO_2 nanoparticles, a maximum number of 86 electrons can be stored in a single nanocrystal, although they believe that the actual limit is likely to be lower than this.

In order to simulate this charge storage, the self-consistent P-KS-CB technique is again applied, now modelling the Fermi level movement on the addition of each electron (Sub-Section 6.2.1). This data is then used in Sub-Section 6.2.2 in a simple tunnelling model, simulating the tip-substrate current and calculating the grain charging behaviour with tip bias.

6.2.1 Determination of the Fermi Level Shift of a 4nm SnO_2 Grain on Charging

The system is defined from the input parameters and results of Section 6-1 and uses the mean ionized donor density ascertained for the E_f= 0eV neutral grain, $n_d = 1.49\times10^{24}m^{-3}$, and the mean surface barrier height of 1.35eV. The position of the surface barrier height is maintained at a constant level over the addition of charge, enabling the Fermi level position for each integer quantity of charge to be determined using the previously described (see Chapter 5) and previously applied (see preceding sub-section) P-KS-CB method. This Fermi level movement is illustrated in Figure 6-21.

A clear stepped pattern emerges, indicative of electron filling in *s*-, *p*-, *d*- and *f*-orbitals. The angular momentum states are generally selected to obey an atomic-like

$$n = n_{node} + l + 1 \qquad (6.9)$$

requirement, see for instance Gasiorowicz [100], the greater angular momentum states lying at

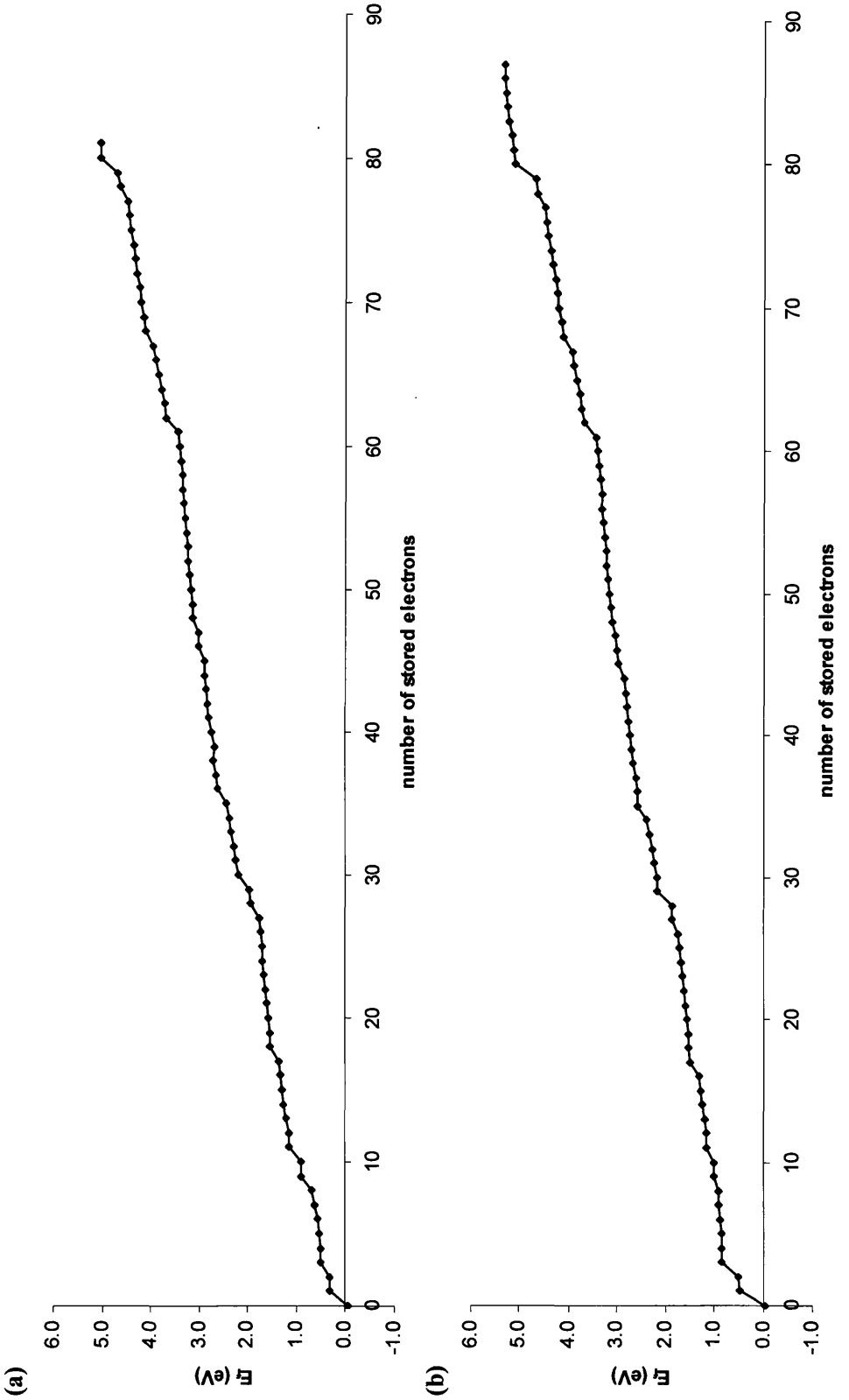


Figure 6-21 Fermi level shift on addition of electrons to a 4nm SnO_2 grain assuming (a) a constant ionised donor density (b) a variable ionised donor density.

higher energy for equal n. n is defined as the principal quantum number, and is the number of nodal surfaces in the wave function plus one. n_{node} represents the number of nodal surfaces within the r dependant part of the wave function and l equals the angular momentum quantum number (as there are l nodal surfaces in a spherical harmonic of degree l – see Section 3.1). In this way, the most physically likely structuring of the levels is achieved. This discrimination forms part of the process of selecting out the suitable eigenstates from the 300+ eigenstates returned by the finite difference solution of the Schrödinger / Kohn-Sham equation (one for each mesh point – see Section 3-2), few of which will even lie within the energetic confines of the defined potential well.

The lowest energy state corresponds to $n = 1$, $l = 0$; the next to $n = 2$, $l = 0$ then $l = 1$ and so forth. For given n there are: n values of l, $l = 0, 1, 2 \ldots n\text{-}1$, for each l; $2l+1$ values of the magnetic quantum number m, $-l \leq m \leq l$; and each wave function has two possible spin states. As a practical example, in hydrogen, the energy levels depend solely on the principal quantum number and so, momentarily neglecting spin for the sake of convention, $n = 1$ is considered non-degenerate (one s-state); $n = 2$ is fourfold degenerate (one s- and three p- states); $n = 3$, nine-fold degenerate (one s-, three p- and five d- states) and so on. Considering spin, of course, doubles the degeneracy. In the quantum dots considered here, each energy level is dependant on both n and l and so each energy state is considered $2(2l+1)$ degenerate, wave functions of equal n and l but differing spin and magnetic quantum numbers all lying at the same energy.

Using the electron configuration notation of $n$orb, where n is again the principal quantum number and 'orb' is the type of orbital (s-, p-, d-, f- etc.), then electron filling of the Wilks *et al.* [6] system illustrated in Figure 6-21(a) occurs in the order:

$$\underline{1s}\ 2s\ 2p\ 3s\ 3p^7\ 3d\ 4s\ 4p\ 4d\ 5s\ 4f\ 5p\ 5d\ 6s\ 5f^{\text{partial}}$$

However, with the thermal electron presence, orbital 'filling' in reality occurs across all orbitals above the Fermi level simultaneously, and so is perhaps a misleading term. Even so, the $n$orb concept does give a good description of the electron structure. Figure 6-22 is a re-print of Figure 6-21(a) with the above labelling convention added.

The $1s$ level has a minimal contribution as thermal effects mean that it is effectively 'skipped' with regard to the charge writing process, being filled from the offing with thermally excited electrons (denoted by the underscore).

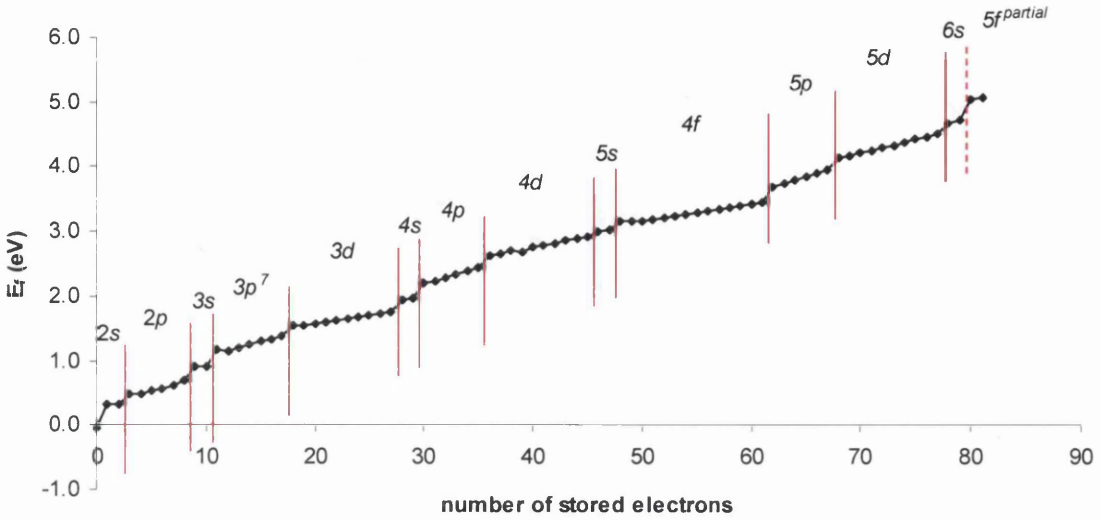


Figure 6-22 Fermi level movement on addition of electrons to a 4nm SnO_2 grain assuming a constant ionised donor density. Re-print of Figure 6-15(a) with added orbital labels.

The 3*p* state appears to contain 7 electrons – this can be explained through the energetic closeness of the early energy levels, and particularly the 3*d* and 3*p* states. Electron ordering and thermal filling is such that added electron number 11 predominantly resides in the 3*s* level, but with a significant high thermal presence in both the 3*p* and 3d levels. The next 6 electrons are then able to be stored within this *p*-orbital, maintaining an increasing thermal presence in the 3*d* state.

The 6*s* $5f^{\text{partial}}$ states alternate position as to which lies lowest energetically. Filling begins with the 5*f* orbital beneath the 6*s* orbital; however, once the 6*s* thermal electron population substantially fills its state, the two orbitals exchange energetic positions, the 6*s* state now energetically the lowest and with a complete electron complement, and the higher 5*f* state partially empty.

This energetic interchange of state positions is analogous to the 4*d*/5*s* fluctuations in atomic electronic structure, seen in the periodic table between Rubidium (^{37}Rb) and Indium (^{49}In). There of course, there is no thermal electron population; it is just a matter of 'energetic' physics, maintaining the lowest possible total atomic energy for any integer electron configuration.

The ability of the applied P-KS-CB method to reach a convergent solution ceases after 81 added electrons, 2 stored electrons 'into' the 5*f* wave functions. This is only 5 electrons short of the total electron complement for the nanocrystal inferred from the experimental work [6].

Throughout Figure's 6-15(a) and 6-16, n_d was maintained at a constant level. This is contrary to what would be implied by Eq. (6.6) if, of course, this equation were valid at these radii. From this relation it would be expected that as the Fermi level rises so too would the number of ionised donors. Although, as has been shown for the case of the neutral grain (see Section 6-1), Eq. (6.6) is not adequate to calculate the necessary ionised donor concentration to maintain the charge neutrality of the grain at these dimensions. However, as the Fermi level rises it certainly would be expected that so too would n_d, up to a maximum of $2N_d$, regardless of the size of grain.

With this is mind, should the denominators of the bulk semiconductor equation Eq. (6.6) be compensated in such a way that it correctly produces the neutral 4nm n_d of the discrete energy spectrum model, for example incorporating a multiplicative factor α such that

$$n_d = \frac{N_d}{1 + g_1 e^{\alpha \frac{(E_d^1 - E_f)}{k_b T}}} + \frac{N_d}{1 + g_2 e^{\alpha \frac{(E_d^2 - E_f)}{k_b T}}} \quad . \tag{6.10}$$

α equal to ~ 0.7336 for the 'standard' system simulated here with $n_d = {}^{neutral}n_d = 1.49\times10^{24}\text{m}^{-3}$, then incorporating this modified equation into the iterative cycles of the P-KS-CB process should give an estimation of n_d's increase with E_f.

This new graph, Figure 6-21(b), differs only slightly from the constant n_d Figure 6-21(a). The variable ionisable donor density reaches a point by ~3 added electrons that the shift in the P-KS-CB satisfying Fermi level causes 100% of the donors to be ionised. Initially this raises the E_f's of Figure 6-21(b) in comparison with the constant case, and makes convergence slightly more difficult, since at this early stage the electronic structure is very much dominated by thermal electrons, and small changes in the Fermi level can result in considerable changes in the electron density. Over the long haul, the increase in n_d causes an increase in N_t, see Figure 6-23, and it happens that the charge balance can be maintained with a slightly lower thermal electron population i.e. the Fermi level of the variable ionised donor density tends to lie below the corresponding position for constant ionised donor density.

The apparent electron structure can be described by

$$\underline{1s}\ 2s\ 2p\ 3s\ 3p\ 3d\ 4s\ 4p\ 4d\ 5s^{2/3}\ 4f^{14/15}\ 5p\ 5d\ 6s\ 5f^{\text{partial}}$$

the $3p$ orbital now containing its correct complement of 6 electrons, and so Figure 6-21(b) pre-empts Figure 6-21(a) for the $3d$ $4s$ $4p$ $4d$ $5s$ plateaus by one added electron. As for the constant case, Figure 6-24 is a re-print of Figure 6-21(b) with the labelling convention added.

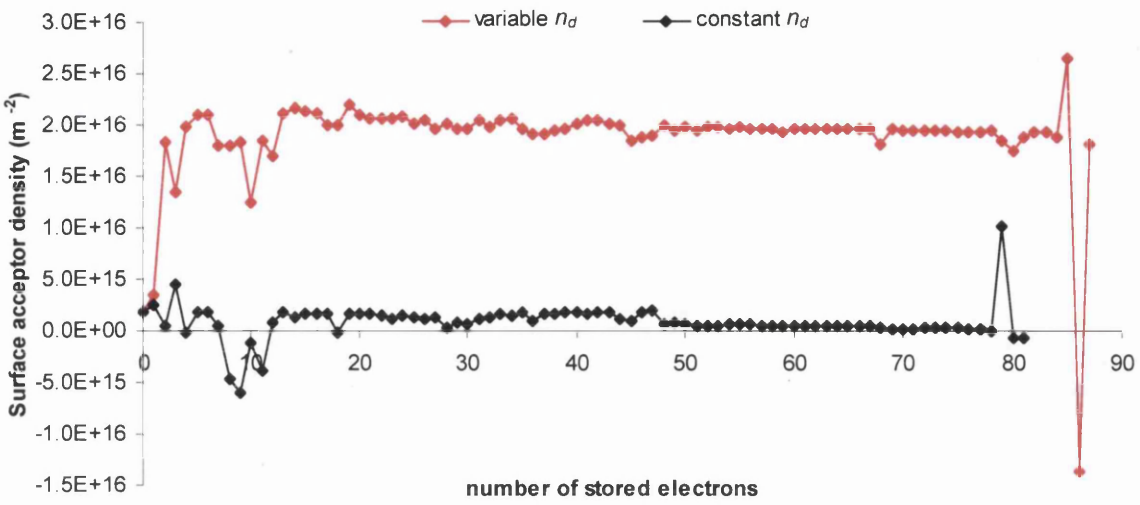


Figure 6-23 Plot of surface acceptor density movement on the addition of electrons to a 4nm SnO_2 grain for both variable and constant ionised donor density (n_d).

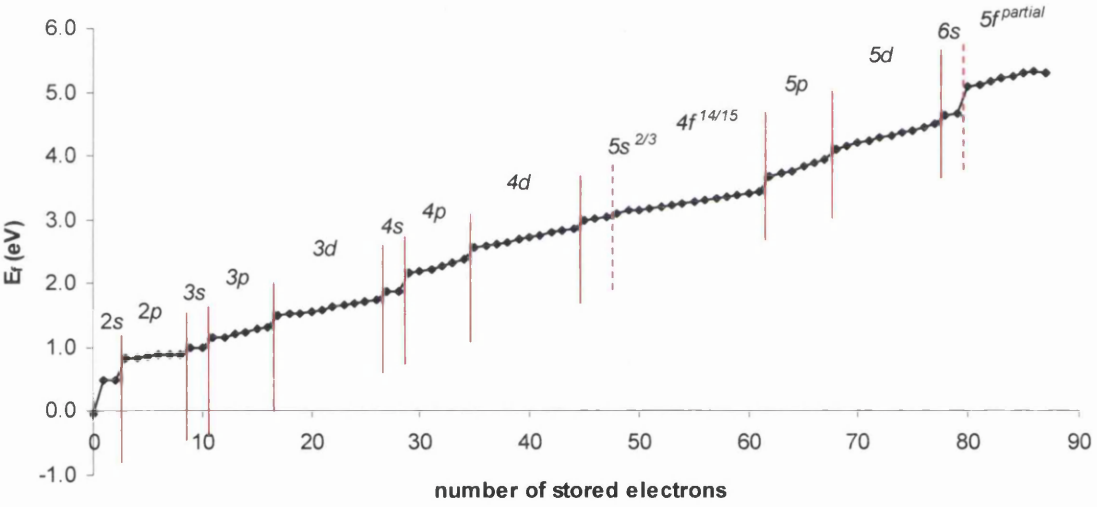


Figure 6-24 Fermi level movement on addition of electrons to a 4nm SnO_2 grain assuming a variable ionised donor density. Re-print of Figure 6-15(b) with added orbital labels.

The 5*s* 4*f* states lie energetically very close together, and during filling alternate position several times as to which lies lowest energetically, analogous to the 6*s* 5*f* $^{\text{partial}}$ alternations of Figure 6-21(a). However, unlike the constant n_d case, due to the closeness of the two states here, a consequence of the subsequent thermal re-ordering during the position switches is a gain of 1 electron in the hybrid 5*s* 4*f* state – 17 electrons being stored instead of 16. This brings Figure 6-21(a) and Figure 6-21(b) back into step for the last few orbital states, where the 6*s* 5*f* $^{\text{partial}}$ orbitals again alternate in position during electron filling.

The convergence of the P-KS-CB methodology now halts at an electron complement of 87 – in good agreement with the 86 electrons inferred from experimental observation [6].

The movement of N_t, as illustrated in Figure 6-23, is slightly chaotic, but is largely dependant on n_d, and as such, can be treated as fairly constant if n_d is. For the variable ionisable donor densities, 100% ionisation is reached by ~3*e*, and so its influence on the variation in N_t can be considered to cease. From Figure 6-23, it can be seen that the greatest instability is found at the start, for less than ~12 electrons, and at the other extreme, greater than ~78 electrons. The charge contained within the surface states is quite ill conditioned, both with respect to variations in v_{sp} at the grain boundary and with respect to variations in the electron density as a whole. The later statement is evident since N_t functions to balance the charge of the total electron population – a complex quantity dependant on energy level position, Fermi level position and the mean of v_s, (the optimised potential, inclusive of v_{sp} and electron-electron interactions). While the thermal electron population acts to smooth out fluctuations in energy level position when considering the movement of the Fermi level, it can be seen to almost have the opposite effect with N_t. At and below 12 additional electrons the high percentage of the total electron contingent made up by thermal electrons is likely to be the dominant reason for N_t's instability. After 78 electrons, nearing the limit of the convergence of the P-KS-CB model, the Coulomb and exchange-correlation effects are very large in magnitude, and have an extremely marked influence on the form of the optimised potential. These are very sensitive quantities to any change in electron density in themselves, and are likely to be the main adverse influence on the surface acceptor density in this range. These two regions of instability are likely to indicate the ranges of greatest inaccuracy in the graphs of Fermi level movement (Figure's 6-15, 6-16 and 6-18).

Interestingly, the 5*s* 4*f* fluctuations, problematic with regard to Fermi level movement, have little effect here – the two states merely being seen as a whole, and as such, treated as a relatively constantly positioned electron sink.

Closer examination of Figure 6-23 reveals a slight downward trend in both profiles, indicative of a small decrease in the thermal electron population. Tentatively, the two surface densities can be approximately modelled via

$$
\begin{aligned}
{}^{Est}N_t^{const\ n_d} &= 1.87\times10^{15} - 2.18\times10^{13}x \\
{}^{Est}N_t^{var\ n_d} &= \begin{cases} 1.87\times10^{15} + 4.60\times10^{15}x & x \le 4 \\ 2.11\times10^{16} - 2.65\times10^{13}x & x > 4 \end{cases}
\end{aligned}
\qquad (6.11)
$$

x representing the number of electrons added to the grain, N_i densities in units of m^{-2}. For clarity, Figure 6-23 is re-printed below as Figure 6-25 with the added trend lines.

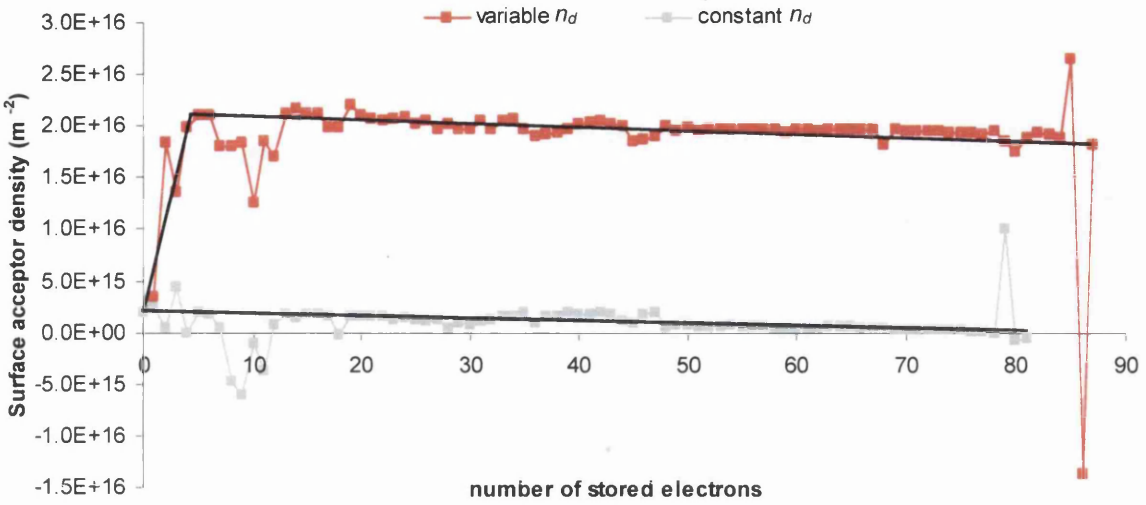


Figure 6-25 Plot of surface acceptor density movement on the addition of electrons to an 4nm SnO_2 grain for both variable and constant ionised donor density (n_d) . Re-print of Figure 6-23 with the added trend lines of Eq. (6.11).

Counter-intuitive as this diminution appears, especially considering the increasing angular momenta available in the higher energy states, it is likely to be caused by the gradual increase in spacing between energy levels as their energy increases up the quantum well. For a simplistic illustration of this see Figure 6-26, where the analytically determined $l = 0$ energy levels of a 5nm spherical, 2.0eV deep square well are plotted (see Section 3-3).

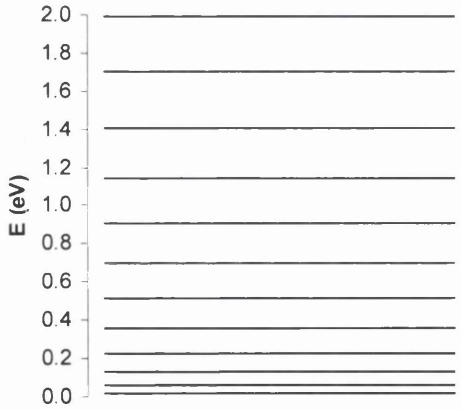


Figure 6-26 Plot of eigenvalues of the $l = 0$ solutions of a 5nm radius spherical, 2.0eV deep square well with the material parameters of a SnO_2 grain. See Chapter 3 for further details.

As the Fermi level moves up the quantum well and thermal electron filling increases in the higher energetic states, the total thermal electron population will decrease as the density of states at these energy levels decreases.

6.2.2 Estimation of Tunnelling Current and Charging Behaviour

To fully evaluate the accuracy of this P-KS-CB model and its derived parameters against Wilks *et al.* [6]'s results, then ideally a simulation of the STM tip-substrate tunnelling current based on the P-KS-CB data should be developed to precisely compare theory with experiment. Unfortunately, the development of a sophisticated tunnelling model is a non-trivial task, and due to time restraints, must be left until a future date. This said however, it is possible to generate a simplified tunnelling program that can provide a first estimation of the tip-substrate current and the dependence of the electron storage on the tip-substrate potential difference.

A detailed discussion of classically forbidden barrier penetration and full tunnelling is not appropriate or necessary for this short, almost introductory, sub-section. For an elementary, cursory overview of the topic recall the form of the wave function of an electron confined within a finite square potential well (see Section 3.3); for electron eigenstates with an energy, E, less than that of the potential V of the barrier wall, the wave function penetrates into the classical forbidden barrier a short distance, as its magnitude exponentially decays to zero. If another identical potential well is situated close enough to the first, then a small proportion of the electron wave will appear in this second well. Since the majority of the probability wave (square of the wave function magnitude) will still exist in the first well then it is most probable that the electron will be found there. However, as both wells are identical then the electron need not favour one over the other and may 'disappear' from the first well and 'reappear' in the second, having 'tunnelled' through the intervening forbidden zone[i]. Indeed, the magnitude of the probability wave existing in the second 'empty' well is greater than would first be expected since in Section 3.3 the second, exponentially increasing, plausible solution of the wave function within the barrier was set to zero to satisfy the boundary condition that $\psi \rightarrow 0$ as $R \rightarrow \infty$. This argument is naturally no longer applicable if this barrier merely connects two wells and the usual form of the wave function in the absence of the second well now gains this exponentially increasing component. For a more rigorous mathematical treatment, the reader is referred to any good quantum mechanics textbook (for instance Merzbacher [37] or Gasiorowicz [100]).

[i] An interesting alternative view can be found in Turton [101]. Through Heisenberg's Uncertainty Principle an electron can be thought of as 'borrowing' enough energy to move over the potential barrier and into the 'empty' well the other side, repaying the energy 'loan'. The 'loan' only lasts for $\sim 10^{-15}$ s but this is long enough for the 'tunnelling' event to occur.

The principles of this simple tunnelling model are outlined in Sub-Section 7.2.3 with a mind to further refinement for future work. Nevertheless, in brief, the movement of electrons through the tunnelling junctions formed by the STM tip – nanocrystal separation and the oxide layer between the nanocrystal and the substrate are taken to be moderated by the Helmholtz free energy changes involved in each electron transfer event. This free energy is defined as the difference between the total energy stored within the system, E_T, and any work done by the system, W. With the assumption that the overall system can be considered to reside in its lowest possible energetic configuration, charge transfer events in which the free energy is decreased are naturally considered most probable.

Shown in Figure 6-27 is the simulated charging behaviour of the familiar 4nm radius SnO_2 nanocrystal for the P-KS-CB variable n_d data determined in the preceding sub-section. The number of electrons deposited in the grain by the STM is plotted against the tip-substrate potential difference. The STM tip is Tungsten and its active region is modelled as a 0.9nm radius sphere (following the approach of Tersoff and Hamman [33]). The tip-grain separation is in practise variable, but taken here as being on average 0.7nm. The grain resides on a 1nm thick SiO_2 oxide layer over the underlying Silicone substrate.

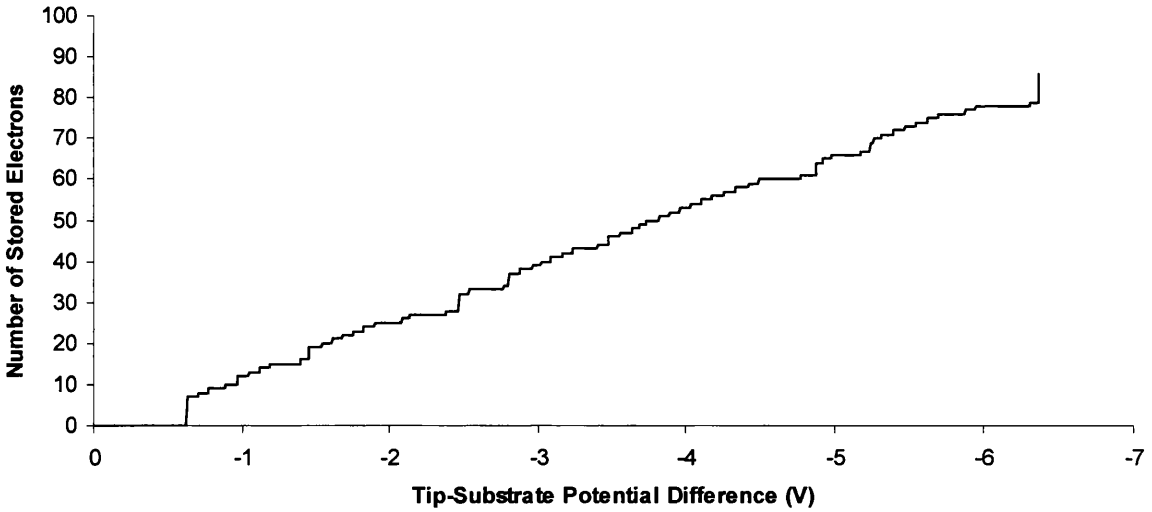


Figure 6-27 Plot of the electron addition to a R = 4nm grain against tip-substrate potential difference for a simple free energy based simulation, using the variable n_d data generated from the P-KS-CB scheme. The tip (W) to grain separation is taken to be 0.7nm, and the oxide layer on the Si substrate is taken to be 1nm thick.

The tip voltage was varied from 0 to -6.365 V, and charge storage seen to occur in 63 separate events, 9 of them multiple electron transfers (corresponding to systems where the n and n+1 Fermi levels lie particularly close together or the two cases where, erroneously, the n+1 Fermi

level has been determined to lie fractionally lower than the n case). The unevenness of the staircase-like pattern, known as the Coulomb staircase[ii], follows from the electron orbital configuration of the Fermi level movement. Deep steps (with respect to the horizontal axis) corresponding to transitions between 'orbitals' (s- to p-, p- to d- etc.), whilst the commoner shallower steps corresponding to the addition of single electrons within specific 'orbitals', the interval between charging events a reflection of the energy difference between the n and $n + 1$ states.

The corresponding tip-substrate current of Figure 6-21 is shown in Figure 6-28.

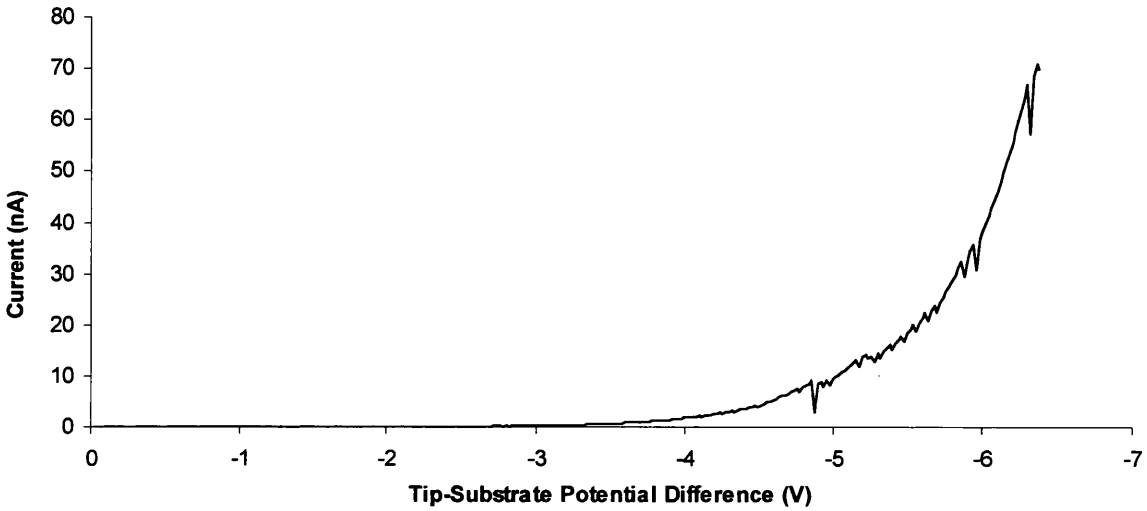


Figure 6-28 Plot of the tip-substrate current against the tip-substrate potential difference for the basic tunnelling simulation of room temperature charge storage on an R = 4nm SnO_2 nanocrystal, using the variable n_d data generated from the P-KS-CB scheme. The tip (W) to grain separation is taken to be 0.7nm, and the oxide layer on the Si substrate is taken to be 1 nm thick.

[ii] The Coulomb staircase [101] is the manifestation of the Coulomb blockade effect, essentially the suppression of charge flow at low bias voltage. Consider two initially uncharged electrodes in close physical proximity connected to a constant current source. Should an electron transfer from one to the other, then the charge difference across the junction will equal $2e$ (the charge on the electrode which has lost the electron of $+e$ and of $-e$ on the electrode which gained the electron), meaning that the tunnelling event has increased the energy of the system from its unchanged state. This is obviously prohibited. This constraint on tunnelling is referred to as the Coulomb blockade. However, although no electron can pass the gap, current may still be considered to flow in the system, causing a build up of negative charge on one electrode and a corresponding amount of positive charge on the other. When this charge reaches a magnitude of $e/2$ on either side, should an electron now tunnel between the two probes, the charge difference across the junction will now equal e, crucially the same as before the charge transfer event. Energy conservation is not violated and so the tunnelling of the electron is allowed. Of course, should two electrons try to cross at once then an energy imbalance arises, as before, and the event is forbidden. In this way, only one electron at a time is allowed passage – hence the term 'single electron transfer'.

The fluctuations in the general exponential-like smooth behaviour corresponding to the grain charging events; the magnitude of the dips increasing as the voltage is increased, with particularly deep minima appearing to be a consequence of multiple electron depositions and transitions between the 'orbitals' of Figure 6-18. The simple scheme used neglects the effects of co-tunnelling (tunnelling events where the overall change in Helmholtz free energy is negative, although the free energy difference over either of the junctions can be positive) and thermionic emission from the STM tip.

Comparison with Wilks *et al.* [6]'s experimental work is again favourable. The authors of this article varied the STM tip voltage from 0 to -6 V, deducing from the differential of the *I-V* graph the occurrence of the plateau regions indicative of the single electron transfer (Coulomb blockade) effect. Figures 6-29 and 6-30 compare the current and *dI/dV* trends in the voltage range -2 to -2.8 V using the adapted experimental data of Ref. [6]'s Figures 5(a) and (b).

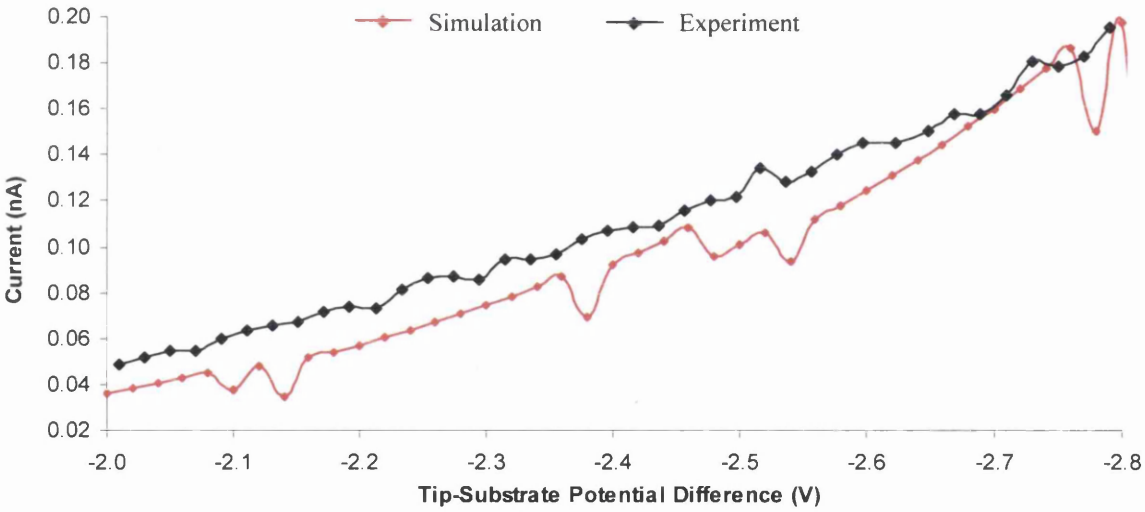


Figure 6-29 Comparison of experimental and theoretically simulated tunnelling current in the range -2.0 to -2.8V during room temperature R = 4nm SnO_2 nanocrystal charging.

The simulated *I-V* curve lies slightly below that measured; this may reflect the previously mentioned neglect of co-tunnelling and thermionic emission or possibly a contribution from charge conducted purely through the surface states of the grain. The differences between the two series of Figure 6-30 are perhaps larger; the simulated maxima and minima (each trough-peak pair corresponding to a single transfer event) are of greater magnitude and less numerous in this region than their measured counter parts. It can but be hoped that better modelling of the tip and assessment of the energy stored within each grain will rectify this.

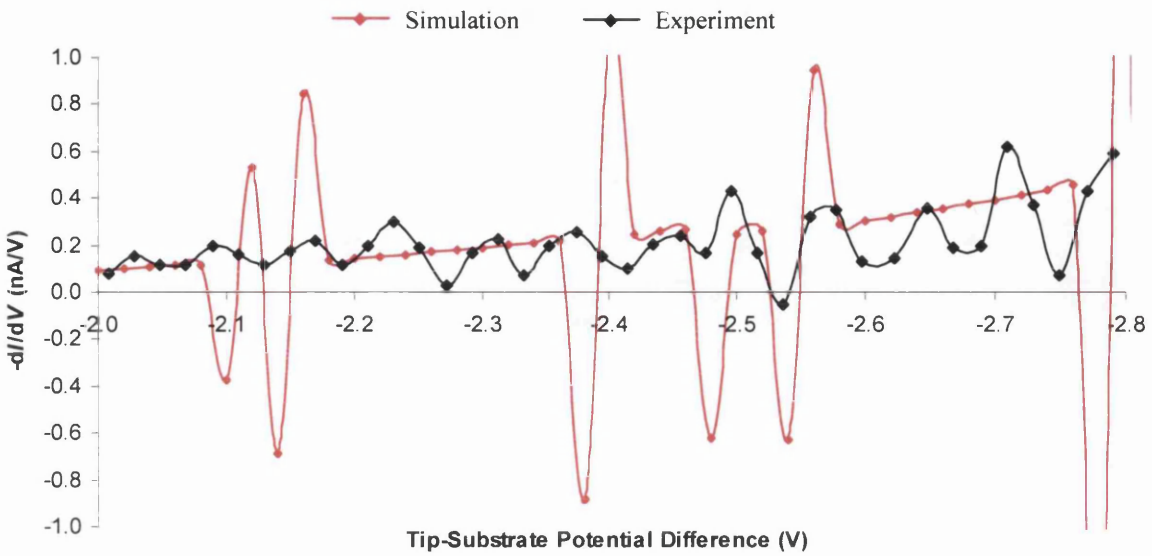


Figure 6-30 Comparison of experimental and theoretically simulated tunneiling current in the range -2.0 to -2.8V during room temperature R = 4nm SnO_2 nanocrystal charging.

Encouragingly, the theoretical simulation places the mode of the voltage intervals between charging events at 0.065V (mean of 0.098V) in excellent agreement with the average period of the experimentally observed dI/dV fluctuations of 0.07V (hence the inferred upper limit of 86 electrons for each grain: 6 divided by 0.07).

6.3 Conclusions

The theoretical methods developed in preceding Chapters were applied to the SnO_2 grains at room temperature as used in experiment by the UWS MNC charge writing group, reported in Maffeis *et al.* [7, 31] and Wilks *et al.* [6], with pleasing results.

Firstly for neutral nanocrystals, using the measured values of the surface barrier heights for 4nm and 15nm radii grains, taken as 1.3eV and 1.4eV respectively, the bulk ionised donor density was extrapolated as ${}^{bulk}n_d = (9.76\pm0.08)\times10^{23}\text{m}^{-3}$ with a depletion width of 41.55±0.95nm. This depletion width and its corresponding surface acceptor density, $N_t = (4.05\pm0.06)\times10^{16}\text{m}^{-2}$, are both in qualitative agreement with that experimentally found by McAleer *et al.* [43]: $N_t \approx 10^{17}\text{m}^{-2}$ Λ = 1-100nm for $n_d \approx 10^{24}$-10^{26}m^{-3}. From the ionised donor density of the bulk semiconductor, the concentration of donors could be set at $(7.91\pm0.07)\times10^{24}\text{m}^{-3}$, and consequentially the upper limit on n_d could be placed at $(1.58\pm0.01)\times10^{25}\text{m}^{-3}$.

Applying the P-KS-CB method in the range 2nm to 20nm, the surface area to volume ratio/quantisation play-off results in a peak ionised donor concentration of $\sim(1.29\pm0.03)\times10^{25}\text{m}^{-3}$ (within its upper limit imposed by consideration of its asymptotic behaviour), and a surface acceptor density of $(3.95\pm0.08)\times10^{16}\text{m}^{-2}$ at 10nm. Moreover, the predicted N_t ratio between the 4nm and 15nm nanocrystals is 5.57±0.37, in excellent agreement with the recorded 5.90 [31].

The relationship between the occupied surface densities of the 4nm, 10nm and bulk radii semiconductor grains are also in good qualitative agreement (with some reservations) with the experimental work of Williams and Coles [41] on the sensitivity of gas sensing films. Preliminary simulations of the inter-granular conduction mechanisms indicate that the variation of the ionised donor density appears to explain the observed sensitivity increase below R = 10nm [41,42] and below R = 3nm [42,45,47] of SnO_2 gas sensing films in general.

The second application of the theoretical methods was to model the charged 4nm SnO_2 grains of Maffeis *et al.* [7] and in particular Wilks *et al.* [6]. Using a mean surface barrier height of 1.35eV and consequent mean n_d of $1.49\times10^{24}\text{m}^{-3}$ as determined for the neutral grains of Ref. [31] analysed above, then on increasing grain charge – simulating the addition of individual electrons to the

grain via the tip of an STM - a clear, stepped movement of the Fermi level was seen. The E_f pattern, barring an occasional alteration due to thermal effects, corresponding to electron storage in the arrangement

$$1s\ 2s\ 2p\ 3s\ 3p\ 3d\ 4s\ 4p\ 4d\ 5s\ 4f\ 5p\ 5d\ 6s\ 5f.$$

The increase in Fermi level is likely to cause an increase in ionised donor density, but approximately allowing for this via a modified Eq. (6.6) causes very little real difference in the E_f profiles, the occupancy of the surface density of states adjusting to compensate for the increase in charge. After ~3 added electrons, n_d reaches $1.58\times10^{25}\text{m}^{-3}$, and the mean surface acceptor density moves from $\overline{N}_t^{const}=8.76\times10^{14}\text{m}^{-2}$ to $\overline{N}_t^{var}=1.91\times10^{16}\text{m}^{-2}$.

The maximum sustainable electron population within one nanocrystal is placed between 81 and 87 electrons in very good agreement with the experimentally inferred limit of 86 [6].

A basic estimate of the charging of, and the tunnelling current through, the 4nm nanocrystals of Ref. [6] using the P-KS-CB data is again in favourable agreement with that measured, both in magnitude and form, and with the modal potential difference interval between charge transfer events calculated to be 0.065 V, compared with the actual interval of 0.07 V.

Overall, it can be concluded that whilst there are many approximations inherent in the theoretical simulations outline here (use of spherical symmetry, use of the KS energy levels in place of the many particle energy levels in order to estimate the thermal electron population etc.), in matching the experimental data as well as they do, these models can be taken to offer a comprehensive and accurate simulation of the surface state properties and charging behaviour of SnO_2 nanoparticles. In doing so, at the very least, they can be taken as strong support of the hypothesis developed in Chapter 5 that both the boundary values used here to solve the non-linear Poisson equation, and not just S_b, are dependent on the material and operating parameters.

Chapter 7 Conclusions and Further Work

This Chapter concludes the thesis. It presents a summary of the theories and results of this work and offers several directions that a continuation of these investigations could take.

Section 7.1 discusses the proposals made within this treatise and analyses the models created to simulate the band structure of the SnO_2 nanocrystals. It treats the data from both the depletion approximation, where the effects of the thermally disassociated electrons are discounted from an assessment of the crystals charge density, and the more complex complete charge density model. It re-introduces the charge balance equation and returns to the pivotal proposal of this work, that the electron density satisfying S_b, $d\varphi / dr\big|_{r=0} = 0$, R, n_d, E_f, T, Q, m^* and ε consistently is unique, and represents the real 'physical' system for those values of parameter. This section summarises the results of the self-consistent solution of the coupled non-linear Poisson equation, Kohn-Sham equations and the charge balance equation, and discusses how these can be used to model the discrete electronic structure of the nanocrystals under charging. The section concludes with an evaluation of the model simulating the charge writing process itself.

Section 7.2 considers the improvements that could be made to the developed methodologies, and examines avenues of future inquiry that might prove fruitful. It discusses orbital dependent correlation, charge leakage between grains and the treatment of the individual electrons as quasiparticles (electron plus polarization cloud) to improve the estimation of the true multi-particle eigenstates. Most importantly, it develops the STM Tip-QD-Substrate tunnelling model briefly used in Chapter 6, and discusses the further refinements that could be made to this essential component of the charge writing simulation.

Section 7.3 concludes the Chapter and indeed the thesis, and presents a few final remarks in closure.

7.1 Synopsis

This work has seen the development of a suite of theoretical models and techniques to obtain a clearer understanding of the fundamental physics behind the charge writing process on spherical SnO_2 nanocrystals.

The basis of any simulation of this nature must begin at the determination of the conduction band structure. For the Maffeis *et al.* [31] system, the depletion approximation can be applied to grains with a radius > 43nm, and as such, it is assumed that the grain is large enough to possess two distinct regions. For the first zone, lying at the centre of the grain and extending from the origin to a radius R_0, the density of the disassociated mobile electrons exactly matches that of the stationary ionised vacancies. Beyond this, from R_0 to R, it is assumed that this region is completely depleted of charge carriers by the surface states and as such is known as the depletion width. In this way, with the boundary conditions that the conduction band bottom must equal the experimentally measured surface barrier height at the boundary of the grain and have a zero first derivative at its centre (or rather R_0 in this approximation), then an unique analytical solution can be formed for Poisson's (linear) equation.

Below 43nm, when the grains become too small to possess a distinct region un-depleted of charge carriers, the curvature of the conduction (and valence) band flatten, and the depth of the potential well formed between the conduction band at the surface and at the centre of the grain, diminishes. In this range of radii, where the semiconductor band bending does not fully evolve and is dominated by the effects of the surface states, the mobile electrons cannot be neglected, and to determine the form of the conduction band bottom, the non-linear Poisson equation must be solved. Due to the non-linear nature of the differential equation, numerical techniques must be employed to achieve this.

This work proposes that in this regime, it is not to be assumed that the same values of the defining parameters R, n_d, E_f, T, Q, m^* and ε corresponding to the first Poisson boundary value, the barrier height S_b, also correspond to a value of zero for the second Poisson equation boundary value, the derivative of the potential at the grain centre.

Furthermore, it is proposed that the potential, and therefore the electron density, satisfying S_b, $d\varphi/dr|_{r=0} = 0$, R, n_d, E_f, T, Q, m^* and ε consistently is unique, and represents the real 'physical' system for those values of parameter.

In this way, by employing the charge balance equation Eq. (5.42), which incorporates the charge within the surface states, and consistently solving it with the non-linear Poisson equation, then it is suggested that the true electron density of the system can be determined.

With reservations, for >25nm to ~ 70 / 80nm the non-linear Poisson equation can be solved in conjunction with the charge balance equation (P-CB) for the experimentally measured parameters of the Maffeis *et al.* system. For grain radii between ~60nm and ~80nm, a relatively constant depletion width of 41.55 ± 0.95 nm is seen to develop.

For smaller radii grains, $R \leq 20$nm (if not at greater radii), the effects of quantisation increase on diminishing size, until they completely dominate any electronic behaviour. In order to model grains of these dimensions, the non-linear Poisson equation is self-consistently solved with both the Kohn-Sham equations and the charge balance equation (P-KS-CB). The Kohn-Sham method maps the interacting many bodied problem onto an auxiliary single-particle problem, its single particle eigenstates minimising the total energy of the system with respect to the electron density and the constraint of constant particle number. For the systems under study, the occupancy of the discrete energy levels is determined by Fermi-Dirac statistics and is dependent on the Fermi level and conduction band bottom; consequential the 'constraint' of constant particle number is actually a function of the electron density. However, it is proposed that the charge balance consistency procedure ensures that its satisfying electron density uniquely (or uniquely for its given method at least) represents the true electron density. In this way, the total particle number is in essence once again a constraint, and the energy minimisation procedure retains its meaning.

The validation of the electron density uniqueness hypotheses and the dependence of the non-linear Poisson equation boundary value $d\varphi/dr|_{r=0}$ on the values of the defining parameters R, n_d, E_f, T, Q, m^* and ε, comes in four parts.

In the first instance, it was shown that the electric field can be uniquely specified by the available data if the potential, and so the charge density, is. As such, $d\varphi/dr|_{r=0}$ will correspond to a given

charge density and so will be dependent on the defining parameters if the charge density is. Therefore, for small grains ($R < \Lambda$) where no neutral zone is present and the whole grain spatially charged, then the charge density, and so the boundary value $d\varphi/dr\big|_{r=0}$, are clearly defining parameter dependent.

The second part of the validating evidence is from the analytical limits, and is more suggestive than perhaps offering irrefutable substantiation. A case can be made that in the limit $R \rightarrow 0$, $d^2\varphi/dr^2\big|_{r=0} \rightarrow 0$; yet in general $\rho(0) \rightarrow \Sigma$ where Σ is some positive number in this same limit unless S_b is equal to one specific value, that of $-\varphi(0)$ for a $R > \Lambda$ grain when $\rho(0) \rightarrow 0$. Thus with $d\varphi/dr\big|_{r=0} = 0$, the inconsistency $\Sigma = 0$ is generally evident unless S_b is equal to one unique value i.e. the parameters consistent with $d\varphi/dr\big|_{r=0} = 0$ are not consistent with general S_b, and are only consistent with one particular value of S_b, suggestive of an unique potential, and as such, an unique electron density.

However, the $\Sigma = 0$ incongruity can also be resolved if, for example, $d\varphi/dr\big|_{r=0}$ and n_d are interdependent and $d\varphi/dr\big|_{r=0} = 0$ as $R \rightarrow 0$ requires that $n_d \rightarrow 0$. Indeed, at the other end of the spectrum, in the depletion approximation, maintaining $d\varphi/dr\big|_{r=0} = 0$ and constant S_b, E_f, T, Q, m^* and ε, then over variation in R, contrary to current opinion [15], either the depletion width must vary or the ionised donor density. The non-linear P-CB method indicates that a constant depletion width and variable ionised donor density are favoured. Additionally, the expression of Samson and Fonstad, Eq. (6.6), determining the density of ionised donors from the donor density, is essential one based on bulk semiconductor statistics; in the nanometric regime, with the effects of quantisation, it would perhaps be unusual if this bulk expression did hold, and n_d did remain constant over the decrease of R, the other defining parameters remaining unchanged. Therefore, to come finally to the point, assuming that n_d does vary, then for constant n_d, E_f, T, Q, m^* and ε, over variation in R, a variation in $d\varphi/dr\big|_{r=0}$ is implied i.e. $d\varphi/dr\big|_{r=0} = 0$ is dependant on the parameters R, n_d, E_f, T, Q, m^* and ε.

The third part of the hypothesis' justification follows from analysis of published work, in particular Malagú *et al.* [15], and indications of discrepancy therein, arising through this issue not being considered and is very encouraging.

The fourth, and final, part of the validation is the success of the models based upon this hypothesis in simulating the experimental data, often far better than existing models not treating this issue. Specifically, from the measured surface barrier heights of 4nm and 15nm radii grains [31], the bulk ionised donor density was extrapolated as ${}^{bulk}n_d = (9.76\pm0.08)\times10^{23}\mathrm{m}^{-3}$ with a depletion width of 41.55±0.95nm and bulk surface acceptor density of $N_t = (4.05\pm0.06)\times10^{16}\mathrm{m}^{-2}$, in qualitative agreement with that experimentally found by McAleer *et al.* [43]. The peak ionised donor concentration was independently determined (from a different model within the suite) to lie at $\sim(1.29\pm0.03)\times10^{25}$ at 10nm, within the upper limit imposed by the doubly ionisable vacancies of density N_d set from ${}^{bulk}n_d$. Most importantly, the N_t ratio between the 4nm and 15nm nanocrystals is found to be 5.57±0.37, in excellent agreement with the recorded 5.90 [31], and improving considerably on other simulations not treating the boundary value consistency issue (see Ref. [31]). This variable ionised donor density behaviour can be used to offer the beginnings of an explanation for the hitherto unsatisfactorily justified, experimentally observed sensitivity increase of polycrystalline gas sensing films on the decrease of the grain radii, particularly below $R = 10$nm [41,42] and $R = 3$nm [42,45,47].

From the ionised donor density satisfying the P-KS-CB model for the given S_b (relative to E_f at zero) and known parameters of a neutral 4nm radius SnO_2 grain, the same model can be used to predict the movement of the Fermi level on the injection of charge to the grains. It estimates the maximum sustainable electron population within one nanocrystal at 81 to 87 electrons; once again in very good agreement with the experimentally inferred limit of 86 [6].

The final program of the suite uses the P-KS-CB charging data, applying Fermi's golden rule and reflecting upon the changes in Helmholtz free energy, to calculate the STM tip-substrate tunnelling current and the voltage dependency of the charge storage events. Both the magnitude and form of the tunnelling current compare favourably with that experimentally measured. The modal potential difference interval between charge transfer events is calculated to be 0.065V, compared with the actual 0.07V.

The evidence in favour of the legitimacy of the boundary value hypothesis and the models that stem from it is considerable.

Overall, this work can be considered to make an original contribution to addressing the experimental SnO_2 data from this new and exciting field of charge writing. It presents the novel

idea of combining the confining potential generated from the non-linear Poisson equation at non-zero temperature, self-consistently with the Kohn-Sham equations and also balancing the charge on the grain to ensure the consistency of the occupied surface states and charge density. Further work, discussed in the next section, suggests some improvements that could be made to the models, and the direction that future investigations could take.

7.2 Further Work

This section details some methods of improvement of the models and possible avenues of future enquiry.

7.2.1 Orbital Dependent Correlation

Due to the rather arbitrary division of the exchange-correlation energy term into its constituent parts E_x and E_c, it is likely to be best to treat both terms, exchange and correlation, in a consistent manner to achieve the best cancellation of errors introduced in the approximation of the correlation term. By modelling the exchange effects as functionals of the single-particle orbitals then the most accurate results are likely to follow if the correlation is treated in a similar manner. However, orbital dependant representations of the correlation energy are not exact and are complex objects indeed. The method of Engel and Dreizler [81] offering the first order approximation of the correlation energy

$$
\begin{aligned}
E_c = -\frac{1}{2}\sum_{i,j}^{occ}\sum_{k,l}^{unoccup} & \frac{1}{\varepsilon_i^{KS}+\varepsilon_j^{KS}-\varepsilon_k^{KS}-\varepsilon_l^{KS}} \int d^3\mathbf{r}_1 d^3\mathbf{r}_2 \frac{\psi_i^{KS*}(\mathbf{r}_1)\psi_k^{KS}(\mathbf{r}_1)\psi_j^{KS*}(\mathbf{r}_2)\psi_l^{KS}(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|} \\
\times \int d^3\mathbf{r}_1 d^3\mathbf{r}_2 & \left(\frac{\psi_k^{KS*}(\mathbf{r}_3)\psi_i^{KS}(\mathbf{r}_3)\psi_l^{KS*}(\mathbf{r}_4)\psi_j^{KS}(\mathbf{r}_4)}{|\mathbf{r}_3-\mathbf{r}_4|} - \frac{\psi_k^{KS*}(\mathbf{r}_3)\psi_j^{KS}(\mathbf{r}_3)\psi_l^{KS*}(\mathbf{r}_4)\psi_i^{KS}(\mathbf{r}_4)}{|\mathbf{r}_3-\mathbf{r}_4|} \right) \\
& + \sum_{k}^{occ}\sum_{l}^{unocc} \frac{1}{\varepsilon_k^{KS}-\varepsilon_l^{KS}} \left| \int d^3\mathbf{r}_1 \psi_k^{KS*}(\mathbf{r}_1)\psi_l^{KS}(\mathbf{r}_1) v_x(\mathbf{r}_1) \right. \\
& \left. + \sum_{j}^{occ} \int d^3\mathbf{r}_1 d^3\mathbf{r}_2 \frac{\psi_k^{KS*}(\mathbf{r}_1)\psi_j^{KS}(\mathbf{r}_1)\psi_j^{KS*}(\mathbf{r}_2)\psi_l^{KS}(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|} \right|
\end{aligned}
\tag{7.1}
$$

might provide a workable expression, although since the correlation effects are of such a significantly smaller magnitude, undue expenditure of computational effort for an improvement of accuracy negligible compared to the errors inherently introduced through the approximations forming the system (and the ~0.2% inaccuracy of the finite difference approach in determining the eigenvalues), would not be cost effective.

7.2.2 Quasiparticles: The GW Approximation

The single particle Kohn-Sham eigenstates are, in the strictest sense, purely Lagrange multipliers (excepting the highest occupied orbital – Koopman's theorem, see Section 4.1). As such, their use as the true eigenstates of the multi-particle system is not conceptually correct, notwithstanding

the experimental and theoretical evidence justifying their use as approximations to ionisation and excitation energies [23,24,25,26,27, 28].

In order to improve on this situation, the next level of sophistication within the model is to add to the independent single electron its Coulomb hole (due to repulsive Coulomb interactions), thus simulating the electron and its surrounding polarisation cloud of positive charge as a single particle-like quasiparticle [102]. Indeed, the Coulomb hole reduces the total charge of the quasiparticle and therefore screens the interactions between quasiparticles. This screened interaction is sufficiently weak that the quasiparticle is effectively almost independent, thus justifying the use of its eigenstates as improved approximations of the true 'excited' many-electron eigenstates.

To simulate the electron-Coulomb hole ensemble, one avenue which has met with considerable success is Hedin's Green's function approach. In keeping with the nature of this chapter, this scheme is given only in outline below, and not rigorously developed. For a detailed derivation of the following equations and their implications, the reader is referred to Hedin's original paper [103] and a good primer on quantum field theory, see for example Kaku [104].

To introduce this method, begin by considering the equation

$$\left(i\frac{\partial}{\partial t} - H\right)\psi = 0 \tag{7.2}$$

where the true Hamiltonian can be separated into two pieces, H_0 the Hamiltonian of the non-interacting system and H_I, the complete interaction effects. To determine the propagator $G(\mathbf{r}, \mathbf{r}', t, t')$ of ψ i.e. the function $G(\mathbf{r}, \mathbf{r}', t, t')$ which propagates the wave function ψ in time from t to $t' > t$ such that

$$\psi(\mathbf{r}',t') = \int d\mathbf{r} G(\mathbf{r},\mathbf{r}';t,t')\psi(\mathbf{r},t) \qquad t > t' \tag{7.3}$$

then

$$\left(i\frac{\partial}{\partial t} - (H_0 + H_I)\right)G(\mathbf{r},\mathbf{r}';t,t') = \delta(\mathbf{r}-\mathbf{r}')\delta(t-t') \tag{7.4}$$

must be solved. The propagator $G(\mathbf{r}, \mathbf{r}', t, t')$ is a Green's function in the correct mathematical sense of the term.

This is not necessarily a simple matter; however, if the Green's function of the non-interacting Hamiltonian, termed G_0, is known, then symbolically[i]

$$\begin{aligned} G &= G_0 + GH_I G_0 \\ &= G_0 + G_0 H_I G_0 + G_0 H_I G_0 H_I G_0 + \cdots \end{aligned} \tag{7.5}$$

which can be written explicitly [104]

$$\begin{aligned} G(\mathbf{r},\mathbf{r}';t,t') &= G_0(\mathbf{r},\mathbf{r}';t,t') + \int dt_1 \int d\mathbf{r}_1 G(\mathbf{r},\mathbf{r}_1;t,t_1) H_I(\mathbf{r}_1,t_1) G_0(\mathbf{r}_1,\mathbf{r}';t_1,t') \\ &= G_0(\mathbf{r},\mathbf{r}';t,t') + \int d^4 r_1 G_0(\mathbf{r},\mathbf{r}_1;t,t_1) H_I(\mathbf{r}_1,t_1) G_0(\mathbf{r}_1,\mathbf{r}';t_1,t') \\ &\quad + \int d^4 r_1 d^4 r_2 G_0(\mathbf{r},\mathbf{r}_1;t,t_1) H_I(\mathbf{r}_1,t_1) G_0(\mathbf{r}_1,\mathbf{r}_2;t_1,t_2) H_I(\mathbf{r}_2,t_2) G_0(\mathbf{r}_2,\mathbf{r}';t_2,t') + \cdots \end{aligned} \tag{7.6}$$

With ψ_0 as the solution of the homogeneous form of Eq. (7.2), $(i\partial/\partial t - H_0)\psi_0 = 0$, then the time evolution of the wave function becomes

$$\begin{aligned} \psi(\mathbf{r},t) &= \psi_0(\mathbf{r},t) + \int d^4 r_1 G_0(\mathbf{r},\mathbf{r}_1;t,t_1) H_I(\mathbf{r}_1,t_1) \psi(\mathbf{r}_1,t_1) \\ &= \psi_0(\mathbf{r},t) + \int d^4 r_1 G_0(\mathbf{r},\mathbf{r}_1;t,t_1) H_I(\mathbf{r}_1,t_1) \psi_0(\mathbf{r}_1,t_1) \\ &\quad + \int d^4 r_1 d^4 r_2 G_0(\mathbf{r},\mathbf{r}_1;t,t_1) H_I(\mathbf{r}_1,t_1) G_0(\mathbf{r}_1,\mathbf{r}_2;t_1,t_2) H_I(\mathbf{r}_2,t_2) \psi_0(\mathbf{r}_2,t_2) + \cdots \\ &\quad + \int d^4 r_1 \cdots d^4 r_n G_0(\mathbf{r},\mathbf{r}_1;t,t_1) H_I(\mathbf{r}_1,t_1) \cdots \times G_0(\mathbf{r}_{n-1},\mathbf{r}_n;t_{n-1},t_n) H_I(\mathbf{r}_n,t_n) \psi_0(\mathbf{r}_n,t_n) + \cdots \end{aligned} \tag{7.7}$$

---

[i] For two operators A and B

$$\begin{aligned} \frac{1}{A+B} &= \frac{1}{1+A^{-1}B}\frac{1}{A} = (1 - A^{-1}B + A^{-1}BA^{-1}B + \cdots)A^{-1} \\ &= A^{-1} - A^{-1}BA^{-1} + A^{-1}BA^{-1}BA^{-1} + \cdots \\ &= A^{-1} - \frac{1}{A+B}BA^{-1} \end{aligned}$$

Identifying

$$\begin{aligned} A &= i\frac{\partial}{\partial t} - H_0 \\ B &= -H_I \\ G &= \frac{1}{A+B} \\ G_0 &= \frac{1}{A} \end{aligned}$$

then the symbolic relations

$$\begin{aligned} G &= G_0 + GH_I G_0 \\ &= G_0 + G_0 H_I G_0 + G_0 H_I G_0 H_I G_0 + \cdots \end{aligned}$$

follow.

from Eq. (AIII.74) [105]. This perturbative expansion can be visualised as the interaction of a particle with a background potential at various points along its path [104] as illustrated in Figure 7-1.

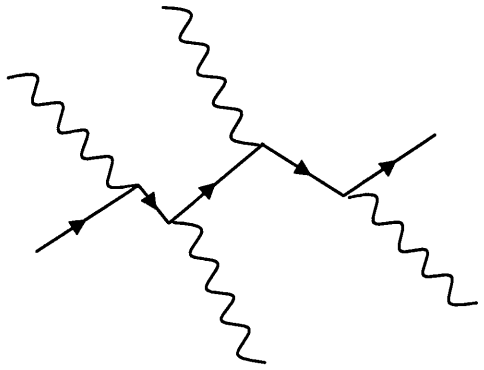


Figure 7-1 Representation of the time evolution of a wave function, the particle interacting with a background potential at various points along its path.

When the Hamiltonian is time independent and energy is conserved, $G(\mathbf{r}, \mathbf{r}'; t, t')$ depends only on the time difference $t - t'$ and as such, can be written $G(\mathbf{r}, \mathbf{r}'; t\text{-}t')$. This can be Fourier transformed into the frequency domain, and so into $G(\mathbf{r}, \mathbf{r}'; \varepsilon)$. Explicitly

$$G(\mathbf{r},\mathbf{r}';\varepsilon) = \int G(\mathbf{r},\mathbf{r}';t-t')e^{i\varepsilon(t-t')}d(t-t') \qquad (7.8)$$

$$G(\mathbf{r},\mathbf{r}';t-t') = \frac{1}{2\pi}\int G(\mathbf{r},\mathbf{r}';\varepsilon)e^{-i\varepsilon(t-t')}d\varepsilon$$

Defining the interaction Hamiltonian, H_I, as the non-local irreducible self-energy operator then

$$\left(-\frac{\nabla^2}{2} + v(\mathbf{r}) - \varepsilon_i\right)G(\mathbf{r},\mathbf{r}';\varepsilon_i) + \int \Sigma(\mathbf{r},\mathbf{r}'';\varepsilon_i)G(\mathbf{r}'',\mathbf{r}';\varepsilon_i)d\mathbf{r}'' = -\delta(\mathbf{r}-\mathbf{r}') \qquad (7.9)$$

The self-energy of a particle is the energetic response of the system experienced by the particle due to its own presence. Along with the Hartree potential, the exchange-correlation potential of density functional theory represents the local, variationally optimised approximation to this non-local, non-Hermitian, dynamic energy operator.

Local components, $[v_H(\mathbf{r}) + v_{xc}(\mathbf{r})]\delta(\mathbf{r} - \mathbf{r}'')$ can be removed from Σ and relocated within the first term of Eq. (7.9). Therefore, recalling the Kohn-Sham optimised local potential, Eq. (4.51)

$$v_s(r) = v(r) + v_H(r) + v_{xc}(r)$$

Eq. (7.9) can be recast as

$$\left(-\frac{\nabla^2}{2} + v_s(\mathbf{r}) - \varepsilon_i\right)G(\mathbf{r},\mathbf{r}';\varepsilon_i) + \int \Delta\Sigma(\mathbf{r},\mathbf{r}'';\varepsilon_i)G(\mathbf{r}'',\mathbf{r}';\varepsilon_i)d\mathbf{r}'' = -\delta(\mathbf{r}-\mathbf{r}') \qquad (7.10)$$

where $\Delta\Sigma$, describing the non-local, dynamical exchange and correlation effects, is equal to Σ - v_H - v_{xc}, and with

$$G_0^{KS}(\mathbf{r}',\mathbf{r}'';\varepsilon_j)=\sum_{j\neq i}\frac{\psi_i^{KS*}(\mathbf{r}')\psi_i^{KS}(\mathbf{r}'')}{\varepsilon_j-\varepsilon_i^{KS}} \tag{7.11}$$

related to the required Green's function G via the perturbative expansion of Eq. (7.6)

$$G(\mathbf{r},\mathbf{r}';\varepsilon_i)=G_0^{KS}(\mathbf{r},\mathbf{r}';\varepsilon_i)+\int dr_2\int dr_1 G_0^{KS}(\mathbf{r},\mathbf{r}_1;\varepsilon_i)\Delta\Sigma(\mathbf{r}_1,\mathbf{r}_2;\varepsilon_i)G_0^{KS}(\mathbf{r}_2,\mathbf{r}';\varepsilon_i)+\cdots \tag{7.12}$$

then the reference system is now the single-particle approximation of Kohn-Sham, and all quantities except Σ, as yet undefined, dependant on the KS single particle eigenstates.

But what of Σ? Hedin [103] developed an approximation for the self-energy Σ^{GW} in the form of an expansion in terms of the screened Coulomb potential W. For simplicity, Hedin's practise of replacing the space-time coordinates $\mathbf{r}_1$, t_1 with the integer 1 and so on will be adopted here. Thus

$$\begin{aligned}
&\int d1=\int d\mathbf{r}_1\int_{-\infty}^{+\infty}dt_1\\
&1^+=(\mathbf{r}_1,t_1+\tau) \qquad\qquad \tau\to 0,\tau>0\\
&\delta(1,2)=\delta(\mathbf{r}_1-\mathbf{r}_2)\delta(t_1-t_2)\\
&v_2(1,2)=v_2(\mathbf{r}_1,\mathbf{r}_2)\delta(t_1-t_2)
\end{aligned} \tag{7.13}$$

$W(1,2)$ is the potential at point 1 due to a charge and its associated polarisation cloud at point 2, and is expressed by

$$\begin{aligned}
W(1,2)&=\int v_2(1,3)\varepsilon^{-1}(3,2)d3\\
&=v_2(1,2)+\int W(1,3)P(3,4)v_2(4,2)d3d4
\end{aligned} \tag{7.14}$$

where ε^{-1} is defined as the propagating dielectric function

$$\varepsilon^{-1}(1,2)=\left(\delta(1,2)-\int v_2(1,3)P(3,2)d3\right)^{-1} \tag{7.15}$$

which connects the screened and unscreened Coulomb interactions. The polarisation propagator $P(1,2)$ can be expanded

$$P(1,2)=-iG(1,2)G(2,1)+\int G(1,3)G(4,1)W(3,4)G(2,4)G(3,2)d3d4+\cdots \tag{7.16}$$

The GW definition of the self-energy follows from Eq. (7.9), relocating only the local Hartree component such that

$$\left(-\frac{\nabla^2}{2}+v(\mathbf{r})+v_H(\mathbf{r})-\varepsilon_i\right)G(\mathbf{r},\mathbf{r}';\varepsilon_i)+\int\Sigma^{GW}(\mathbf{r},\mathbf{r}'';\varepsilon_i)G(\mathbf{r}'',\mathbf{r}';\varepsilon_i)d\mathbf{r}''=-\delta(\mathbf{r}-\mathbf{r}') \tag{7.17}$$

and is given by Hedin's screened potential expansion

$$\Sigma^{GW}(1,2) = G(1,2)W(1^+,2) - \int G(1,3)G(3,4)G(4,2)W(1,4)W(3,2)d3d4 + \cdots \qquad (7.18)$$

For those readers familiar with Feynman diagrams, the three lowest order contributions from the expansion of P are represent in diagrammatic form in Figure 7-2 and similarly for the self-energy Σ in Figure 7-3

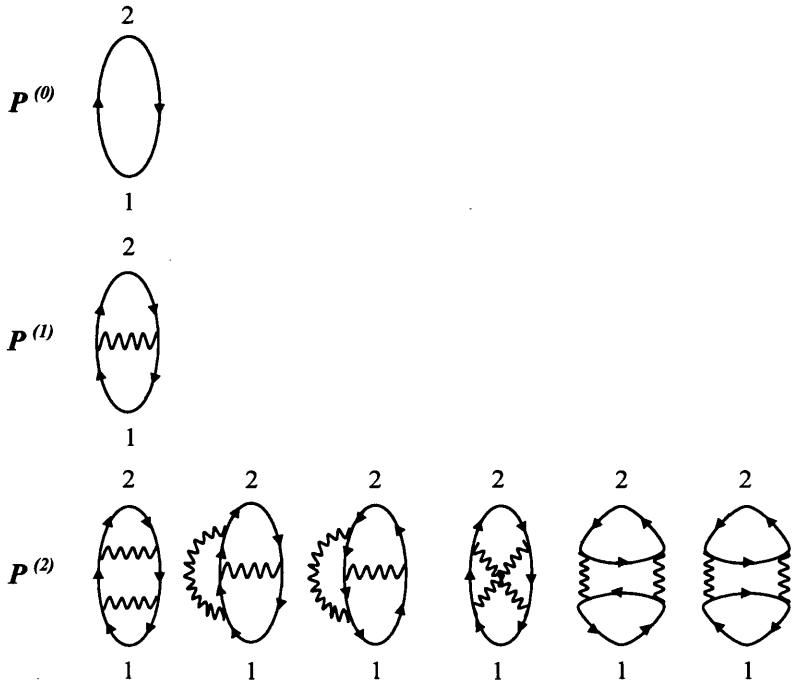


Figure 7-2 Diagrammatic representation of the first three orders in W of contributions from the expansion of the polarisation function P

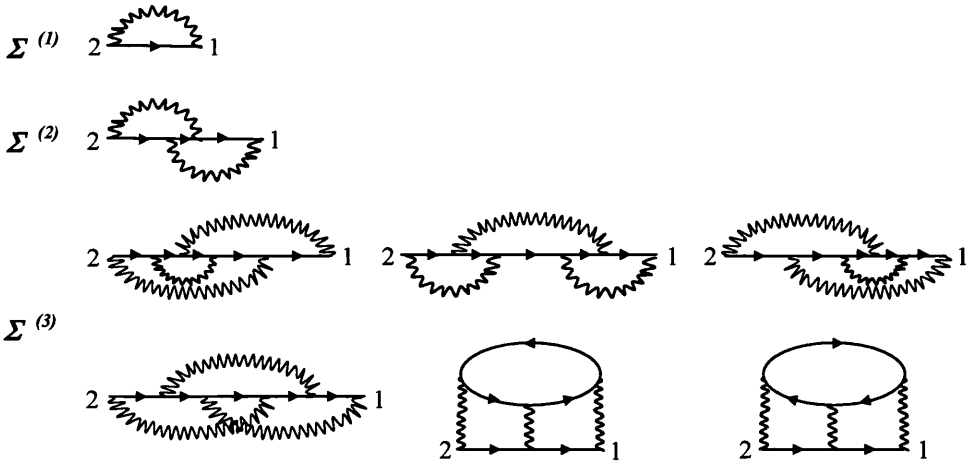


Figure 7-3 Diagrammatic representation of the first three orders in W of contributions from the expansion of the self energy Σ^{GW}

The single-particle Green's function $G(\mu, \nu)$ is depicted by an arrow containing solid line from points ν to μ as in Hedin [103], and the interaction, here the screened Coulomb interaction $W(\mu, \nu)$, as a wavy line between μ and ν.

Defining the vertex function Γ as

$$\Gamma(1,2;3) = \delta(1,2)\delta(1,3) - \int\int\int\int \frac{\delta\Sigma^{GW}(1,2)}{\delta G(4,5)} G(5,6)\Gamma(6,7;3)G(7,4)d4d5d6d7 \qquad (7.19)$$

P and Σ can be written

$$P(1,2) = -i\int\int G(1,3)\Gamma(3,4;2)G(4,1)d3d4 \qquad (7.20)$$

$$\Sigma^{GW}(1,2) = i\int\int W(1^+,3)G(1,4)\Gamma(4,2;3)d3d4 \qquad (7.21)$$

and

$$G(1,2) = G_0(1,2) + \int\int G_0(1,3)\Sigma^{GW}(3,4)G(4,2)d3d4 \qquad (7.22)$$

where G_0 satisfies

$$\left(-\frac{\nabla^2}{2} + v(\mathbf{r}) + v_H(\mathbf{r}) - \varepsilon_i\right) G_0(\mathbf{r},\mathbf{r}';\varepsilon_i) = -\delta(\mathbf{r}-\mathbf{r}') \qquad (7.23)$$

Along with Eq. (7.14), the set of equations Eq. (7.19) to (7.23) are known as Hedin's GW integro-differential equations. In principle, their self-consistent solution solves the many-body problem exactly, although in practice this is a very demanding task computationally [29] and approximations must be made.

Conventionally the random phase approximation (RPA) is often used, where the polarisation function $P(1,2)$ is approximated by its zeroth order in W term, $P_0 = -iG(1,2)G(2,1)$. Correspondingly then, the vertex function $\Gamma(1,2;3)$ is reduced to $\delta(1,2)\delta(1,3)$, simplifying the expression for the self energy.

Indeed, truncating in this manner has less effect than might be supposed. Should the self-energy in the vertex function Eq. (7.19) be approximated to first order, yielding

$$\Gamma(1,2;3) = \delta(1,2)\delta(1,3) + \int\int W(1^+,2)G(1,4)G(5,2)\Gamma(4,5;3)d4d5 \qquad (7.24)$$

then P, Eq. (7.20), reduces to the sum of the leading terms in each order, see Figure 7-4

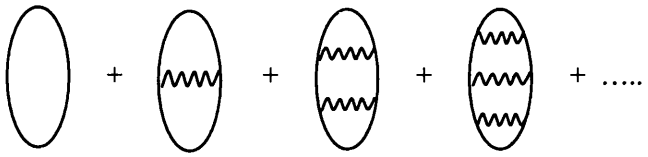


Figure 7-4 The leading diagrams of zeroth, first, and second order in W etc. for P. Often referred to as the ladder-bubble sum.

and likewise for Σ, Eq. (7.21), see Figure 7-5

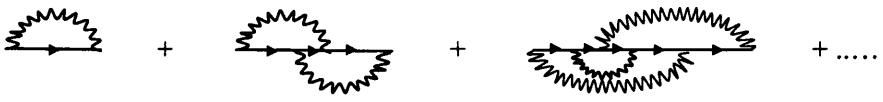


Figure 7-5 The leading diagrams of first, second, and third order in W etc. for Σ.

Hedin [103] reports that this Σ offers no real improvement on Σ estimated purely to first order in W. He concludes that while an infinite sum is required for P in the case of very low electron densities (see also Economou [105]), and that the leading terms appearing in Figure 7-4 will not suffice alone, in the case of higher densities however, it is the quality of G and W that is most important.

A further simplification that is often made [102,106] is the approximation of Green's function G by the optimised Kohn-Sham DFT Green's function G_0^{KS}.

The estimation of the self-energy then follows from the set of equations:

$$P_0(1,2) = -iG_0^{KS}(1,2)G_0^{KS}(2,1) \tag{7.25}$$

$$\varepsilon_{RPA}(1,2) = \delta(1,2) - \int v_2(1,3)P_0(3,2)d3 \tag{7.26}$$

$$W_0(1,2) = \int v_2(1,3)\varepsilon_{RPA}^{-1}(3,2)d3 \tag{7.27}$$

$$\Sigma^{GW}(1,2) = iG_0^{KS}(1,2)W_0(1^+,2) \tag{7.28}$$

The GW self energy as previously mentioned excludes Coulomb effects but is inclusive of all exchange and correlation effects, local and non-local. Working with the assumption that the Kohn-Sham eigenstates already give good approximations to the quasiparticle states, then it is reasonable to use first-order perturbation theory to obtain the correct energy levels of the quantum well. Applying the small self-energy correction $\Sigma(\mathbf{r}, \mathbf{r}'; \varepsilon_i) - v_{xc}(\mathbf{r})\delta(\mathbf{r} - \mathbf{r}')$ then

$$\varepsilon_i \approx \varepsilon_i^{KS} + \left\langle \psi_i^{KS} \middle| \Sigma^{GW}(\varepsilon_i) - v_{xc} \middle| \psi_i^{KS} \right\rangle \tag{7.29}$$

In order to circumvent Eq. (7.29)'s non-linearity, the finite difference-like approximation [106]

$$\Sigma^{GW}(\mathbf{r},\mathbf{r}';\varepsilon_i) \approx \Sigma^{GW}(\mathbf{r},\mathbf{r}';\varepsilon_i^{KS}) + (\varepsilon_i - \varepsilon_i^{KS})\frac{\partial \Sigma^{GW}(\mathbf{r},\mathbf{r}';\varepsilon_i^{KS})}{\partial \varepsilon} \tag{7.30}$$

can be used to recast Eq. (7.29) as

$$\varepsilon_i \approx \varepsilon_i^{KS} + Z_i \left\langle \psi_i^{KS} \middle| \Sigma^{GW}(\varepsilon_i^{KS}) - v_{xc} \middle| \psi_i^{KS} \right\rangle \tag{7.31}$$

where Z_i is referred to as the quasiparticle renormalisation factor

$$Z_i = \frac{1}{1 - \left\langle \psi_i^{KS} \middle| \partial \Sigma^{GW}(\varepsilon_i^{KS})/\partial \varepsilon \middle| \psi_i^{KS} \right\rangle} \tag{7.32}$$

and is equal to the quasiparticle weight

$$Z_i = \int |\psi_i(\mathbf{r})|^2 d\mathbf{r} < 1 \tag{7.33}$$

where ψ_i are the quasiparticle wave functions.

Separating out the Coulomb interaction from the screened potential such that

$$\begin{aligned} \Sigma^{GW}(1,2) &= iG_0^{KS}(1,2)v_2(1,2) + iG_0^{KS}(1,2)[W_0(1^+,2) - v_2(1,2)] \\ &= \Sigma_x^{GW}(1,2) + \Sigma_c^{GW}(1,2) \end{aligned} \tag{7.34}$$

then the expectation value of Σ_x^{GW} can be analytically assessed, yielding

$$\left\langle \psi_i^{KS} \middle| \Sigma_x^{GW} \middle| \psi_i^{KS} \right\rangle = -\sum_j \int\int \frac{\psi_i^{KS*}(\mathbf{r})\psi_j^{KS}(\mathbf{r})\psi_j^{KS*}(\mathbf{r}')\psi_i^{KS}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}' \tag{7.35}$$

The correlation component, explicitly written in the energy domain as

$$\Sigma_c^{GW}(\mathbf{r},\mathbf{r}';\varepsilon) = \frac{i}{2\pi}\int_{-\infty}^{\infty} G_0^{KS}(\mathbf{r},\mathbf{r}';\varepsilon+\varepsilon')[W_0(\mathbf{r},\mathbf{r}';\varepsilon') - v_2(\mathbf{r},\mathbf{r}')]e^{i\varepsilon'\tau}d\varepsilon' \tag{7.36}$$

where τ is an infinitesimal positive time, is usually evaluated with a complex contour integral [106], and is used to set both Z_i and $\left\langle \psi_i^{KS} \middle| \Sigma_c^{GW} \middle| \psi_i^{KS} \right\rangle$.

The above discussion outlines the necessary steps to improve the KS eigenstates in order that they accurately represent the energy levels within the quantum dot, assuming of course that the approximations made to Hedin's *GW* equations are acceptable within this system. The computational implementation of the GW scheme is a subject of considerable intest in the Literature, and in particular, Rojas *et al.* [107], Rieger *et al.* [108] and Steinbeck *et al.* [109] outline some additional simplification to accelerate convergence for larger systems and several potentially useful computational details.

Broadly speaking, the assessment of the *GW* self-energy can fall into one of three categories (self-consistent ($\Sigma = iGW$), partially self-consistent ($\Sigma = iGW_0$), and non-self-consistent ($\Sigma = iG_0W_0$) and only further investigation of this topic will determine the optimum scheme to employ for accurate results for minimum computational effort. The bulk of immediate future work should therefore be focused in this field, although it is possible that with the level of error inherent within the framework of approximations describing the charged nanocrystal, the extra computational effort in self-energy correcting the KS eigenstates will not be at all cost effective and sensible.

7.2.3 STM Tip-QD-Substrate Tunnelling

The ability of a particle to tunnel through a classically forbidden region has been a topic of interest from the very first days of quantum mechanics; rigorous discussions of its basic mathematical detail can be found in any good quantum mechanics textbook (for instance Merzbacher [37], Bohm [64] or Gasiorowicz [100], and a short non-mathematical overview of barrier penetration can be found in Sub-Section 6.2.2). Complex treatments from the literature are varied and diverse: from the time evolution operator method [110] to the non-equilibrium Green's functions of Keldysh [111] (common in many modern tunnelling simulations, for example Ref.'s [112,113,114]) and those of Caroli *et al.* [115]. Duke [116] offers a comprehensive discussion of various tunnelling methods, and much work has been done with TB (Tight Binding - see for example Ref. [117]) and LACO (linear combination of atomic orbitals – see for instance Ref. [118]) models. As an elementary exploration into this wide and varied field, a simple (in comparison at least) perturbation theory route has been followed, similar in its basics to that of Wasshuber [32]. In the spirit of the chapter, this tunnelling scheme is again only given in outline below, and not rigorously detailed.

The STM tip, the quantum dot under study and the substrate can be viewed as one interconnected system, and one which can be represented as a simple electronic circuit of two capacitors, each representing one of the potential barriers (or tunnelling junctions), in series – as illustrated in Figure's 7-6 and 7-7. In this way, tunnelling events can visualised as charge transfers across the said capacitors.

The charge on junction one, J_1, tip to nanocrystal, is

$$q_1 = C_1 V_{J1} \tag{7.37}$$

on junction two, J_2, nanocrystal to substrate,

$$q_2 = C_2 V_{J2} \tag{7.38}$$

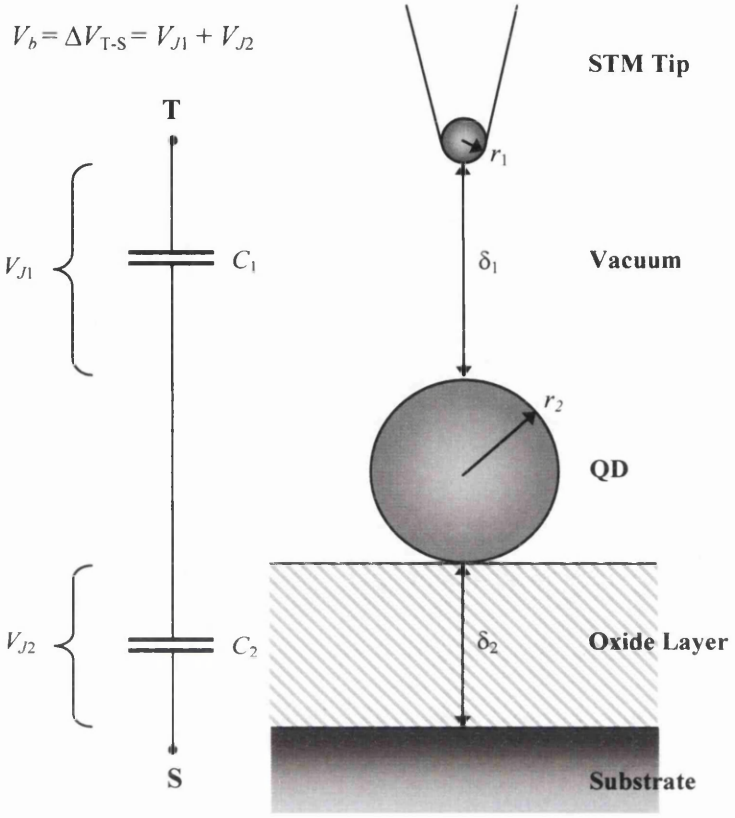


Figure 7-6 Schematic of STM Tip - QD - Substrate system. Active region of Tip modelled as a sphere of radius r_1 and the quantum dot treated as a sphere of radius r_2. Tip-QD and QD-Substrate separations denoted by δ_1 and δ_2 respectively. The potential difference between the Tip-Substrate is V_b, with the voltage drop over junction 1 (Tip-QD), J_1, represented as V_{J1} and likewise V_{J2} for junction 2 (QD-Substrate), J_2. Each junction is modelled as a capacitor, C_1 representing J_1, and C_2 J_2.

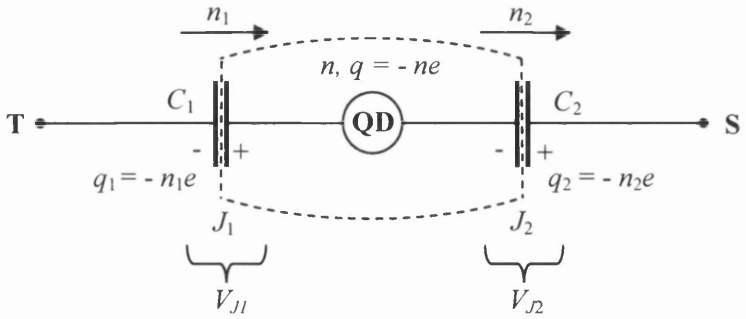


Figure 7-7 Circuit schematic representing Tip-QD-Substrate system of Figure 7-6. n_1 represents the number of electrons tunnelling across J_1, the charge on J_1 simply $q_1 = -n_1e$, and similarly for junction 2. The number of electrons stored within the QD is n with a charge of q.

and on the nanocrystal itself

$$q = q_1 - q_2 \tag{7.39}$$

q comprising of the n electrons residing on the nanocrystal, the difference between the n_1

electrons tunnelling into the quantum dot, and the n_2 electrons tunnelling out of the dot to the substrate.

With the natural assumption that the overall system can be considered to reside in its lowest possible energetic configuration, then for a given tip-substrate potential difference, only those electron transfers events that reduce the overall energy can be considered to be statistically likely. Measuring the system via its Helmholtz free energy, F, defined as the difference between the total energy stored within the system, E_T, and any work done by the system, W,

$$F = E_T - W \tag{7.40}$$

then the change in this energy for a tunnelling event is a measure of the probability of this tunnelling event. As mentioned above, given that physical systems will always reside in their lowest possible energy configurations, charge transfer events in which the free energy decreases i.e. $\Delta F\ (= F_f - F_i) < 0$, are logically favoured.

For the very simple treatment applied to the P-KS-CB data in Section 6.2, the total stored energy was the energy stored within the SnO_2 grain of n electrons. In a sophisticated analysis, the total energy of the n electrons in their particular eigenstates, the energy of the thermal electrons, the Coulomb and exchange-correlation energies would all need to be evaluated, along with allowances made for the effect of the shift in the Fermi level between the n and $n \pm 1$ systems on all these quantities. Importantly, this Fermi level movement can no longer seen as a movement 'up' the energy well, but rather that the energy levels and confining potential must be dragged down relative to the Fermi level, 'fixed' at a voltage V_{J2} above the zero of the system. Hence, E_T is not a particularly simple object to evaluate. The elementary approach used in Section 6.2 is based on the assumption that since it is the change in free energy that is being sought, a reasonable approximation of ΔE_T can be obtained via the following argument. As the thermal electron population (electrons excited into energy levels above the Fermi level) remains relatively constant over charge transfer events, see Section 6.2 for detail, then if the energy levels are seen as moving down, and not the Fermi level moving up the energy well, then to first approximation, the energy of the electrons (without interaction effects) due to their presence in these eigenstates, will remain fairly constant with regard to small increases or decreases in the population of electrons. Consequentially, if the energy of electron-electron interactions are simply estimated by the classical potential energy of the charge stored within the two capacitors, then the change in E_T can be expressed as the change in the capacitative energy of the system plus the energy taken to move the Fermi level (merely the difference of the two Fermi levels in the old, freely moving

Fermi level reference frame)[ii].

From the classical expression for the energy stored in a capacitor, the total capacitative energy of the system is

$$E_{cap} = \frac{V_{J1}^2 C_1 + V_{J2}^2 C_2}{2} \tag{7.41}$$

Since $q = ne$ then

$$V_{J1} = \frac{C_2 V_b + ne}{C_1 + C_2} \tag{7.42}$$

and similarly

$$V_{J2} = \frac{C_1 V_b - ne}{C_1 + C_2} \tag{7.43}$$

therefore Eq. (7.41) can be written

$$E_{cap} = \frac{C_1 C_2 V_b^2 + (ne)^2}{2(C_1 + C_2)} \tag{7.44}$$

The tunnelling of an electron over either of the junctions will cause a change in the voltages across the junctions of $\pm e/(C_1 + C_2)$ from Eq. (7.42) and Eq. (7.43). Taking the capacitance of each junction as constant, then this voltage change causes a corresponding polarisation charge of $q_{pol} = C\Delta V$. The charge that then must be 'replaced' is $-e - q_{pol}$, which of course is equal to the negative of the charge gained/lost over the junction through which a tunnelling event did not occur. Therefore, the work done by the system is $W = -(-e - q_{pol})V_b$, or in general

$$W = -\frac{eV_b(n_1 C_2 + n_2 C_1)}{C_1 + C_2} \tag{7.45}$$

As such, the change in Helmholtz free energy for a single electron transfer event over junction 1 or 2 for the elementary approach used in Section 6.2 will therefore be

$$\Delta F_{J1}^{\pm} = E_f^{n\pm1} - E_f + \frac{e(e \pm 2ne \pm V_b C_2)}{2(C_1 + C_2)} \tag{7.46}$$

[ii] Indeed, this idea has been simply employed by Johnson *et al.* [119] in their T~0K 2-dimensional QD system to estimate the movement of the Fermi level on the addition of a single electron to the dot i.e.

$$E_f(N+1) - E_f(N) = \frac{e^2}{C} + (E_{N+1} - E_N)$$

C the capactitance between the QD and the gate, and E_N the total energy of the N electron system etc.

$$\Delta F_{J2}^{\pm} = E_f^{n\mp1} - E_f + \frac{e\left(e \mp 2ne \pm V_b C_1\right)}{2\left(C_1 + C_2\right)} \tag{7.47}$$

the +ve on the ΔF superscript indicating a tunnelling event from left to right across the junction (as depicted in Figure 7-7), and conversely the –ve; right to left.

Assuming that the STM tip is locally spherical [33] then the capacitances C_1 and C_2 follow readily from the image charge method applied to two spheres of dissimilar radii (see Appendix IV, Mutual Capacitance of Two Spheres using the Image Charge Method). For the system of Figure 7-8 with the parameters of Section 6.2 (W tip with a radius of curvature of 0.9nm, Si substrate, tip-nanocrystal separation (δ_1) of ~0.7nm, SiO_2 thickness (δ_2) of 1nm and a R = 4nm SnO_2 nanocrystal), the capacitances $C_1 \approx 9.55\times10^{-20}$ F and $C_2 \approx 4.44\times10^{-18}$ F follow.

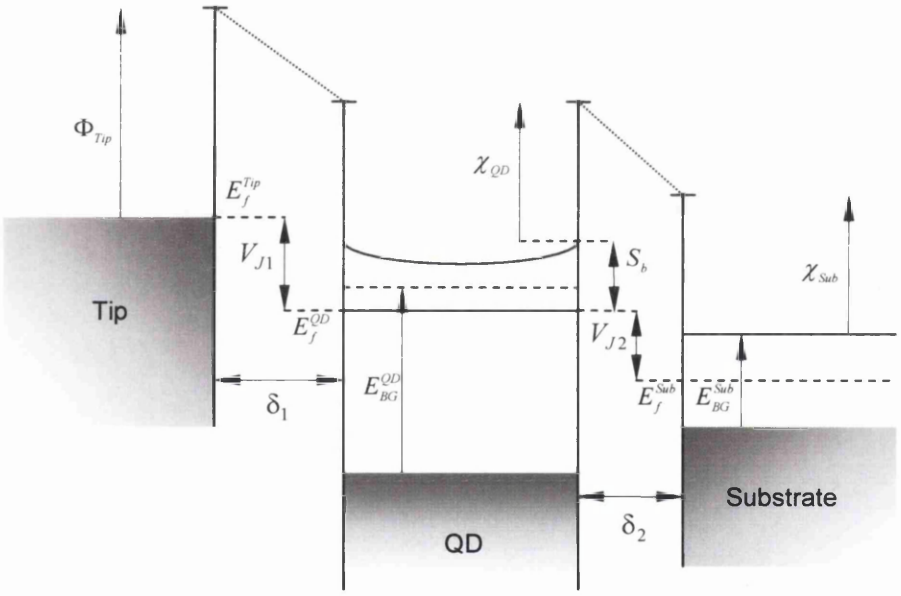


Figure 7-8 Cross sectional energy diagram of the Tip-QD-Substrate system. Labelling conventions remain the same as Figure 7-6 with the addition of χ representing the electron affinity, Φ the work function, and E_{BG} the semiconductor band gap.

Then, using the expression for tunnelling current developed in Appendix IV, Fermi's golden rule:

$$I = e\Gamma(\Delta F) = \iint \frac{2\pi e}{\hbar}\left|H_{fi}\right|^2 D_i D_f f_i(1-f_f)\delta(E_i - E_f + \Delta F)dE_i dE_f \tag{7.48}$$

the i and f subscripts representing the initial and final states of the tunnelling electron, a first approximation of the tunnelling current can readily be achieved.

The density of states of the tip and the substrate are calculated with standard expressions for the

density of states of a metal and semiconductor respectively [35]:

$$D_{tip}(E) = \frac{1}{2\pi^2}\left(\frac{2m_{tip}}{\hbar^2}\right)^{3/2}\sqrt{E} \tag{7.49}$$

$$D_{sub}(E) = \frac{1}{2\pi^2}\left(\frac{2m_{sub}}{\hbar^2}\right)^{3/2}\sqrt{E - E_{CBB}} \tag{7.50}$$

The tunnelling matrix element H_{fi} ($=\langle\psi_f|H_1|\psi_i\rangle$) is more difficult to assess. For a detailed simulation, a suitable approach would be to use the

$$H_{fi} = -\frac{\hbar}{2m}\int_S d\boldsymbol{\sigma}\cdot\left(\psi_f^*\nabla\psi_i - \psi_i^*\nabla\psi_f\right) \tag{7.51}$$

expression of Bardeen [120] as in Tersoff and Hamann [33], the integral carried out over any surface within the barrier region. However, an adequate estimate of the matrix element can be made from the use of the Wentzel-Kramers-Brillouin (WKB) approximation that

$$\frac{\psi_f}{\psi_i} \approx \exp\left(-\int dx\sqrt{(2m/\hbar^2)(V(x)-E)}\right) \tag{7.52}$$

$V(x)$ describing the potential barrier, see Blakemore [35] or Bohm [64]. For the parameters and P-KS-CB data of Section 6.2, this simplification and those of Eq. (7.48) results in the Coulomb staircase charging profile shown in Figure 6-27, reprinted below as Figure 7-9, and the accompanying current-voltage plot shown in Figure 6-28, reprinted below as Figure 7-10.

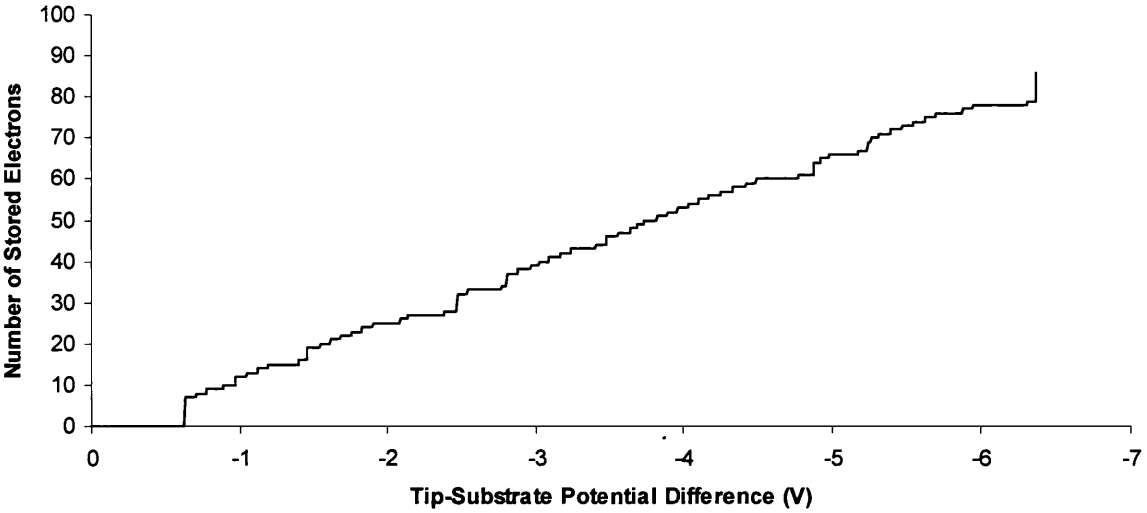


Figure 7-9 Plot of the electron addition to a R = 4nm grain against tip-substrate potential difference for a simple free energy based simulation, using the variable n_d data generated from the P-KS-CB scheme. The tip (W) to grain separation is taken to be 0.7nm, and the oxide layer on the Si substrate is taken to be 1 nm thick. Re-print of Figure 6-27.

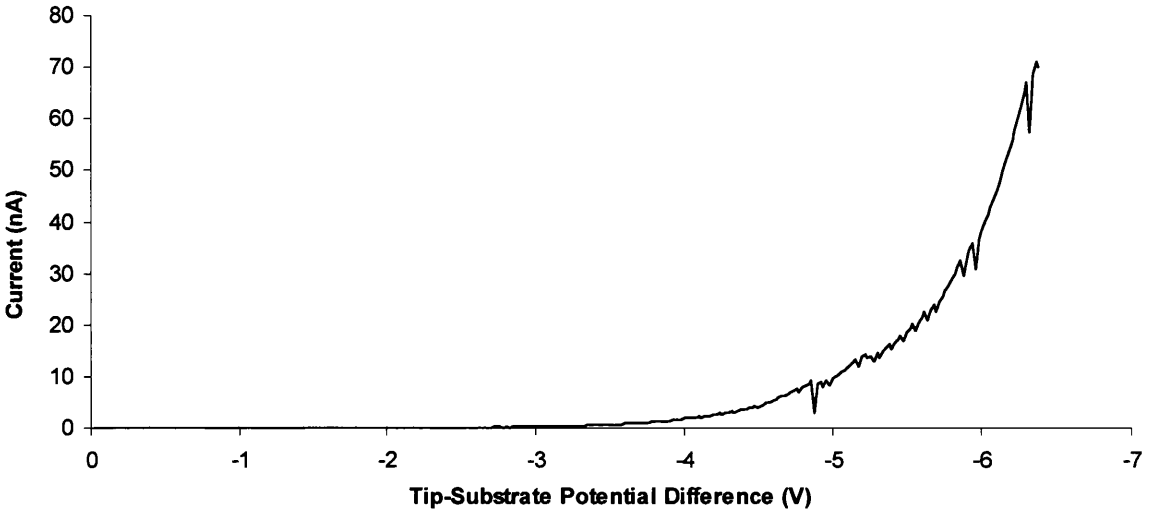


Figure 7-10 Plot of the tip-substrate current against the tip-substrate potential difference for the basic tunnelling simulation of room temperature charge storage on an R = 4nm SnO_2 nanocrystal, using the variable n_d data generated from the P-KS-CB scheme. The tip (W) to grain separation is taken to be 0.7nm, and the oxide layer on the Si substrate is taken to be 1nm thick. Re-print of Figure 6-28.

As discussed in the referenced section, these simulations are in very good agreement with the experimental data (see Figures 6-29 and 6-30) even with this primitive model of the electron tunnelling; most importantly, the modal voltage interval between charging events is calculated at 0.065V compared to the average period of the experimentally observed dI/dV fluctuations of 0.07V.

Given the success of even the very simplistic application of this first-order perturbation theory approached shown here, it is possible that more sophisticated non-equilibrium Green's function approaches (Keldysh [111] etc.) and the like, will not be necessary for the particular circumstances and parameters of interest here.

Future work would focus on such things as improving the calculation of the tunnelling matrix element, a more accurate assessment of the energy differences of the n and $n \pm 1$ states of the QD, and improving the modelling of the tip-QD junction (more realistic tip geometry and so a more accurate C_1). Also important would be the inclusion of additional potential sources of current; for example, including the effects of co-tunnelling (tunnelling events where the overall change in Helmholtz free energy is negative, although the free energy difference over either of the junctions can be positive) and thermionic emission from the STM tip. It may also be possible that under these particular circumstances non-charging electron transfer through the surface electron states will provide a sizable contribution to the tunnelling current. For this scenario, the change in

Helmholtz free energy would not include the energy shift of the Fermi level between n and $n \pm 1$ systems, making transfer events more likely at lower biases. However, more detail would have to be known about the energetic structure of the surface states, or at the very least, the density of the unoccupied surface states, before any quantitative assessment of this current can be made.

7.2.4 Charge Leakage

The charge writing treatment to date evaluates the SnO_2 nanocrystal in isolation. However, in polycrystalline films the individual quantum dots rest in close proximity to each other. Charge transfer between grains is not only likely to influence the charging process and the duration over which charge can be stored, but intimately affect the shape of the band structure and the entire discrete energy spectrum. The first stage in the assessment of this phenomenon might be to treat inter grain tunnelling through the same energy considerations as were used to evaluate the tip-grain-substrate tunnelling, though any in-depth treatment of the charge writing process must eventually move beyond the spherical symmetry of the individual grains, and treat the charge distribution across the grain 'clusters' as a whole.

Consideration would naturally need to be given at this point to the sinter-necks between grains and their effect on the discrete energy levels. In studying this part of the charging process, it may be possible to also further the primitive gas sensing film sensitivity model of Chapters 2 and 6 to something a little more meaningful, and treat the increases in sensor responsiveness below R = 10nm and R = 3nm in detail.

7.3 Concluding Remarks

There can be no doubt that the potential impact of the charge writing process is considerable, its possible applications wide ranging; from the nanoelectronic to the biomedical. The recent experimental advances in this field, particularly with such things as the STM charge injection of 4nm radius SnO_2 nanocrystals, bring the true dawn of this budding technology ever closer.

This work has aimed to shed a little more light on the relatively neglected underlying physics of the process, and has perhaps gone some way to explaining and interpreting some of this nanoscale experimental SnO_2 data. This concluding Chapter has summarised the models used and developed over the preceding Chapters, and has offered a final synopsis and discussion of their results and implications. It has indicated what has been achieved, and what possible avenues could be pursued to further extend this work, and improve the understanding of this new and exciting branch of science.

Appendices

Appendix I

AI.1 Approximation of the Fermi-Dirac Integral

The family of Fermi Dirac integrals taking the form

$$F_j(y_0) = \int_0^\infty \frac{y^j dy}{1 + e^{y - y_0}} \tag{AI.1}$$

and their close cousins

$$F_j(y_0) = \Gamma(j+1)\mathscr{F}_j(y_0) \tag{AI.2}$$

arise from the evaluation of the statistical magnitudes of an electron gas, integrating over all possible states weighted by the Fermi-Dirac function. They have many applications: Einstein's relation for degenerate semiconductors is a function of $F_{-1/2}$, the 'supply' function is dependent on F_0, and $F_{3/2}$ can appear in electronic energy density expressions. However, perhaps the most widespread occurrence is that of $F_{1/2}$ in the carrier density expression Eq. (1.35).

Generalised approximations of the Fermi-Dirac integrals can be found for $-1/2 < j < 5/2$ with an error between 0.7 % and 1.2 % [121], and higher accuracy expressions can commonly be found for j = 3/2 (0.63%) and j = 1/2 (0.53%), see Aymerich-Humet *et al.* [122]. Blakemore [57] presents a review of approximations for $\mathscr{F}_{1/2}(y_0)$ in particular, several of which are of higher order degrees of accuracy.

This work uses the Chebyshev polynomial approximations to $F_{1/2}$ of Werner and Raymann [123],

and as such, the Fermi-Dirac integral is represented with the two expansions:

$$F_{1/2}(y_0) = e^{y_0} \sum_{n=0}^{5} a_n e^{n y_0} \qquad -\infty < y_0 \leq +1 \qquad \text{(AI.3)}$$

$$F_{1/2}(y_0) = y_0^{3/2} \left(\frac{2}{3} + \sum_{n=0}^{5} \frac{b_n}{y_0^{2n+2}} \right) \qquad +1 < y_0 < +\infty \qquad \text{(AI.4)}$$

With the coefficients of Table AI-1, these give $F_{1/2}$ to 0.02 % and 0.05 % respectively; adequate for the purposes of this work.

Table AI-1 Table of coefficients for the Werner and Raymann approximation of $F_{1/2}$

n	a_n	b_n
0	+ 0.88607596	+ 0.843500
1	- 0.30871705	+ 0.710809
2	+ 0.14638520	- 3.712456
3	- 0.05843877	+ 6.705628
4	+ 0.01431771	- 5.594877
5	- 0.00150176	+ 1.777787

AI.2 Gamma and Beta Functions

In the infinite limit, the gamma function, plotted in Figure AI-1, can be defined as [65]

$$\Gamma(z) = \lim_{n \to \infty} \frac{1 \cdot 2 \cdot 3 \cdots n}{z(z+1)(z+2)\cdots(z+n)} n^z \qquad z \neq 0, -1, -3 \ldots \qquad \text{(AI.5)}$$

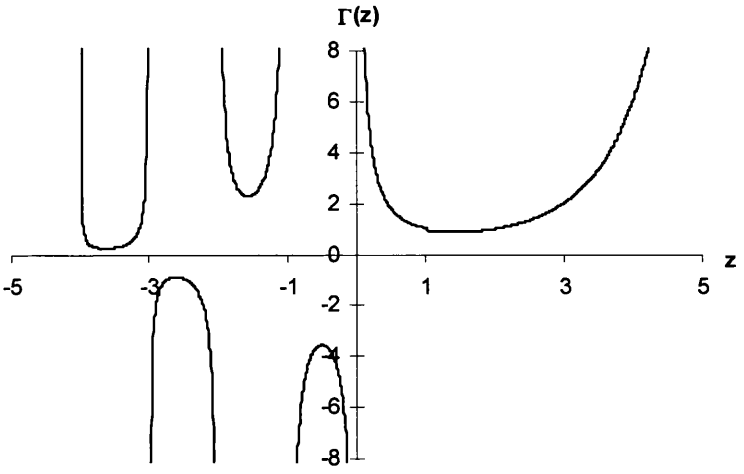


Figure AI-1 Plot of the gamma function, $\Gamma(z)$.

Replacing z with $z + 1$ in Eq. (AI.5) then

$$\Gamma(z+1) = \lim_{n\to\infty} n\frac{z}{z+n+1}\frac{1\cdot 2\cdot 3\cdots n}{z(z+1)(z+2)\cdots(z+n)}n^z \qquad \text{(AI.6)}$$
$$= z\Gamma(z)$$

For $z = 1$

$$\Gamma(1) = \lim_{n\to\infty}\frac{1\cdot 2\cdot 3\cdots n}{1\cdot 2\cdot 3\cdots n(n+1)}n^1 = 1$$

then applying Eq. (AI.6)

$$\Gamma(2) = 1$$
$$\Gamma(3) = 2\Gamma(2) = 2$$

and so on until

$$\Gamma(n) = (n-1)! \qquad \text{(AI.7)}$$

for n a positive integer. From this, the gamma function can be interpreted as a generalisation of the elementary factorial function, extending the latter to negative and non-integer values. For this reason, the gamma function is also referred to the factorial function [50].

One alternative definition[i] of the gamma function, called the Euler integral, is

$$\Gamma(z) = \int_0^\infty e^{-t}t^{z-1}dt \qquad \Re(z) > 0 \qquad \text{(AI.8)}$$

the restriction on the real component of z necessary to prevent divergence, see Arfken [65]. Using the above, then

$$m!n! = \lim_{a\to\infty}\int_0^a e^{-u}u^m du\int_0^a e^{-v}v^n dv \qquad \Re(m) < -1 \text{ and } \Re(n) < -1$$

and substituting x^2 and y^2 for u and v respectively yields

$$m!n! = \lim_{a\to\infty} 4\int_0^a e^{-x^2}x^{2m+1}dx\int_0^a e^{-y^2}y^{2n+1}dy$$

which in polar coordinates, with $x = r\cos\theta$ and $y = r\sin\theta$ can be written

$$m!n! = \lim_{a\to\infty} 4\int_0^a e^{-r^2}r^{2m+2n+3}dr\int_0^{\pi/2}\cos^{2m+1}\theta\sin^{2n+1}\theta\,d\theta = 2(m+n+1)!\int_0^{\pi/2}\cos^{2m+1}\theta\sin^{2n+1}\theta\,d\theta$$

---

[i] The third definition of the gamma function is known as Weierstrass's form:

$$\Gamma(z) = \lim_{n\to\infty}\frac{1}{z}\prod_{m=1}^{n}(1+z/m)^{-1}n^z$$

Rearranged, this is generally referred to as the beta function,

$$B(m+1,n+1)=2\int_0^{\pi/2}\cos^{2m+1}\theta\sin^{2n+1}\theta d\theta=\frac{m!n!}{(m+n+1)!}=\frac{\Gamma(m+1)\Gamma(n+1)}{\Gamma(m+n+2)} \qquad \text{(AI.9)}$$

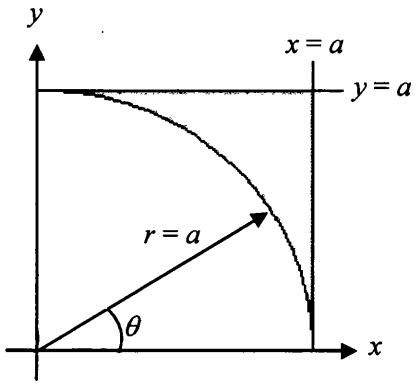


Figure AI-2 Diagram of Cartesian and polar coordinates. Shaded region indicates area excluded from beta function integral on transformation between coordinate systems.

The transformation to polar coordinates results in the shaded region of Figure AI-2 being neglected from the integration. However, this is of no importance as within this region the maximum value of the integrand is $e^{-a^2}a^{2m+2n+3}$, which vanishes so rapidly as $a \rightarrow \infty$ that the integral over that region vanishes anyway.

The beta function is of particular use in Chapter 3 in evaluating the integral of $\sin^{2l+1}\theta$. Letting $m = -½$ in Eq. (AI.9) then

$$2\int_0^{\pi/2}\sin^{2n+1}\theta d\theta=\frac{\Gamma(1/2)\Gamma(n+1)}{\Gamma(n+3/2)}$$

where

$$\Gamma(n+1)=n!$$

$$\begin{aligned}\Gamma(n+3/2)&=(n+1/2)\Gamma(n+1/2)\\&=(n+1/2)(n-1/2)(n-3/2)....(1/2)\Gamma(1/2)\\&=\frac{(2n+1)(2n-1)(2n-3)\ldots1}{2^{n+1}}\Gamma(1/2)\\&=\frac{(2n+1)!!}{2^{n+1}}\Gamma(1/2)\end{aligned}$$

and

$$\Gamma(1/2)=2\int_0^{\infty}e^{-t^2}dt=\sqrt{\pi}$$

therefore

$$\int_0^{\pi} \sin^{2n+1}\theta \, d\theta = \frac{\sqrt{\pi}\, 2^{n+1}\, n!}{\sqrt{\pi}(2n+1)!!} \qquad \text{(AI.10)}$$

$$= 2\frac{(2n)!!}{(2n+1)!!}$$

using $2^n n! = (2n)!!$

Appendix II

AII.1 Orthogonality of the Associated Legendre Functions for Identical *m*

From the definition of $P_l^m(x)$ using Eq. (3.61) and Rodrigues' formula Eq. (3.62), it is possible to write [65]

$$\int_{-1}^{1} P_p^m(x)P_q^m(x)dx = \frac{(-1)^m}{2^{p+q}\,p!q!}\int_{-1}^{1}(x^2-1)^m\frac{d^{p+m}}{dx^{p+m}}(x^2-1)^p\frac{d^{q+m}}{dx^{q+m}}(x^2-1)^q dx \qquad \text{(AII.1)}$$

If $p \neq q$, initially assume that $p < q$. Then on integrating by parts $q + m$ times

$$\int_{-1}^{1} P_p^m(x)P_q^m(x)dx = \frac{(-1)^m(-1)^{q+m}}{2^{p+q}\,p!q!}\int_{-1}^{1}(x^2-1)^q\frac{d^{q+m}}{dx^{q+m}}\left((x^2-1)^m\frac{d^{p+m}}{dx^{p+m}}(x^2-1)^p\right)dx \qquad \text{(AII.2)}$$

as all integrated parts will equal zero provided they contain a $(x^2\text{-}1)$ factor. Using Leibniz's theorem (see footnote [v] of Section 3.3), the RHS integrand can be expanded as

$$(x^2-1)^q\frac{d^{q+m}}{dx^{q+m}}\left((x^2-1)^m\frac{d^{p+m}}{dx^{p+m}}(x^2-1)^p\right) = (x^2-1)^q\sum_{i=0}^{q+m}\frac{(q+m)!}{i!(q+m-i)!}\frac{d^{q+m-i}}{dx^{q+m-i}}(x^2-1)^m\frac{d^{p+m+i}}{dx^{p+m+i}}(x^2-1)^p \qquad \text{(AII.3)}$$

The maximum power of x within the term $(x^2\text{-}1)^m$ is x^{2m} and x^{2p} within $(x^2\text{-}1)^p$, and so the inequalities

$$\begin{aligned} q+m-i &\le 2m \\ p+m+i &\le 2p \end{aligned} \qquad \text{(AII.4)}$$

must hold, implying that

$$q \le p \qquad \text{(AII.5)}$$

contradicting the initial assumption $p < q$. Therefore, the RHS integrand must equal zero. Similarly, should the assumption that $p > q$ be made, then after integrating $p + m$ times it is concluded that $p \le q$, implying that the RHS must again equal zero.

Now let $p = q$, then from Eq. (AII.3) $i = q - m$, and therefore Eq. (AII.2) becomes

$$\int_{-1}^{1}\left(P_p^m(x)\right)^2 dx = \frac{(-1)^{p+2m}(p+m)!}{2^{2p}\,p!p!(p-m)!(2m)!}\int_{-1}^{1}(x^2-1)^p\frac{d^{2m}}{dx^{2m}}(x^2-1)^m\frac{d^{2p}}{dx^{2p}}(x^2-1)^p dx \qquad \text{(AII.6)}$$

With the substitutions

$$\begin{aligned} \frac{d^{2m}}{dx^{2m}}(x^2-1)^m &= \frac{d^{2m}}{dx^{2m}}(x^{2m} - mx^{2m-2} + \cdots) \\ &= (2m)! \end{aligned} \qquad \text{(AII.7)}$$

and

$$\int_{-1}^{1}(x^2-1)^p\,dx=(-1)^p\int_0^{\pi}\sin^{2p+1}\theta d\theta$$
$$=2\frac{(-1)^p(2n)!!}{(2n+1)!!}$$
$$=\frac{(-1)^p 2^{2p+1}\,p!\,p!}{(2p+1)!} \qquad \text{(AII.8)}$$

where the double factorials have been reduced to normal factorials using the relations,

$$(2n)!!=2^n n! \quad \text{and} \quad (2n+1)!!=\frac{(2n+1)!}{2^n n!} \qquad \text{(AII.9)}$$

Eq. (AII.6) can be reduced to

$$\int_{-1}^{1}\left(P_p^m(x)\right)^2\,dx=\frac{2}{2p+1}\frac{(p+m)!}{(p-m)!} \qquad \text{(AII.10)}$$

Combining the two case $p = q$ and $p \neq q$, then

$$\int_{-1}^{1}P_p^m(x)P_q^m(x)\,dx=\frac{2}{2p+1}\frac{(p+m)!}{(p-m)!}\delta_{p,q} \qquad \text{(AII.11)}$$

The orthogonality of the Legendre polynomials follows as a special case of Eq. (AII.11) (m = 0). While it is possible to develop an orthogonality relation for associated Legendre functions of identical lower index and differing upper, see for instance Arfken and Weber [65], this is not required in the context of this work.

AII.2 Introduction to Eigenvalues and Eigenvectors

For a linear transformation A, represented as a matrix, a scalar λ is considered its eigenvalue and a nonzero vector x its eigenvector if

$$Ax=\lambda x \qquad \text{(AII.12)}$$

This is a very important relation, particularly in quantum mechanics, since any observable in nature is taken to be representable as a linear operator A and that any measurement of that observable must equal one of the eigenvalues of A. As such, a short treatment of Eq. (AII.12) and a discussion of the properties of A's eigenstates will not only be informative, but of considerable use at several points within the main body of this work.

Two matrices, A and B, are considered to be similar if there exists an invertible matrix P such that

$$B = P^{-1}AP \tag{AII.13}$$

Conventionally, B is described as being related to A via a similarity transformation.

Importantly, if A and B are similar, then from Eq. (AII.12)

$$P^{-1}Ax_i = \lambda_i P^{-1}x_i$$

which can be written

$$P^{-1}APP^{-1}x_i = \lambda_i P^{-1}x_i$$

with the identity $I = PP^{-1}$. Defining the vector y_i as $y_i = P^{-1}x_i$, then using Eq. (AII.13)

$$By_i = \lambda_i y_i$$

This means that if A and B are similar, then A and B must have the same eigenvalues.

Moreover, if $X_i = (p_{1i}, p_{2i} \ldots p_{ni})$ is an eigenvector of A, defined over the vector space V_n, corresponding to the eigenvalue λ_i, where $i = 1, \ldots, n$, so that $AX_i = \lambda_i X_i$, then if the eigenvectors span V_n, the matrix P, with elements p_{ij}, is a diagonalising matrix for A such that

$$D = P^{-1}AP = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix} \tag{AII.14}$$

since, if the n vectors X_i span V_n, they are linearly independent, and therefore det $P \neq 0$ and P^{-1} exists (a proof of which can be found in Byron and Fuller [69]). Then as $D = d_{ij} = \lambda_i \delta_{ij}$ and $AX_i = \lambda_i X_i$,

$$\begin{aligned} \sum_k a_{ik} p_{kj} &= \lambda_i p_{ij} \\ &= \sum_k p_{ik} \delta_{kj} \lambda_j \\ &= \sum_k p_{ik} d_{kj} \end{aligned}$$

for i and j extending from 1 to n. Therefore

$$AP = PD$$

which, since P has an inverse, yields Eq. (AII.14)[i],

$$D = P^{-1}AP$$

---

[i] Interestingly, $(X_i, X_j) = \sum_{k=1}^{n} p_{ki}^* p_{kj} = \delta_{ij}$ i.e. $P^\dagger P = I$ then $P^\dagger = P^{-1}$ and so the diagonalising matrix P will always be unitary.

Unfortunately, this is not particularly useful in itself as a method for practically computing the eigenvalues of A, as the diagonalising matrix P can only be constructed if the eigenvectors of A are known, and generally the eigenvalues of a matrix are needed first to determine the eigenvectors.

Forward then the QR-factorisation method, which as its name suggests is based on the factorisation of a matrix into two components, Q which is unitary and R which is upper triangular[ii]. It is an iterative method that gradually transforms the matrix into a format from which its eigenvalues can easily be extracted. As applicable to the finite difference approximations of this work, the matrix A is taken to be real and tridiagonal.

Let $A_0 = A$ and factor $A_0 = Q_0R_0$, then forming a matrix A_1 such that $A_1 = R_0Q_0$; next factor $A_1 = Q_1R_1$ and compute $A_2 = R_1Q_1$ and so forth, determining

$$A_s = Q_sR_s \tag{AII.15}$$

then computing

$$A_{s+1} = R_sQ_s \tag{AII.16}$$

Q^{-1} is required to exist by our definition of Q (although incidentally in a real space unitary Q is orthogonal Q), therefore $R_s = Q_s^{-1}A_s$ and $A_{s+1} = Q_s^{-1}A_sQ_s$, hence A_{s+1} is similar to A_s. Following this through to its conclusion, then A_{s+1} is similar to $A_0 = A$ for all s, implying that all A's have the same eigenvalues.

Digressing for a moment, the actual factorisation of A_s into Q_sR_s is achieved using Given's method (see for example Wilkinson [124]). In essence, plane rotations are used to zero matrix subdiagonal entries such that multiplying A_s from the left by the matrix C_2 sets the a_{21} element of C_2A_s (denoted by $a_{21}^{(2)}$) to zero, then multiplying by the matrix C_3, acts to set the element a_{32} of $C_3C_2A_s$ ($a_{32}^{(3)}$) to zero and so on, until

$$C_nC_{n-1}\ldots C_3C_2A_s = R_s \tag{AII.17}$$

where R_s is upper triangular as required.

C_j is constructed simply: in rows j-1 and j, and columns j-1 and j, lies the 2×2 submatrix

[ii] all elements below the main diagonal are equal to zero

$$\begin{bmatrix} \cos\theta_j & \sin\theta_j \\ -\sin\theta_j & \cos\theta_j \end{bmatrix} \tag{AII.18}$$

with all other elements along the main diagonal set to 1 and all other entries set to zero. θ_j is such that

$$a_{jj-1}^{(j)} = -\sin\theta_j a_{j-1j-1}^{(j-1)} + \cos\theta_j a_{jj-1}^{(j-1)} = 0$$

i.e.

$$\theta_j = \arctan\left(a_{jj-1}^{(j-1)} / a_{j-1j-1}^{(j-1)}\right) \tag{AII.19}$$

These C_j are orthogonal and as such their product, and the inverse of this product - Q_s, is also orthogonal

$$Q_s = (C_n C_{n-1} \ldots C_3 C_2)^{-1} = C_2^T C_3^T \ldots C_{n-1}^T C_n^T \tag{AII.20}$$

which yields

$$A_{s+1} = C_n C_{n-1} \ldots C_3 C_2 A_s C_2^T C_3^T \ldots C_{n-1}^T C_n^T \tag{AII.21}$$

from Eq. (AII.16)

Returning to the main argument, repeated application of Eq. (AII.15) and Eq. (AII.16) will generate a series of A's with identical eigenvalues. Importantly, as $s \to \infty$ A_s is found to become upper triangular. A lengthy discussion of this trend and formal proofs of convergence of A_s to its upper triangular form, for a variety of eigenvalue constraints, can be found in Wilkinson [124]. However, these facts are not germane to the discussion in hand but, nonetheless, the interested reader might find satisfaction with the following, less rigorous argument:

As $s \to \infty$, then for convergence, it is required that $A_{s+1} \to A_s$ i.e. $Q_s^{-1} A_s Q_s \to A_s$. While this condition would be satisfied for any A_s and Q_s that commute, for it to hold true in general for any real tridiagonal matrix A, then given the form of Q_s, Q_s must tend to I i.e $\theta_j \to 0$ for all j. This implies that a_{jj-1} for all j must tend to zero, and thus while the upper diagonals of A can become nonzero in the earlier iterations (and will unless A, and therefore A_s, is symmetric), its lower diagonals remain at zero, and asymptotically A_s's subdiagonal tends to zero, thus in the asymptotic limit, A_s becomes upper triangular.

To see the advantage of an upper triangular A_s over a tridiagonal A_s in ascertaining A's eigenvalues, the concept of determinants needs to be briefly discussed. A determinant of an $n \times n$ matrix A is a scalar quantity associated with that matrix and is written

$$\det A = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} \tag{AII.22}$$

For $n = 1$, the determinant is

$$\det A = a_{11} \tag{AII.23}$$

and for $n \geq 2$, it can be represented as

$$\det A = a_{j1}C_{j1} + a_{j2}C_{j2} + \cdots + a_{jn}C_{jn} \tag{AII.24}$$

or

$$\det A = a_{1k}C_{1k} + a_{2k}C_{2k} + \cdots + a_{nk}C_{nk} \tag{AII.25}$$

where j and k run from 1 to n, and

$$C_{jk} = (-1)^{j+k} M_{jk} \tag{AII.26}$$

where M_{jk} is the determinant of the $(n\text{-}1)\times(n\text{-}1)$ submatrix of A, formed by deleting the j^{th} row and k^{th} column of A. Thus for a 2×2 matrix

$$\det A = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

Four important properties of determinants that play important parts either in this Appendix or in other Chapters are:

1. Interchange of two rows of a determinant reverses the sign of the determinant.
2. Multiplication of a row of a determinant by a constant c multiplies the value of the determinant by c.
3. The value of the determinant is zero if any of the rows or columns are proportional (or identical) to each other.
4. The determinants of similar matrices are equal.

As proof of statement 1, consider the case $n = 2$, then clearly (1) holds since

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

and

$$\begin{vmatrix} c & d \\ a & b \end{vmatrix} = bc - ad$$

For the order n matrix A, let B be obtained from A via the interchange of two rows, then by Eq. (AII.24) expanding about row j, j not one of the interchanged rows

$$\det A = \sum_{k=1}^{n} (-1)^{j+k} a_{jk} M_{jk}$$

$$\det B = \sum_{k=1}^{n} (-1)^{j+k} a_{jk} N_{jk}$$

The determinant N_{jk} is of course obtained from the determinant M_{jk} by the interchange of the same two rows which converted A to B. These two determinants are of order n-1. Then by induction, if (1) holds for the n-1 case it must hold for determinants of order n. So clearly N_{jk} = - M_{jk} and det A = - det B, and statement 1 is proven for determinants of any order.

For statement 2, if B is obtained from A via the multiplication of row j of A by a constant c, then expanding about this row

$$\begin{aligned} \det B &= \sum_{k=1}^{n} (-1)^{j+k} b_{jk} M_{jk} \\ &= \sum_{k=1}^{n} (-1)^{j+k} c a_{jk} M_{jk} \\ &= c \sum_{k=1}^{n} (-1)^{j+k} a_{jk} M_{jk} = c \det A \end{aligned}$$

and statement 2 is proved.

Considering statement 3, if in matrix A, row j is equal to c times row i, then by (2) det A = $c$det B where for matrix B, row j is identical to row i. Interchanging these two rows will leave det B unchanged, but by (1) det B will change sign, implying that det B = - det B i.e. det B = 0 and therefore det A is also zero, and statement 3 is proved.

To prove statement 4, for similar matrices A and B, $B = P^{-1}AP$ then

$$\begin{aligned} \det B &= \det(P^{-1}AP) \\ &= \det P^{-1} \det A \ \det P \\ &= \det A \end{aligned} \qquad \text{(AII.27)}$$

since det (AB) = det (BA) = det A det B (see Kreyszig [50]). Therefore the determinants of similar matrices are equal.

Returning to the main discussion and applying the above aside to the case at hand, from Eq.'s

(AII.24) and (AII.26) the determinant of a diagonal matrix must simply be the product of its diagonal elements, and therefore a product of its eigenvalues. So if a general matrix can be diagonalised by a similarity transform i.e. if it has a spanning set of eigenvectors, then by Eq. (AII.27) its determinant and that of its diagonalised form must be the same, and since its eigenvalues must remain unchanged then

$$\det A = \lambda_1 \lambda_2 \lambda_3 \ldots \lambda_n \tag{AII.28}$$

for a general, diagonalisable $n \times n$ matrix A[iii].

Furthermore, if A is upper triangular Eq.'s (AII.24) and (AII.26) clearly indicate that its determinant is again purely dependant on the product of its diagonal elements. Then since

$$(A - \lambda_i I)x_i = 0 \tag{AII.29}$$

it follows that

$$\begin{aligned} \det(A - \lambda_i I) &= \begin{vmatrix} a_{11} - \lambda_i & a_{12} & \cdots & a_{1n} \\ & a_{22} - \lambda_i & \cdots & a_{2n} \\ & & \ddots & \vdots \\ 0 & & & a_{nn} - \lambda_i \end{vmatrix} \\ &= \prod_{j=1}^{n} (a_{jj} - \lambda_i) \\ &= 0 \end{aligned} \tag{AII.30}$$

implying that A's diagonal elements are again its eigenvalues.

Ergo, the QR factorisation technique, as described above, will, when applied to a real tridiagonal matrix, transform it such that it becomes an upper triangular matrix and therefore will have its eigenvalues lying along its main diagonal. Further details of the QR method and strategies to accelerate its converge can be found in Wilkinson [124] and Stewart [125].

For a numerical example, the QR-factorisation method, as described above, applied to the matrix

$$\begin{bmatrix} 6 & -\sqrt{18} & 0 & 0 \\ -\sqrt{18} & 7 & \sqrt{2} & 0 \\ 0 & \sqrt{2} & 6 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

[iii] This in fact true for any matrix A – see Kryszig [50] and Byron and Fuller [69]

ascertains its eigenvalues to within $\sim 1\times10^{-5}$ % of their exact values, namely 11, 6, 3 and 2, after around 10 iterations.

Having extracted the eigenvalues of a matrix, the next stage is to determine its eigenvectors. A particularly useful technique is inverse iteration. Essentially, from a non-zero starting vector $x^{(0)}$, inverse iteration, or the inverse power method as it is sometimes known, generates a sequence of gradually improving estimates through solving

$$(A-\sigma I)x^{(k+1)} = x^{(k)} \qquad k = 0,1,2\ldots \tag{AII.31}$$

for $x^{(k+1)}$, where σ represents an approximation of the eigenvalue corresponding to the eigenvector being determined. Practically, $x^{(k+1)}$ is normalised after each iterative step so that actually $x^{(k)}x^{(k)}=1$ in Eq. (AII.31). Expanding $x^{(k+1)}$ and $x^{(k)}$ as linear combinations of A's exact eigenvectors, y_i, $Ay_i = \lambda_i y_i$ so that

$$\begin{aligned} x^{(k+1)} &= \sum_i \alpha_i y_i \\ x^{(k)} &= \sum_i \beta_i y_i \end{aligned} \tag{AII.32}$$

then

$$\sum_i \alpha_i(\lambda_i-\sigma)y_i = \sum_i \beta_i y_i$$

therefore

$$\alpha_i = \frac{\beta_i}{\lambda_i-\sigma}$$

and

$$x^{(k+1)} = \sum_i \frac{\beta_i y_i}{\lambda_i-\sigma} \tag{AII.33}$$

This means that if σ is close to an eigenvalue λ_n, then $x^{(k+1)}$ will be richer in y_n than in any other eigenvector, and, provided β_n is not too small, will approximate y_n well up to a normalisation factor.

To enhance the estimate of the eigenvalue for iterative step k, consider the action of the exact eigenvector in Eq. (AII.31):

$$(A-\sigma_k I)y_n = (\lambda_n-\sigma_k)y_n \tag{AII.34}$$

now, substitute the improved vector $x^{(k+1)}$ for y_n and let λ_n be the improved eigenvalue estimate, σ_{k+1}, then the LHS of Eq. (AII.34) is equal to $x^{(k)}$ and so

$$\sigma_{k+1} = \sigma_k + \frac{1}{x^{(k)} x^{(k+1)}} \tag{AII.35}$$

using the $x^{(k)}x^{(k)} = 1$ normalisation relation [52].

To solve Eq. (AII.31) for $x^{(k+1)}$, a selection of techniques can be used, including the iterative Gauss-Seidel method. For an $n \times n$ matrix system $Ax = b$, it returns the vector x and is particularly competitive computationally, like with all iterative methods, if the matrix A is large and sparse.

For this first stage of the Gauss-Seidel method, rearrange the system of equations so that no diagonal coefficient is zero, then equate the elements a_{jj} to 1, j =1 to n, by diving through each equation by the current value of a_{jj}, then factor

$$A = I + L + U \tag{AII.36}$$

where L and U are lower and upper tridiagonal matrices respectively with their main diagonals set to zero. Since $Ax = b$ and $Ix = x$ then

$$x = b - Lx - Ux \tag{AII.37}$$

As this is an iterative procedure, the accuracy of $x^{(k+1)}$ will in general be greater than that of $x^{(k)}$. This improved data of the advancing cycle can, at times, be used to enhance the accuracy of the current cycle. For instance, during iteration k after working with row 1 of the matrix, there exists an improved x_1^{k+1} which can then be used in the remainder of the k cycle and so forth. Explicitly,

$$x^{(k+1)} = b - Lx^{(k+1)} - Ux^{(k)} \tag{AII.38}$$

The convergence of any iterative sequence is dependant on the connection between $x^{(k)}$ and $x^{(k+1)}$. Rearranging Eq. (AII.38) so that

$$(I + L)x^{(k+1)} = b - Ux^{(k)}$$

then multiplying by $(I+L)^{-1}$ on the left yields

$$x^{(k+1)} = Cx^{(k)} + (I + L)^{-1}b \tag{AII.39}$$

where $C = -(I+L)^{-1}U$, and is referred to as the iteration matrix. If the spectral radius of C (denoted as $\rho(C)$ and equal to the highest magnitude eigenvalue of C, $\max|\zeta|$) is less than one, then the sequence converges: as [iv, v]

[iv] For each eigenvalue ζ_s of C there is a corresponding eigenvector $\mathbf{u}_s$ which can be normalised to any vector norm, thus $\|\mathbf{u}_s\| = 1$. Therefore $\|C\| \geq \|C\mathbf{u}_s\| = \|\zeta_s\mathbf{u}_s\| = |\zeta_s|$. This relationship holds for all the eigenvalues of C, including the largest, hence Eq. (AII.40) [126].

[v] A vector norm for a vector $\mathbf{x}$ is denoted by $\|\mathbf{x}\|$. Common norms are the l_1-norm:

$$\rho(C) \leq \|C\| \tag{AII.40}$$

then it is possible to find an infinitesimally small quantity $\varepsilon > 0$ such that $\|C\| \leq \rho(C)+\varepsilon < 1$ with $\rho(C) < 1$. Given that $\|AB\| \leq \|A\|\,\|B\|$, then $\|C^k\| \leq \|C\|^k < 1$ so

$$\lim_{k\to\infty} \| C^k \| = 0 \tag{AII.41}$$

and Eq. (AII.39) will tend to some fixed value i.e. it will converge. Since

$$\lim_{k\to\infty} \| C \|^k \geq \lim_{k\to\infty} \| C^k \|$$

a condition that is used practically as sufficient for convergence is

$$\|C\| < 1 \tag{AII.42}$$

The spectral radius of C can be used to accelerate the convergence of the Gauss-Seidel technique. From Eq. (AII.38)

$$x^{(k+1)} = x^{(k)} + b - Lx^{(k+1)} - (U + I)x^{(k)}$$

then the addition of a factor $\omega > 1$ creates a Successive Over-Relaxation (SOR) formula for the Gauss-Seidel method

$$x^{(k+1)} = x^{(k)} + \omega\,(b - Lx^{(k+1)} - (U + I)x^{(k)}) \tag{AII.43}$$

Kreyzig [50] recommends the value

$$\omega = \frac{2}{1+\sqrt{1-\rho(C)}} \tag{AII.44}$$

$$\| \mathbf{x} \|_1 = | x_1 | + | x_2 | + \ldots + | x_n |$$

the Euclidian or l_2-norm:

$$\| \mathbf{x} \|_2 = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}$$

and the l_∞-norm

$$\| \mathbf{x} \|_\infty = \max_j | x_j |$$

Similarly, $\|A\|$ represents the norm of a square matrix, and again, there are different choices of norm. Commonly there is the Frobenius norm

$$\| A \| = \sqrt{\sum_i \sum_j a_{ij}^2}$$

the column sum norm

$$\| A \| = \max_j \sum_i | a_{ij} |$$

i.e. take the sum of $|a_{ij}|$ in column j, j = 1,2...n, and then take the largest of these sums, and the row sum norm

$$\| A \| = \max_i \sum_j | a_{ij} |$$

i.e. take the sum of $|a_{ij}|$ in row i, i = 1,2...n, and then take the largest of these sums

As a numerical example, these two iterative methods, GS and SOR, applied to the matrix-vector system

$$\begin{bmatrix} 1 & -1/4 & -1/4 & 0 \\ -1/4 & 1 & 0 & -1/4 \\ -1/4 & 0 & 1 & -1/4 \\ 0 & -1/4 & -1/4 & 1 \end{bmatrix} x = \begin{bmatrix} 50 \\ 50 \\ 25 \\ 25 \end{bmatrix}$$

where the exact value for x is

$$x = \begin{bmatrix} 87.5 \\ 87.5 \\ 62.5 \\ 62.5 \end{bmatrix}$$

generate vectors accurate to $\sim 5\times10^{-7}$ % and $\sim 1\times10^{-13}$ % respectively, after only 15 iterations.

Utilising the SOR method within the inverse iteration technique, the eigenvectors of the matrix

$$\begin{bmatrix} -2 & 2 & -3 \\ 2 & 1 & -6 \\ -1 & -2 & 0 \end{bmatrix}$$

can be found, agreeing with exact eigenvalues

$$x_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} \quad x_2 = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \quad x_3 = \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}$$

to similar accuracy.

AII.3 Derivatives of Analytic Functions

If $f(z)$ is analytic on and within a closed contour C, its value at each point within C is determined by its values on the bounding curve C. This is known as Cauchy's integral formula and can be expressed as [69]

$$f(z_0) = \frac{1}{2\pi i} \oint_C \frac{f(z)}{z - z_0} dz \qquad \text{(AII.45)}$$

where z_0 is any point within C.

Given Eq. (AII.45), then by the definition of the first derivative

$$f'(z_0) = \lim_{z_1 \to z_0} \frac{f(z_1) - f(z_1)}{z_1 - z_0}$$

$$= \frac{1}{2\pi i} \lim_{z_1 \to z_0} \oint_C \left[\frac{f(z)}{z - z_1} - \frac{f(z)}{z - z_0} \right] \frac{dz}{z_1 - z_0}$$

$$= \frac{1}{2\pi i} \lim_{z_1 \to z_0} \oint_C \frac{f(z)}{(z - z_1)(z - z_0)} dz$$

therefore

$$f'(z_0) - \frac{1}{2\pi i} \oint_C \frac{f(z)}{(z - z_0)^2} dz = \frac{1}{2\pi i} \lim_{z_1 \to z_0} \oint_C f(z) \left[\frac{1}{(z - z_1)(z - z_0)} - \frac{1}{(z - z_0)^2} \right] dz$$

$$= \frac{1}{2\pi i} \lim_{z_1 \to z_0} (z_1 - z_0) \oint_C \frac{f(z)}{(z - z_1)(z - z_0)^2} dz \qquad \text{(AII.46)}$$

From the bounds of absolute magnitude

$$\left| \frac{1}{2\pi i} \lim_{z_1 \to z_0} (z_1 - z_0) \oint_C \frac{f(z)}{(z - z_1)(z - z_0)^2} dz \right| \le \frac{1}{2\pi} \lim_{\varepsilon \to 0} \varepsilon \oint_C \frac{|f(z)||dz|}{|(z - z_0) - \varepsilon e^{i\theta}||z - z_0|^2}$$

where $\varepsilon e^{i\theta}$ has been set equal to $z_1 - z$. Using the *ML*-inequality[vi] [50] and replacing $|z - z_0|$ by its minimum value, μ, and $|f(z)|$ by its maximum value, M, then

$$\frac{1}{2\pi} \lim_{\varepsilon \to 0} \varepsilon \oint_C \frac{|f(z)||dz|}{|(z - z_0) - \varepsilon e^{i\theta}||z - z_0|^2} \le \frac{1}{2\pi} \frac{ML}{\mu^2} \lim_{\varepsilon \to 0} \frac{\varepsilon}{\mu - \varepsilon} = 0$$

and thus the LHS of Eq. (AII.46) is zero and so

$$f'(z_0) = \frac{1}{2\pi i} \oint_C \frac{f(z)}{(z - z_0)^2} dz$$

Indeed, for the nth derivative

$$f^{(n)}(z_0) = \frac{n!}{2\pi i} \oint_C \frac{f(z)}{(z - z_0)^{n+1}} dz \qquad \text{(AII.47)}$$

The requirement that $f(z)$ be analytic implies the existence of all other derivatives, the n^{th} derivative being continuous within C as a consequence of the existence of the $(n+1)^{th}$ derivative. A function is deemed analytic in a domain if it possesses a continuous derivative everywhere within that domain, and so all the derivatives of $f(z)$ are also analytic.

[vi] *ML*-inequality:

$$\left| \oint_C f(z) dz \right| \le ML$$

where M is a constant such that for all values of $|f(z)|$ on C, $|f(z)| \le M$, and L is the length of contour C.

Appendix III

AIII.1 Introduction to the Calculus of Variations

An important problem of calculus is the determination of the stationary values of a function f. For a function of several variables, $f(x_1, x_2, \ldots x_n)$, this requires that

$$\frac{\partial f}{\partial x_i} = 0 \qquad i = 1,2,\ldots,n \tag{AIII.1}$$

generally a task of no particular difficulty. However, suppose the problem is modified so that the stationary value of a function f must be found, but now subject to the constraint that a second function $g(x_1, x_2, \ldots x_n)$ must be constant. This is not straight forward: the partial derivative of f with respect to x_i taken in Eq. (AIII.1) requires that all other variables, x_j where $j \neq i$, remain constant, however, the existence of a constraint means that it is not possible to vary one variable without also altering at least one of the other independent variables.

The differential of f

$$df = \sum_i^n \frac{\partial f}{\partial x_i} dx_i \tag{AIII.2}$$

must vanish for all small displacements from the stationary point. In the absence of a constraint, dx_i can be selected independently, and so it is possible to choose $dx_i \neq 0$, $dx_j = 0$ then $\partial f/\partial x_i = 0$ and so forth. Introducing the constraint, then

$$dg = \sum_i^n \frac{\partial g}{\partial x_i} dx_i \tag{AIII.3}$$

and must always equal zero. Adding Eq. (AIII.2) and a multiple of Eq. (AIII.3) yields

$$df + \lambda dg = \sum_i^N \left(\frac{\partial f}{\partial x_i} + \lambda \frac{\partial f}{\partial x_i} \right) dx_i = 0 \tag{AIII.4}$$

which to be true for all choices of dx_i, implies that

$$\frac{\partial f}{\partial x_i} + \lambda \frac{\partial g}{\partial x_i} = 0 \qquad i = 1,2,\ldots,n \tag{AIII.5}$$

This set of equations is identical to those generated in finding the stationary values of the function h, where $h = f + \lambda g$, without any constraints. λ is referred to as a Lagrange undetermined multiplier. If a set of constraints exist, g_k, then Eq. (AIII.5) becomes

$$\frac{\partial f}{\partial x_i} + \sum_k \lambda_k \frac{\partial g_k}{\partial x_i} = 0 \qquad i = 1,2,\ldots,n \tag{AIII.6}$$

introducing a separate Lagrange multiplier, λ_k, for each constraint.

What if this were to be extended, so that it is the form of the function f, or even functions f_i, that are required such that a second quantity, itself a function of the f_i's (known as a functional), takes a stationary value?

Classically, this is the root of an alternative, but equivalent, formulisation of Newton's equations of motion, known as the action principle. In essence, the foundation of this alternative methodology is the evaluation of all possible paths between two points in time and the selection of the one with the minimum 'action'. This then is the route taken by the system and called the classical or Newtonian path [104].

Beginning with a simple scenario in the absence of any constraints; consider a classical mechanical system specified by the coordinates $q_i(t)$, i = 1 to n with a potential $V(q_1, q_2, \ldots, q_n, t)$. Its motion will be determined by its Lagrangian, L, defined (non-relativistically) as the difference between the kinetic and potential energy of a system i.e.

$$L(t, q_i, \dot{q}_i) = T(q_i, \dot{q}_i) - V(q_i, t) \tag{AIII.7}$$

the dot denoting differentiation with respect to time [69].

The functional of the system known as the action is defined as the integral of the Lagrangian between time t_1 and t_2

$$S = \int_{t_1}^{t_2} L dt \tag{AIII.8}$$

and the Newtonian path is then the motion of the system which minimises the action,

$$\delta S = 0 \tag{AIII.9}$$

with respect to the functions $q_i(t)$, i = 1 to n. The requirement that the action be minimised is alternatively known as Hamilton's principle.[i]

[i] To be exact of course, the condition $\delta S = 0$ is only necessary for a minimum. The solutions q_i could also generate a point of inflection or a maximum in the functional S. Only on the physical interpretation of the situation is the extremum of S interpreted as a minimum.

To calculate δS, deform the path by $\delta q_i(t)$, see Figure A3-1,

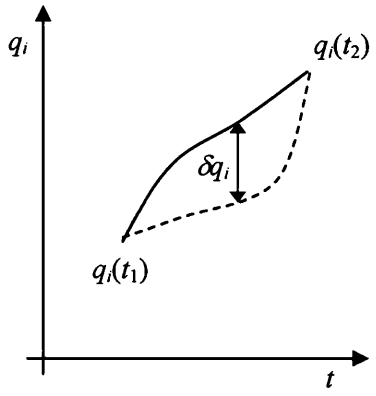


Figure AIII-1 Representation of varied path of $q_i(t)$ through the fixed points $q_i(t_1)$ and $q_i(t_2)$

keeping both $q_i(t_1)$ and $q_i(t_2)$ fixed i.e.

$$\delta q_i(t_1) = \delta q_i(t_1) = 0 \tag{AIII.10}$$

then

$$\delta S = \sum_i \int_{t_1}^{t_2} dt \left(\frac{\partial L}{\partial q_i} \delta q_i + \frac{\partial L}{\partial \dot{q}_i} \delta \dot{q}_i \right) = 0 \tag{AIII.11}$$

varying the Lagrangian with respect to both changes in position and velocity. Integrating the second term of the integral by parts yields

$$\delta S = \sum_i \int_{t_1}^{t_2} dt \left[\left(\frac{\partial L}{\partial q_i} + \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} \right) \delta q_i + \frac{d}{dt} \left(\delta q_i \frac{\partial L}{\partial \dot{q}_i} \right) \right] = 0 \tag{AIII.12}$$

In addition, from Eq. (AIII.10), the last term can be seen to equal zero, and so in general, the action is minimised if

$$\frac{\partial L}{\partial q_i} + \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} = 0 \qquad i = 1,2,...,n \tag{AIII.13}$$

These equations are known as the Euler-Lagrange equations, and their solutions give the form of q_i for which the action is a minimum. For instance, for an arbitrary potential $V(x, y, z)$ with

$$T = \frac{1}{2} m \left(\dot{x}^2 + \dot{y}^2 + \dot{z}^2 \right)$$

then

$$L = \frac{1}{2} m \left(\dot{x}^2 + \dot{y}^2 + \dot{z}^2 \right) - V(x, y, z)$$

with the Euler-Lagrange equations

$$m\ddot{x} = -\frac{\partial V}{\partial x}$$
$$m\ddot{y} = -\frac{\partial V}{\partial y}$$
$$m\ddot{z} = -\frac{\partial V}{\partial z}$$

As $-\partial V/\partial x$ is simply the force on the particle in the x-direction, vectorially the three equations can be written as

$$m\ddot{\mathbf{r}} = \mathbf{F}$$

Quantum mechanically of course, the minimum action is only the most likely path of a system, there being a finite probability of alternate behaviour. The action has now been promoted to the fundamental object of the system, and indeed, a generalised version of the Euler-Lagrange equations and Hamilton's principle for an infinite number of degrees of freedom is the corner stone of field theory.

To move beyond the simple classical scenario discussed above, consider that although Eq. (AIII.13) minimises an integrand with n variables, $q_1(t)$, $q_2(t)\ldots$ $q_n(t)$, they are dependent variables as they all depend on the independent variable t. To add more independent variables, introduce the Lagrangian density, $\mathscr{L}$, whose spatial integral yields the Lagrangian. In one spatial dimension (in addition to time) and for one dependent variable $y(x, t)$, the action principle means that the function y that describes the motion of the system is the one which causes

$$S = \int_{t_1}^{t_2} L\,dt = \int_{t_1}^{t_2}\int_{x_1}^{x_2} \mathscr{L}\,dx\,dt \qquad \text{(AIII.14)}$$

to be an extremum with respect to those functions $y(x, t)$ which satisfy the given values of y on the boundary of the domain D. D is defined on the two dimensional xt-plane by the rectangle $x_1 \le x \le x_2$, $t_1 \le t \le t_2$. The Euler-Lagrange equation for this functional is then

$$\frac{\partial \mathscr{L}}{\partial y} + \frac{\partial}{\partial x}\frac{\partial \mathscr{L}}{\partial\left(\frac{\partial y}{\partial x}\right)} + \frac{\partial}{\partial t}\frac{\partial \mathscr{L}}{\partial\left(\frac{\partial y}{\partial t}\right)} = 0 \qquad \text{(AIII.15)}$$

Time has now lost any particular significance as an independent variable within the variational framework, and the action principle and the Euler-Lagrange equation have been generalised for a Lagrangian density simply of two arbitrary independent variables.

Generalising this to a system of n independent variables x_i i.e. extending the space to n-dimensions, and including m dependent variables, y_j, then the relevant Lagrangian density is of the form

$$\mathscr{L} = \mathscr{L}\left[x_1, x_2, \cdots, x_n, y_1, y_2, \cdots, y_m, \frac{\partial y_1}{\partial x_1}, \frac{\partial y_1}{\partial x_2}, \cdots, \frac{\partial y_1}{\partial x_n}, \cdots, \frac{\partial y_m}{\partial x_1}, \frac{\partial y_m}{\partial x_2}, \ldots, \frac{\partial y_m}{\partial x_n}\right] \qquad \text{(AIII.16)}$$

where $y_j = y_j(x_1, x_2, \ldots, x_n)$. For the action

$$S = \int_D \mathscr{L}\left(x_i, y_j, \frac{\partial y_j}{\partial x_i}\right) dx_1 dx_2 \cdots dx_n \qquad \text{(AIII.17)}$$

to be minimised, $\delta S = 0$, the set of Euler-Lagrange equations which give the extremising functions y_j are

$$\frac{\partial \mathscr{L}}{\partial y_j} + \sum_i^n \frac{\partial}{\partial x_i} \frac{\partial \mathscr{L}}{\partial\left(\frac{\partial y_j}{\partial x_i}\right)} = 0 \qquad j = 1,2,\ldots,m \qquad \text{(AIII.18)}$$

It is now a convenient place to reintroduce the concept of constraints. Analogous to finding the stationary value of a function $f(x_1, x_2, \ldots x_n)$, satisfying the set of constraints $g_k(x_1, x_2, \ldots x_n)$, the problem here is to determine the stationary value of the multiple integral

$$S = \int_D \mathscr{L}\left(x_i, y_j, \frac{\partial y_j}{\partial x_i}\right) dx_1 dx_2 \cdots dx_n$$

satisfying the integral constraints

$$J_k = \int_D g_k\left(x_i, y_j, \frac{\partial y_j}{\partial x_i}\right) dx_1 dx_2 \cdots dx_n \qquad \text{(AIII.19)}$$

where J_k is a constant, for the general case of n independent variables (x_i), m dependent variables (y_i), and p integral constraints (J_k). As in the case of the simple function f, form the new function

$$h = \mathscr{L} + \sum_{k=1}^{p} \lambda_k g_k \qquad \text{(AIII.20)}$$

where λ_k are p constant Lagrange multipliers, then the task of extremising S given all J_k, is identical to applying the Euler-Lagrange equations to h

$$\frac{\partial h}{\partial y_j} + \sum_i^n \frac{\partial}{\partial x_i} \frac{\partial h}{\partial\left(\frac{\partial y_j}{\partial x_i}\right)} = 0 \qquad j = 1,2,\ldots,m \qquad \text{(AIII.21)}$$

As an example, regard the Lagrangian density

$$\mathscr{L} = \frac{1}{2}\left(\frac{\partial \psi^*}{\partial x}\frac{\partial \psi}{\partial x} + \frac{\partial \psi^*}{\partial y}\frac{\partial \psi}{\partial y} + \frac{\partial \psi^*}{\partial z}\frac{\partial \psi}{\partial z} \right) + V(x,y,z)\psi^*\psi \qquad \text{(AIII.22)}$$

with three independent variables (x, y and z) and two dependent variables (the real, ψ_1, and imaginary, ψ_2, parts of the complex wave function ψ (= $\psi(x, y, z)$)) and where V is chosen to be real.

To minimise

$$S = \int_D \mathscr{L} dx dy dz \qquad \text{(AIII.23)}$$

subject to the normalisation constraint

$$J = \int_D \psi^*\psi dx dy dz = 1 \qquad \text{(AIII.24)}$$

form the function h,

$$\begin{aligned} h &= \mathscr{L} + \lambda\psi^*\psi \\ &= \frac{1}{2}\sum_{i=1}^{3}\left[\left[\frac{\partial \psi_1}{\partial x_i} \right]^2 + \left[\frac{\partial \psi_2}{\partial x_i} \right]^2 \right] + (V+\lambda)(\psi_1^2 + \psi_2^2) \end{aligned} \qquad \text{(AIII.25)}$$

re-labelling x, y, z as x_1, x_2 and x_3 for convenience, and the two Euler-Lagrange equations

$$\lambda\psi_j + V\psi_j - \frac{1}{2}\sum_{i=1}^{3}\frac{\partial}{\partial x_i}\left(\frac{\partial \psi_j}{\partial x_i} \right) = 0 \qquad j = 1,2 \qquad \text{(AIII.26)}$$

follow. These equations can be written in the more familiar form of the time-independent Schrödinger equation

$$-\frac{1}{2}\nabla^2\psi + V\psi = E\psi \qquad \text{(AIII.27)}$$

should the Lagrange multiplier be identified with the negative of the energy eigenvalue, $\lambda = -E$ [65].

Indeed, this rather neatly leads on to the perhaps commoner form of the variational principle in quantum mechanics, where, instead of working with the Lagrangian formulation and its independent variables of 'position' (y_j) and 'velocity' ($\partial y_j / \partial x_i$), it is often more useful to work in the alternative and comparable Hamiltonian formulation and its independent variables of 'position' (y_j) and 'momentum' ($\partial L / \partial(\partial y_j / \partial x_i)$).

In this way, Schrödinger's equation Eq. (AIII.27), or $(H - E)\psi = 0$, follows from the equivalent route of seeking the function ψ such that the expectation value of the Hamiltonian, $\langle\psi|H|\psi\rangle$ in the Dirac notation of Chapter 4, is minimised i.e. $\delta\langle\psi|H|\psi\rangle = 0$, subject to the normalisation constraint $\langle\psi|\psi\rangle = 1$ [70]. It is therefore required that

$$\delta\left[\langle\psi|H|\psi\rangle + \lambda\langle\psi|\psi\rangle\right] = 0$$

where λ is the unusual lagrange multiplier. Given that

$$\langle\psi|H|\psi\rangle + \lambda\langle\psi|\psi\rangle = \int_D \left(\psi^* H\psi + \lambda\psi^*\psi\right) dxdydz = \int_D h\, dxdydz$$

and

$$\begin{aligned}\psi^* H\psi &= -\frac{\psi^*\nabla^2\psi}{2} + V\psi^*\psi \\ &= \frac{1}{2}\nabla\psi^*\cdot\nabla\psi + V\psi^*\psi \\ &= \frac{1}{2}\sum_{i=1}^{3}\frac{\partial\psi^*}{\partial x_i}\frac{\partial\psi}{\partial x_i} + V\psi^*\psi\end{aligned}$$

then proceeding as before, it can be seen that h is the same as in the Lagrangian formulism, see Eq. (AIII.25), and as such, the Euler-Lagrange equations applied to h are given by Eq. (AIII.26), and identifying $\lambda = -E$ then once again $(H - E)\psi = 0$ results.

AIII.2 Introduction to Green's Functions

Green's functions perform a pivotal role in many branches of modern quantum physics. An extensive topic, linking differential and integral equations, Green's function theory cannot be comprehensively be covered in one short Appendix and for further detail introducing this rather vast subject, the interested reader is referred in particular to Economou [105] and Byron and Fuller [69].

AIII.2.1 Time-Independent Green's Functions

The Green's function of the linear, Hermitian, time-independent, differential operator L is defined as the solution of the differential equation

$$(\lambda - L)G(\mathbf{r},\mathbf{r}';\lambda) = -\delta(\mathbf{r} - \mathbf{r}') \qquad \text{(AIII.28)}$$

subject to certain boundary conditions. The operator L is taken to possess a complete set of orthonormal eigenfunctions, $\{\varphi_n\}$, such that

$$(\lambda_n - L)\varphi_n(\mathbf{r}) = 0 \tag{AIII.29}$$

where λ_n are the corresponding eigenvalues of the eigenfunctions φ_n. $\{\varphi_n\}$ satisfy the same boundary conditions as G.

Once G is known, then it can be used to solve equations of the form

$$(\lambda - L)\psi(\mathbf{r}) = -f(\mathbf{r}) \tag{AIII.30}$$

generating the particular solution with the integral(s)

$$\psi(\mathbf{r}) = \int G(\mathbf{r}, \mathbf{r}'; \lambda) f(\mathbf{r}') d\mathbf{r}' \qquad \lambda \neq \{\lambda_n\} \tag{AIII.31}$$

$$\psi(\mathbf{r}) = \varphi_i(\mathbf{r}) + \int G(\mathbf{r}, \mathbf{r}'; \lambda_i) f(\mathbf{r}') d\mathbf{r}' \qquad \lambda_i = \{\lambda_n\} \tag{AIII.32}$$

The proof of this is straightforward. For the general inhomogeneous differential equation given in Eq. (AIII.30), expand ψ and f in terms of the complete set of homogeneous eigenfunctions, such that

$$\psi(\mathbf{r}) = \sum_{n=0}^{\infty} \alpha_n \varphi_n(\mathbf{r}) \tag{AIII.33}$$

$$f(\mathbf{r}) = \sum_{n=0}^{\infty} \beta_n \varphi_n(\mathbf{r}) \tag{AIII.34}$$

and similarly for the Green's function G:

$$G(\mathbf{r}, \mathbf{r}'; \lambda) = \sum_{n=0}^{\infty} \varphi_n(\mathbf{r}) a_n^{\lambda}(\mathbf{r}') \tag{AIII.35}$$

Substitution of Eq. (AIII.35) into Eq. (AIII.28), with the expansion of the delta function as

$$\delta(\mathbf{r} - \mathbf{r}') = \sum_{n=0}^{\infty} \varphi_n(\mathbf{r}) \varphi_n^*(\mathbf{r}') \tag{AIII.36}$$

yields

$$\lambda \sum_{n=0}^{\infty} \varphi_n(\mathbf{r}) a_n^{\lambda}(\mathbf{r}') - \sum_{n=0}^{\infty} \lambda_n \varphi_n(\mathbf{r}) a_n^{\lambda}(\mathbf{r}') = -\sum_{n=0}^{\infty} \varphi_n(\mathbf{r}) \varphi_n^*(\mathbf{r}') \tag{AIII.37}$$

and substituting both Eq. (AIII.33) and Eq. (AIII.34) into Eq. (AIII.30) results in

$$\sum_{n=1}^{\infty} (\lambda \alpha_n - \lambda_n \alpha_n + \beta_n) \varphi_n = 0$$

which, from the linear independence of φ_n, implies that

$$\alpha_n (\lambda_n - \lambda) = \beta_n \tag{AIII.38}$$

If λ does not equal any of the discrete eigenvalues λ_n then

$$\psi(\mathbf{r}) = \sum_{n=0}^{\infty} \frac{\varphi_n(\mathbf{r})\beta_n}{\lambda_n - \lambda}$$
$$= \sum_{n=0}^{\infty} \frac{\varphi_n(\mathbf{r})}{\lambda_n - \lambda} \int \varphi_n^*(\mathbf{r}') f(\mathbf{r}') d\mathbf{r}'$$

using φ_n's orthonormality. The orthonormal nature of the eigenfunctions also implies that

$$a_n^{\lambda}(\mathbf{r}')(\lambda_n - \lambda) = \varphi_n^*(\mathbf{r}') \tag{AIII.39}$$

from Eq. (AIII.37). Then on interchange of the summation and integral

$$\psi(\mathbf{r}) = \int \sum_{n=0}^{\infty} \frac{\varphi_n(\mathbf{r})\varphi_n^*(\mathbf{r}')}{\lambda_n - \lambda} f(\mathbf{r}') d\mathbf{r}'$$
$$= \int G(\mathbf{r}, \mathbf{r}'; \lambda) f(\mathbf{r}') d\mathbf{r}' \tag{AIII.40}$$

as required, and where the Green's function satisfying Eq. (AIII.28) is given by

$$G(\mathbf{r}, \mathbf{r}'; \lambda) = \sum_{n=0}^{\infty} \frac{\varphi_n(\mathbf{r})\varphi_n^*(\mathbf{r}')}{\lambda_n - \lambda} \tag{AIII.41}$$

on combining Eq. (AIII.35) and Eq. (AIII.39).

When λ is equal to one of the eigenvalues λ_m, a solution to Eq. (AIII.30) only exists if $\beta_n = 0$, as can be seen from Eq. (AIII.38). For m not equal to n then

$$\psi(\mathbf{r}) = \sum_{m \neq n} \frac{\varphi_m(\mathbf{r})\beta_m}{\lambda_m - \lambda} \tag{AIII.42}$$

is a solution to Eq. (AIII.30). However, given that

$$(\lambda - L)\varphi_n(\mathbf{r}) = 0 \tag{AIII.43}$$

is true, then ψ - φ_n is also a solution of Eq. (AIII.30) since

$$(\lambda - L)[\psi(\mathbf{r}) - \varphi_n(\mathbf{r})] = (\lambda - L)\psi(\mathbf{r}) - (\lambda - L)\varphi_n(\mathbf{r})$$
$$= -f(\mathbf{r}) - 0 \tag{AIII.44}$$

Therefore, in general

$$\psi(\mathbf{r}) = \varphi_n(\mathbf{r}) + \int G(\mathbf{r}, \mathbf{r}'; \lambda_n) f(\mathbf{r}') d\mathbf{r}' \tag{AIII.45}$$

where the Green's function $G(\mathbf{r}, \mathbf{r}'; \lambda_n)$ is given by

$$G(\mathbf{r}, \mathbf{r}'; \lambda_n) = \sum_{m \neq n} \frac{\varphi_m(\mathbf{r})\varphi_m^*(\mathbf{r}')}{\lambda_m - \lambda_n} \tag{AIII.46}$$

However, in this case, although referred to in the literature (see for instance Krieger *et al.* [21] or Engel and Driezler [81]) as a Green's function, $G(\mathbf{r}, \mathbf{r}'; \lambda_n)$ is not in the strictest mathematical

sense, as it does not satisfy Eq. (AIII.28) for all $\{\varphi_n\}$.

Applying the same orthogonality arguments for $m \neq n$ as in the $\lambda \neq \lambda_n$ case, then from Eq. (AIII.37)

$$a_m^{\lambda_n}(\mathbf{r}') = \frac{\varphi_m^*(\mathbf{r}')}{\lambda_m - \lambda_n} \qquad m \neq n$$

and Eq. (AIII.28) is satisfied, but for $m = n$, Eq. (AIII.37) implies that

$$\varphi_n^*(\mathbf{r}') = 0$$

which is inconsistent with the definition of an eigenvector. Indeed, Economou [105] states that

> "...$G(z)$ is uniquely defined if and only if $z \neq \{\lambda_n\}$. If z coincides with any of the discrete eigenvalues of L, G does not exist..."

where $G(z)$ is an abbreviation of $G(\mathbf{r}, \mathbf{r}'; z)$, Economou's z representing the λ of this appendix.

This said, considering the limit where λ in $G(\mathbf{r}, \mathbf{r}'; \lambda)$ tends to λ_n :

$$\begin{aligned}
\lim_{\lambda\to\lambda_n}(\lambda - L)G(\mathbf{r},\mathbf{r}';\lambda) &= \lim_{\lambda\to\lambda_n}(\lambda - L)\sum_{m=0}^{\infty}\frac{\varphi_m(\mathbf{r})\varphi_m^*(\mathbf{r}')}{\lambda_m - \lambda} \\
&= \lim_{\lambda\to\lambda_n}(\lambda - L)\sum_{m\neq n}\frac{\varphi_m(\mathbf{r})\varphi_m^*(\mathbf{r}')}{\lambda_m - \lambda} + \lim_{\lambda\to\lambda_n}(\lambda - \lambda_n)\frac{\varphi_m(\mathbf{r})\varphi_m^*(\mathbf{r}')}{\lambda_n - \lambda} \\
&= -\sum_{m=0}^{\infty}\varphi_m(\mathbf{r})\varphi_m^*(\mathbf{r}') \\
&= -\delta(\mathbf{r} - \mathbf{r}') \qquad \text{(AIII.47)}
\end{aligned}$$

it can be seen that $G(\mathbf{r}, \mathbf{r}'; \lambda_n)$ does however satisfy

$$(\lambda_n - L)G(\mathbf{r},\mathbf{r}';\lambda_n) = -\delta(\mathbf{r} - \mathbf{r}') + \varphi_n(\mathbf{r})\varphi_n^*(\mathbf{r}') \qquad \text{(AIII.48)}$$

Should the operator L have a continuous spectrum of eigenstates, of which λ is a member, then $G(\mathbf{r}, \mathbf{r}'; \lambda)$ can normally be considered to exist, but will not be unique as any general solution of the homogeneous equation can be combined with G to also form a solution. As L is Hermitian its eigenvalues will be real, and so a ready avenue in working with such Green's functions is to extend them into the complex plane, integrating around the branch cut resulting from the continuous eigenstates. See Economou [105] for instance for further details.

To exemplify, regard the problem in Chapter 4 of calculating the perturbations, $\delta\varepsilon$ and $\delta\psi$, of the eigenstates of Eq. (4.46),

$$\left[-\frac{\nabla^2}{2}+v_s(\mathbf{r})\right]\psi_i(\mathbf{r})=\varepsilon_i\psi_i(\mathbf{r})$$

resulting from a small change, δv_s, in the potential v_s.

Explicitly, Eq. (4.46) becomes

$$\left[-\frac{\nabla^2}{2}+v_s(\mathbf{r})-\varepsilon_i\right]\left(\psi_i(\mathbf{r})+\delta\psi_i(\mathbf{r})\right)+\left(\delta v_s(\mathbf{r})-\delta\varepsilon\right)\left(\psi_i(\mathbf{r})+\delta\psi_i(\mathbf{r})\right)=0 \quad \text{(AIII.49)}$$

which, using Eq. (4.46) once again and from

$$\begin{aligned} 0&\approx\delta v_s\delta\psi_i \\ 0&\approx\delta\varepsilon_i\delta\psi_i \end{aligned} \quad \text{(AIII.50)}$$

Eq. (AIII.49) reduces to

$$\left[-\frac{\nabla^2}{2}+v_s(\mathbf{r})-\varepsilon_i\right]\delta\psi_i(\mathbf{r})=\left(\delta\varepsilon-\delta v_s(\mathbf{r})\right)\psi_i(\mathbf{r}) \quad \text{(AIII.51)}$$

multiplying by ψ_i^* and integrating over all space yields

$$\delta\varepsilon_i=\int\psi_i^*(\mathbf{r})\delta v_s(\mathbf{r})\psi_i(\mathbf{r})d\mathbf{r} \quad \text{(AIII.52)}$$

The Green's function of Eq. (AIII.46) is taken as the solution of the equation

$$\left[-\frac{\nabla^2}{2}+v_s(\mathbf{r})-\varepsilon_i\right]G(\mathbf{r},\mathbf{r}';\varepsilon_i)=\delta(\mathbf{r}-\mathbf{r}')-\psi_i(\mathbf{r})\psi_i^*(\mathbf{r}') \quad \text{(AIII.53)}$$

and then from Eq. (AIII.32), the solution of

$$\left[-\frac{\nabla^2}{2}+v_s(\mathbf{r})-\varepsilon_i\right]\left(\psi_i(\mathbf{r})+\delta\psi_i(\mathbf{r})\right)=\left(\delta\varepsilon-\delta v_s(\mathbf{r})\right)\psi_i(\mathbf{r}) \quad \text{(AIII.54)}$$

will be

$$\psi_i(\mathbf{r})+\delta\psi_i(\mathbf{r})=\psi_i(\mathbf{r})+\int G(\mathbf{r},\mathbf{r}';\varepsilon_i)\left(\delta\varepsilon-\delta v_s(\mathbf{r}')\right)\psi_i(\mathbf{r}')d\mathbf{r}' \quad \text{(AIII.55)}$$

This can be written as

$$\delta\psi_i(\mathbf{r})=-\int G(\mathbf{r},\mathbf{r}';\varepsilon_i)\delta v_s(\mathbf{r}')\psi_i(\mathbf{r}')d\mathbf{r}' \quad \text{(AIII.56)}$$

on realising that

$$\int G(\mathbf{r},\mathbf{r}';\varepsilon_i)\psi_i(\mathbf{r}')d\mathbf{r}'=0 \quad \text{(AIII.57)}$$

AIII.2.2 Time-Dependent Green's Function

This section will limit itself to linear partial differential equations of first order in time only, as applicable to equation of the form of the time-dependant Schrödinger equation for example.

Analogous to the time-independent situation, for a partial differential equation of the form

$$-\frac{\partial \psi}{\partial \tau} - L\psi = 0 \tag{AIII.58}$$

where τ is a time variable, the associated Green's function is defined as the solution of the equation

$$-\left(\frac{\partial}{\partial \tau} + L\right) G(\mathbf{r}, \mathbf{r}'; \tau, \tau') = -\delta(\mathbf{r} - \mathbf{r}')\delta(\tau - \tau') \tag{AIII.59}$$

L is a linear, Hermitian, time-independent, differential operator which is assumed to possess a complete set of orthonormal eigenfunctions, $\{\varphi_n\}$, such that

$$(\lambda_n - L)\varphi_n(\mathbf{r}) = 0 \tag{AIII.60}$$

To solve Eq. (AIII.58), proceed as before, and expand ψ in terms of φ_n

$$\psi(\mathbf{r}, \tau) = \sum_{n=0}^{\infty} \alpha_n(\tau)\varphi_n(\mathbf{r}) \tag{AIII.61}$$

the τ dependence of ψ introduced through τ dependent α_n. Substituting Eq. (AIII.61) into Eq. (AIII.58) yields

$$\sum_n \left[\frac{\partial \alpha_n(\tau)}{\partial \tau} + \lambda_n \alpha_n(\tau)\right]\varphi_n(\mathbf{r}) = 0$$

φ_n are linearly independent so

$$\frac{\partial \alpha_n(\tau)}{\partial \tau} + \lambda_n \alpha_n(\tau) = 0 \tag{AIII.62}$$

and the solution of Eq. (AIII.62) is simply

$$\alpha_n(\tau) = e^{-\lambda_n(\tau - \tau')}\alpha_n(\tau') \tag{AIII.63}$$

where $\alpha_n(\tau')$ follows from φ_n's orthogonality:

$$\alpha_n(\tau') = \int \varphi_n^*(\mathbf{r})\psi(\mathbf{r}, \tau')d\mathbf{r}$$

Consequently

$$\psi(\mathbf{r}, \tau) = \sum_{n=0}^{\infty} \varphi_n(\mathbf{r})e^{-\lambda_n(\tau - \tau')} \int \varphi_n^*(\mathbf{r}')\psi(\mathbf{r}', \tau')d\mathbf{r}'$$

which can be written as

$$\psi(\mathbf{r},\tau)=\int G_1(\mathbf{r},\mathbf{r}';\tau,\tau')\psi(\mathbf{r}',\tau')d\mathbf{r}' \tag{AIII.64}$$

where

$$G_1(\mathbf{r},\mathbf{r}';\tau,\tau')=\sum_{n=0}^{\infty}\varphi_n(\mathbf{r})\varphi_n^*(\mathbf{r}')e^{-\lambda_n(\tau-\tau')} \tag{AIII.65}$$

The function G_1 is not a true Green's function, satisfying as it does

$$-\left(\frac{\partial}{\partial\tau}+L\right)G_1(\mathbf{r},\mathbf{r}';\tau,\tau')=0 \tag{AIII.66}$$

but is often referred to as the propagator of the function since it propagates ψ from time τ' to time τ, where $\tau > \tau'$. The subscript 1 on the function G_1 indicates that it pertains to an equation of first order in time.

But what of G proper? Eq. (AIII.65) indicates that as τ approaches τ' then

$$\lim_{\tau\to\tau'} G_1(\mathbf{r},\mathbf{r}';\tau,\tau')=\delta(r-r') \tag{AIII.67}$$

recalling the expansion of the delta function in Eq. (AIII.36). From the definition of G, Eq. (AIII.59), G must satisfy Eq. (AIII.66) everywhere except at the point $\mathbf{r} = \mathbf{r}'$, $\tau = \tau'$. Motivated by this observation, propose that G_1 be combined with the Heaviside function $\theta(\tau\text{-}\tau')$ so that

$$G(\mathbf{r},\mathbf{r}';\tau,\tau)=G_1(\mathbf{r},\mathbf{r}';\tau,\tau)\theta(\tau-\tau') \tag{AIII.68}$$

where the Heaviside function is defined by

$$\theta(\tau-\tau')=\begin{cases}1 & \tau-\tau'>0\\ 0 & \tau-\tau'\le 0\end{cases} \tag{AIII.69}$$

and as such its derivative is a delta function

$$\frac{\partial}{\partial\tau}\theta(\tau-\tau')=\delta(\tau-\tau') \tag{AIII.70}$$

Substituting Eq. (AIII.68) into Eq. (AIII.59) yields

$$\begin{aligned}-\left(\frac{\partial}{\partial\tau}+L\right)G(\mathbf{r},\mathbf{r}';\tau,\tau') &= -\frac{\partial}{\partial\tau}\left[G_1(\mathbf{r},\mathbf{r}';\tau,\tau')\theta(\tau-\tau')\right]-L\left[G_1(\mathbf{r},\mathbf{r}';\tau,\tau')\theta(\tau-\tau')\right]\\ &= -\left(\frac{\partial}{\partial\tau}+L\right)G_1(\mathbf{r},\mathbf{r}';\tau,\tau')\theta(\tau-\tau')-G_1(\mathbf{r},\mathbf{r}';\tau,\tau')\delta(\tau-\tau')\\ &= -G_1(\mathbf{r},\mathbf{r}';\tau,\tau')\delta(\tau-\tau')\end{aligned} \tag{AIII.71}$$

making use of both Eq. (AIII.66) and Eq. (AIII.70).

Since the delta function is equal to zero except at $\tau = \tau'$, G_1 need only be evaluated at $\tau = \tau'$, and

therefore, according to Eq. (AIII.67), Eq. (AIII.71) reduces to

$$-\left(\frac{\partial}{\partial\tau}+L\right)G(\mathbf{r},\mathbf{r}';\tau,\tau')=-\delta(\mathbf{r}-\mathbf{r}')\delta(\tau-\tau') \qquad \text{(AIII.72)}$$

as required.

As G is a true Green's function, then for the partial differential equation

$$-\left(\frac{\partial}{\partial\tau}+L\right)\psi=-F(\mathbf{r},\tau) \qquad \text{(AIII.73)}$$

the wave function ψ follows from

$$\psi(\mathbf{r},\tau)=\psi_0(\mathbf{r},\tau)+\int G(\mathbf{r},\mathbf{r}';\tau,\tau')F(\mathbf{r}',\tau')d\mathbf{r}'d\tau' \qquad \text{(AIII.74)}$$

where ψ_0 is the solution of the homogeneous equation. For the Schrödinger equation for example, the term $F(\mathbf{r}, \tau)$ might represent a perturbation of some kind – see Chapter 7. This chapter also highlights the property of the time-dependent Green's function that its dependence solely on the time difference τ - τ' allows it to be related to the time independent Green's function, discussed in the earlier subsection.

AIII.3 The Addition Theorem

The proof of the addition theorem of spherical harmonics

$$P_n(\cos\gamma)=\frac{4\pi}{2n+1}\sum_{m=-n}^{n}Y_n^m(\theta_1,\phi_1)Y_n^{m*}(\theta_2,\phi_2)$$

is layed out in Arfken and Weber [65] and is the method followed here.

For the coordinate systems depicted in Figure AIII-2, x_2, y_2, z_2 are related to x_1, y_1, z_1 by the rotations θ_2 and ϕ_2, and a given point can be specified by both (θ_1, ϕ_1) and (γ, ψ), in the axes x_1, y_1, z_1 and x_2, y_2, z_2 respectively.

Assuming that a function $f(\theta, \phi)$, to be evaluated over the surface of a sphere, can be expanded in a Laplace series[ii], then

[ii] Laplace series: convergent double series of spherical harmonics, such that a suitable function f, evaluated on a surface of a sphere, can be represented

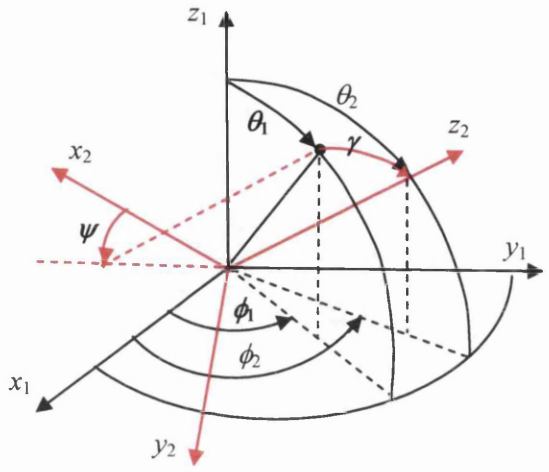


Figure AIII-2 Representation of a point on two different coordinate axes, (θ_1, ϕ_1) denoting its position relative to x_1, y_1, z_1 and (γ, ψ) denoting its position relative to x_2, y_2, z_2. The axes x_2, y_2, z_2 are obtained from x_1, y_1, z_1 by the rotations θ_2 and ϕ_2. Angular coordinates of the point are in bold for clarity.

$$f(\theta_1,\phi_1)=\begin{cases} Y_n^m(\theta_1,\phi_1) & \text{relative to } x_1,y_1,z_1 \\ \sum_{m=-n}^{n} a_{nm}Y_n^m(\gamma,\psi) & \text{relative to } x_2,y_2,z_2 \end{cases} \qquad \text{(AIII.75)}$$

dropping the n summation since the Legendre polynomials are an eigenfunctions of L^2 with eigenvalues $n(n+1)$.

Multiplying by $Y_m^{0*}(\gamma,\psi)$ and integrating, then

$$\int f(\theta_1,\phi_1)Y_n^{0*}(\gamma,\psi)d\Omega_{\gamma\psi} = a_{n0}$$

or

$$\int Y_n^m(\theta_1,\phi_1)Y_n^{0*}(\gamma,\psi)d\Omega_{\gamma\psi} = a_{n0} \qquad \text{(AIII.76)}$$

Now, if the Legendre polynomial $P_n(\cos\gamma)$ has the Laplace expansion

$$P_n(\cos\gamma)=\sum_{m=-n}^{n} b_{nm}Y_n^m(\theta_1,\phi_1) \qquad \text{(AIII.77)}$$

where the dependence on θ_2 and ϕ_2 is built into b_{nm}, then multiplying by the complex conjugate of

---

$$f(\theta,\phi)=\sum_{nm} a_{nm}Y_n^m(\theta,\phi)$$

Y_n^m and integrating over the sphere yields

$$\int P_n(\cos\gamma)Y_n^{m*}(\theta_1,\phi_1)d\Omega_{\theta_1,\phi_1} = b_{nm}$$

which can be written

$$\sqrt{\frac{4\pi}{2n+1}}\int Y_n^0(\gamma,\psi)Y_n^{m*}(\theta_1,\phi_1)d\Omega = b_{nm} \qquad \text{(AIII.78)}$$

The subscripts on the solid angle Ω have been dropped since over the range of integration the choice of polar axis is immaterial.

Equating Eq. (AIII.76) and Eq. (AIII.78) implies that

$$b_{nm}^* = a_{n0}\sqrt{\frac{4\pi}{2n+1}} \qquad \text{(AIII.79)}$$

Then on realising from Eq. (AIII.75) that at $\gamma = 0$, $P_n(1) = 1$ while $P_n^m(1) = 0\, m \neq 0$ then evidently

$$f(\theta_1,\phi_1)\big|_{\gamma=0} = a_{n0}\sqrt{\frac{2n+1}{4\pi}} \qquad \text{(AIII.80)}$$

and so

$$\begin{aligned} b_{nm}^* &= \frac{4\pi}{2n+1} f(\theta_1,\phi_1)\big|_{\gamma=0} \\ &= \frac{4\pi}{2n+1} Y_n^m(\theta_1,\phi_1) \\ &= \frac{4\pi}{2n+1} Y_n^m(\theta_2,\phi_2) \end{aligned} \qquad \text{(AIII.81)}$$

which, on taking its complex conjugate allows Eq. (AIII.77) to be written as

$$P_n(\cos\gamma) = \frac{4\pi}{2n+1}\sum_{m=-n}^{n} Y_n^m(\theta_1,\phi_1)Y_n^{m*}(\theta_2,\phi_2)$$

as required.

Appendix IV

AIV.1 Mutual Capacitance of Two Spheres Using the Image Charge Method

The image charge method is an approach that can be used to calculate capacitance in a situation where the geometry of the problem can be complicated. Essentially, the method seeks to simplify the problem by 'removing' the problematic charge distributions and 'replacing' them with a set of 'image' point charges that duplicate the 'old' potential.

For the approximation of the tunnelling behaviour used in the main body of this work, it is assumed that the capacitance of the two junctions can be adequately described by their mutual capacitance only, and this capacitance is, in general, independent of the voltage applied across the junction. In this way, the voltage across the junction being modelled can be set arbitrarily, and so it is acceptable to place one side of the junction at 0V.

Consider the case of the two conducting spheres in Figure AIV-1.

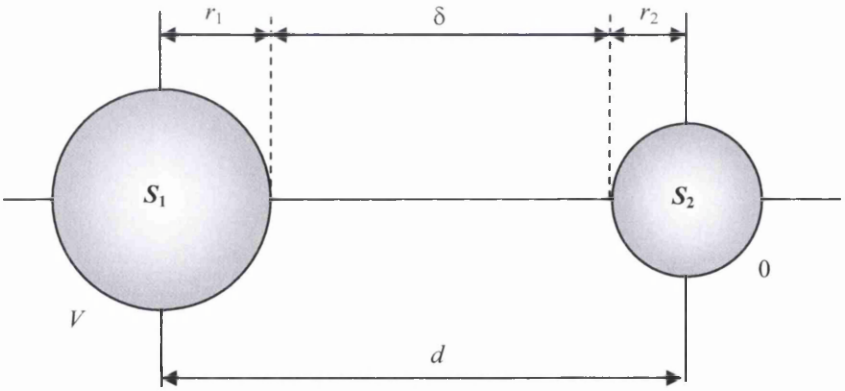


Figure AIV-1 Schematic of two conducting spheres, labelled S_1 and S_2, of radii r_1 and r_2 respectively, situated a distance δ apart. The separation of the centres is d.

The potential at the surface of sphere one, S_1, can be duplicated by a point charge, q_0, at its centre satisfying

$$q_0 = 4\pi V r_1 \varepsilon_1 \qquad \text{(AIV.1)}$$

and the same can be said for sphere two, S_2, except that this charge is of course zero.

However, if the two spheres were removed, then the two point charges cannot simply replace the spheres as matters stand because the charge q_0 will perturb the potential at the former surface of S_2. Another point charge is now needed, say q_1, within the former S_2 and of opposite polarity to q_0 in order to negate the effect of q_0 and restore the potential at the former surface of S_2 to zero, see Figure AIV-2.

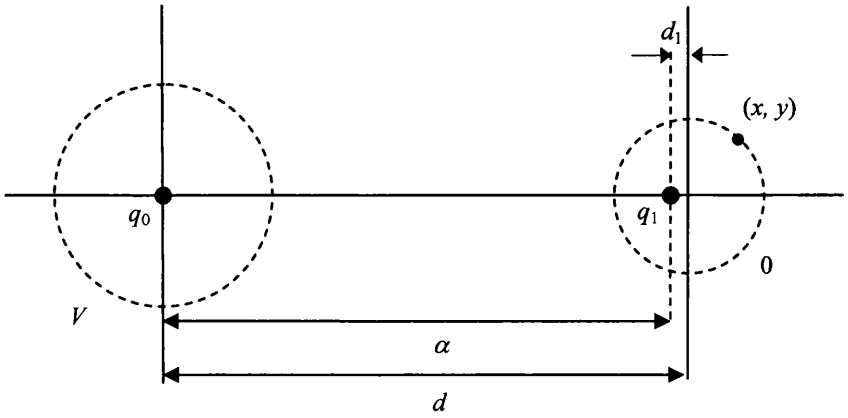


Figure AIV-2 Representation of first steps of the image charge method applied to Figure AIV-1. S_1 has been replaced by the point charge q_0, and S_2 has also been removed. The image point charge q_1 has been placed within the former locus of S_2 and lies a distance α from q_0 and a distance d_1 from the centre of S_2. The point (x, y) represents an arbitrary point on what would be the surface of S_2.

Therefore, for an arbitrary point (x, y) on the former S_2's surface, the relationship

$$0 = \frac{q_0}{4\pi\varepsilon}\frac{1}{\sqrt{x^2+y^2}} + \frac{q_1}{4\pi\varepsilon}\frac{1}{\sqrt{(x-\alpha)^2+y^2}} \qquad \text{(AIV.2)}$$

must hold. Having 'removed' the two spheres, the permittivity of the space in which they were imbedded is now uniform, and as such can be cancelled out from the above expression, making it largely immaterial. The 3-dimensional problem can be solved in the 2-dimensions of Figure AIV-2 without any loss of generality; the rotational symmetry around the horizontal x-axis ensuring that (x, y) represents any point on the surface of a sphere.

Letting

$$q_1 = -\beta q_0 \qquad \text{(AIV.3)}$$

then Eq. (AIV.2) can be written

$$\frac{1}{x^2+y^2} = \frac{\beta}{(x-\alpha)^2+y^2}$$

which implies that

$$x^2 + y^2 + \frac{2x\alpha}{\beta^2 - 1} = \frac{\alpha^2}{\beta^2 - 1}$$

and therefore

$$\left(x - \frac{\alpha}{1-\beta^2}\right)^2 + y^2 = \frac{\alpha^2\beta^2}{(1-\beta^2)^2} \qquad \text{(AIV.4)}$$

the equation of a circle of radius $\alpha\beta/(1-\beta^2)$ centred on $(\alpha/(1-\beta^2), 0)$. Ergo, the radius of S_2 must equal $\alpha\beta/(1-\beta^2)$ and d must equal $\alpha/(1-\beta^2)$, and as such $\beta = r_2 / d$, so therefore

$$q_1 = -\frac{r_2}{d} q_0 \qquad \text{(AIV.5)}$$

$$d_1 = d - \alpha = \frac{r_2^2}{d} \qquad \text{(AIV.6)}$$

Given that $r_2 < d$ then $\beta < 1$, implying that $|q_1| < |q_0|$. Also, since $d - \alpha = \beta r_2$, then $d - \alpha < r_2$, and therefore the charge q_1 must lie within S_2; as mentioned earlier, pre-empting this result.

The old boundary of S_2 is now back at 0V, however, q_1 has perturbed the potential at the old boundary of S_1. Therefore, an image charge of q_1, q_2, is needed within S_1 to compensate for q_1, as illustrated in Figure AIV-3.

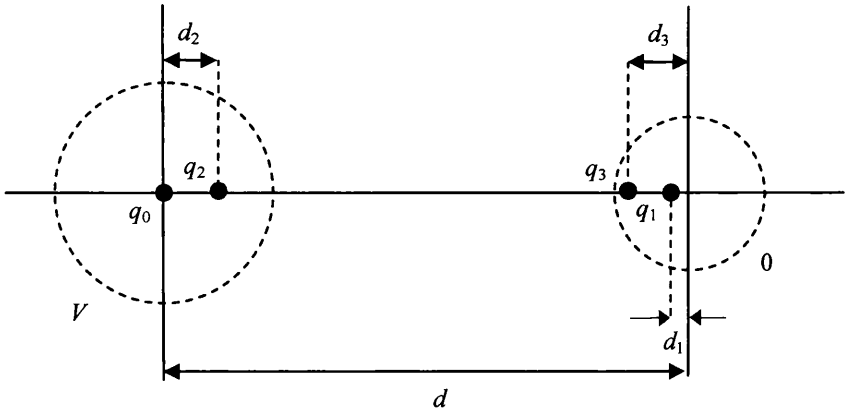


Figure AIV-3 Representation of first steps of the image charge method applied to Figure AIV-1. Further image charges have been placed within the old loci of their representative spheres: q_2 a distance d_2 from the old centre of S_1, and q_1 and q_3, residing d_1 and d_3 respectively from the old centre of S_2. As described in the text, this process continues indefinitely.

Repeating the previous reasoning, then r_1 must now equal $\alpha\beta/(1-\beta^2)$ and $(d - d_1)$ must equal $\alpha/(1-\beta^2)$. Thus $\beta = r_1 / (d - d_1)$ and therefore

$$q_2 = \frac{r_1 r_2}{d^2 - r_2^2} q_0 \qquad \text{(AIV.7)}$$

$$d_2 = \frac{dr_1^2}{d^2 - r_2^2} \tag{AIV.8}$$

using Eq.'s (AIV.5) and (AIV.6). q_2 perturbs the potential at S_2, and an image charge of q_2, q_3, is needed and so on.

The mutual capacitance of S_1 and S_2 is then the sum of the charges contained with S_2, q_Σ, divided by the potential difference between the two spheres, V:

$$C_{mutual} = \frac{q_\Sigma}{V} = \frac{q_1 + q_3 + q_5 + \cdots}{V} \tag{AIV.9}$$

The series q_Σ is rapidly convergent, $|q_{2n+1}| < |q_{2n-1}|$ n a positive integer, making the image charge method a practical and useful tool.

Applied to the Tip-QD junction as described in Chapters 6 and 7, for a tungsten sphere of radius 0.9nm separated from a SnO_2 sphere of radius 4nm by a distance of ~ 0.7nm, then $C_{mutual} \approx 9.55\times10^{-20}$F.

The second junction described in the above Chapters, QD-substrate, is between a sphere and a conducting plane. Consider now a point charge q situated a distance $d/2$ above a conducting plane at 0V, see Figure AIV-4(a).

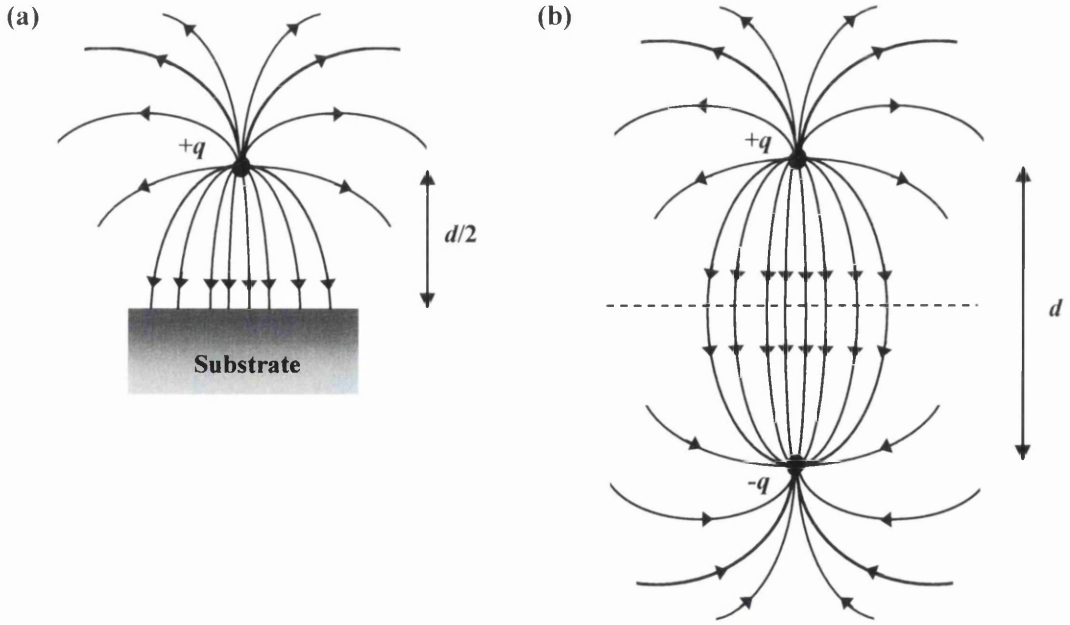


Figure AIV-4 Diagram representing (a) a point charge $+q$ situated a distance $d/2$ above a substrate, representative of a conducting plane, and (b) a point charge $+q$ and its image $-q$ separated by a distance d. The lines emanating/terminating from the point charge(s) indicate the electric field.

Removing the plane, the potential can be recreated simply by replacing it with a mirror charge $-q$, placed a distance $d/2$ below the former plane, see Figure AIV-4(b).

In this way, the image charge method applied to calculating the capacitance of two spheres can be applied to the problem of the capacitance between a sphere and a plane of separation $d/2$ by replacing the plane by a second sphere, of equal radius to the first, the pair separated by a distance d. For the QD-substrate junction of Chapters 6 and 7 with a SnO_2 sphere of radius 4 nm separated from the Si substrate by a 1nm thick oxide layer, the mutual capacitance is $C_{mutual} \approx 4.44\times10^{-18}$F.

AIV.2 Fermi's Golden Rule

For the time-dependent Schrödinger equation (TDSE), represent the full (perturbed) Hamiltonian of the system as the sum of the unperturbed reference Hamiltonian H_0 and a small time dependant perturbation H_1

$$(H_1 + H_0)\psi(t) = i\hbar\frac{\partial\psi(t)}{\partial t} \tag{AIV.10}$$

For the unperturbed case the solutions $H_0\phi_n = \hbar\omega_n\phi_n$ are assumed to be known, $\hbar\omega_n$ the eigenvalues and ϕ_n the eigenvectors of H_0 fulfilling the usual normalisation conditions.

Given that the perturbation, H_1, is small, it is possible to describe the wave function of the perturbed case as a series expansion of the unperturbed case

$$\psi(t) = \sum_n c_n(t)\phi_n e^{-i\omega_n t} \tag{AIV.11}$$

then from the TDSE

$$i\hbar\sum_n \frac{dc_n(t)}{dt}\phi_n e^{-i\omega_n t} + \hbar\sum_n \omega_n c_n(t)\phi_n e^{-i\omega_n t} = \sum_n c_n(t)H_1\phi_n e^{-i\omega_n t} + \hbar\sum_n \omega_n c_n(t)\phi_n e^{-i\omega_n t} \tag{AIV.12}$$

implying that

$$i\hbar\sum_n \frac{dc_n(t)}{dt}\phi_n e^{-i\omega_n t} = \sum_n c_n(t)H_1\phi_n e^{-i\omega_n t} \tag{AIV.13}$$

Multiplying by $\phi_m^* e^{+i\omega_m t}$ and integrating over all **r**, naturally recalling the orthogonality condition that $\int\phi_n^*\phi_m d^3\mathbf{r} = \delta_{nm}$, then

$$i\hbar\frac{dc_m(t)}{dt} = \sum_n c_n(t)H_{mn}e^{i\omega_{mn}t} \tag{AIV.14}$$

with

$$\omega_{mn} = \omega_m - \omega_n$$
$$H_{mn} = \int \phi_m^* H_1 \phi_n d^3 r \tag{AIV.15}$$

When the perturbation is turned on at time $t = 0$ let the system be in an initial state k, therefore $c_k(0) = 1$ while $c_n(0) = 0$ $n \neq k$. Assuming that for a first-order approximation, scattering out of this initial state over time is negligible, then in general $c_k(t) = 1$. Ergo, from Eq. (AIV.14) it follows that

$$i\hbar \frac{dc_m(t)}{dt} = H_{mk} e^{i\omega_{mk} t} \tag{AIV.16}$$

and so

$$c_m(t) = \frac{-i}{\hbar} \int_0^t H_{mk} e^{i\omega_{mk} t} dt = \frac{H_{mk}}{\hbar \omega_{mk}} \left(1 - e^{i\omega_{mk} t}\right) \tag{AIV.17}$$

Therefore, the probability at time t that ψ is an eigenstate of H_0 with energy $\hbar\omega_n$ is

$$\begin{aligned} P_n(t) &= \left|\langle \phi_n | \psi(t) \rangle\right|^2 \\ &= \left|c_n(t)\right|^2 \\ &= \frac{\left|H_{nk}\right|^2}{\hbar^2 \omega_{nk}^2} \left(2 - (e^{i\omega_{nk} t} + e^{-i\omega_{nk} t})\right) \\ &= \frac{\left|H_{nk}\right|^2}{\hbar^2} 4t^2 \operatorname{sinc}^2\left(\frac{\omega_{nk} t}{2}\right) \end{aligned} \tag{AIV.18}$$

When $t \to \infty$, a time long enough to ensure that the scattering process, in this case the tunnelling event, has been completed, then from the imaginary part of

$$\oint \frac{e^{i2\omega z}}{(\omega z)^2} dz$$

integrated over a closed semi-circular contour in the upper half of the complex plane, where its residue was obtained from its Laurent expansion, it follows that

$$\int_{-\infty}^{\infty} 2^2 \frac{\sin^2(\omega_{nk} t / 2)}{\omega_{nk}^2} d\omega_{nk} = 2\pi t \tag{AIV.19}$$

that is, for $t \to \infty$, $4t^2 \operatorname{sinc}(\omega_{nk} t / 2)$ has the properties of a delta function, $2\pi t \delta(\omega_{nk})$. The transmission probability per unit time from state k to state n i.e. the rate of change of $P_n(t)$ is

$$\Gamma_{k \to n} = \frac{d}{dt} \left|c_n(t)\right|^2 = \frac{2\pi}{\hbar} \left|H_{nk}\right|^2 \delta(\hbar\omega_{nk}) = \frac{2\pi}{\hbar} \left|H_{nk}\right|^2 \delta(E_n - E_k) \tag{AIV.20}$$

using $\delta(ax) = |a|^{-1}\delta(x)$. This implies that a scattering process only takes place if energy is conserved, see for example Gasiorowicz [100]. Eq. (AIV.20) is a form of the famous Fermi's Golden Rule [70].

For the tunnelling processes described in Chapters 6 and 7, work has to be done by the power source for the electron to tunnel between sites, and energy is taken to charge the grain by the arriving electron. With each tunnelling event then, there is an associated change in the energy of the system. This has been measured via the Helmholtz free energy, and as such, this change in the free energy associated with the evolution of the system from state k to state n, ΔF, should be reflected in the matrix element H_{nk} i.e.

$$H_{nk} \to H_{nk} e^{-i\Delta F t/\hbar} \tag{AIV.21}$$

and so following through Eq.'s (AIV.14) to (AIV.20), the rate of tunnelling between the initial state k to the final state n is given by

$$\Gamma_{k\to n} = \frac{2\pi}{\hbar}\left|H_{nk}\right|^2 \delta\left(E_n - E_k - \Delta F\right) \tag{AIV.22}$$

In this way, summing over all the available states, the total rate of tunnelling from the occupied states, labelled i, on one side of a barrier to the unoccupied states, labelled j, on the other side of the barrier is found from

$$\Gamma(\Delta F) = \frac{2\pi}{\hbar}\sum_i\sum_f \left|H_{fi}\right|^2 f_i(1-f_f)\delta\left(E_f - E_i - \Delta F\right) \tag{AIV.23}$$

f_n representing the Fermi occupation factor. Taking the states to be densely packed then the summations can be converted into integrals in momentum or **k**-space i.e.

$$\sum \to 2\int \frac{d^3\mathbf{k}}{(2\pi)^3}$$

per unit volume, realising that there are two possible electron states, due to spin, for each state of Eq. (AIV.23). The differential wave vector element can be simplified via $d^3\mathbf{k} = d\Omega_k k^2\, dk$, $d\Omega_k$ the solid angle differential, so that

$$2\int \frac{d^3\mathbf{k}}{(2\pi)^3} \to \int \frac{k^2 dk}{\pi^2}$$

The number of electron states per unit volume within an infinitesimal energy range is given by

$$D(E)dE = \left(k/\pi\right)^2 dk$$

see Blakemore [35] for details, and so

$$\sum \to \int D(E)dE$$

where $D(E)$ represents the density of electron states, allowing the summations over momentum of Eq. (AIV.23) to be converted into energy integrals such that

$$\Gamma(\Delta F) = \frac{2\pi}{\hbar} \int_{E_i}^{\infty} \int_{E_f}^{\infty} D(E_i) D(E_f) \left| H_{fi} \right|^2 f_i (1 - f_f) \delta\left(E_i - E_f + \Delta F\right) dE_i dE_f \quad \text{(AIV.24)}$$

giving an usable and practical expression for the tunnelling rate of electrons through a potential barrier. The charge carried by each electron is $-e$, and as such the conventional tunnelling current is simply $I = e\Gamma$.

Bibliography

[1] P. Ball, Technology Review **104** 31 (2001)

[2] R.F. Service, Science **309** 95 (2005)

[3] A. Parviz, B. Ryan, G.M. Whitesides, IEEE Trans Adv Pack **26** 233 (2003)

[4] E. Winfree, F. Liu, L.A. Wenzler and N.C. Seeman, Nature **394** 539 (1998)

[5] W.M.D. Wright and D.G. Chetwynd, Nanotechnology **9** 133 (1998)

[6] S.P. Wilks, T.G.G. Maffeïs, K.S. Teng, G.T. Owen and M.W. Penny, J. Vacuum Science and Technology B **22**(4) (2004) 1995

[7] T.G.G. Maffeïs, G.T. Owen, M.W. Penny, H. Ferkel and S.P. Wilks. Appl. Surf. Sci. **234** 2 (2004)

[8] I. Amlani, A.O. Orlov, G. Toth, G.H. Bernstein, C.S. Lent and G.L. Snider, Science **284** 289 (1999)

[9] G.H. Bernstein, I. Amlani, A.O. Orlov, C.S. Lent and G.L. Snider, Nanotechnology **10** 166 (1999)

[10] C.S. Lent, P.D. Tougaw, W. Porod and B.G. Bernstein, Nanotechnology **4** 49 (1993)

[11] M. Otobe, H. Yajima and S. Oda, Appl. Phys. Lett. **72** 1089 (1998)

[12] H. Osman, J. Schmidt, K. Svensson, R.E. Palmer, Y. Shigeta and J.P. Wilcoxon, Chem. Phys. Lett. **330** 1 (2000)

[13] P. Mesquida and A. Stemmer, Advanced Materials **13**(18) 1395 (2001)

[14] M. Penny (2006) Scanning Tunnelling Microscopy/Scanning Tunneling Spectroscopy Characterisation of Nanopatterened Tin Dioxide. PhD Thesis, MNC, UW Swansea.

[15] C.Malagu, V.Guidi, M.Stefancich, M.C.Carotta and G.Martinelli, J. Appl. Phys. **91** 808 (2002)

[16] C. Malagu, M.C. Carotta, S. Galliera, V. Guidi, T.G.G. Maffeis, G. Martinelli, G.T. Owen and S.P. Wilks, Sens. Actuators B **103** 50 (2004)

[17] C. Malagu, M.C. Carotta, H. Fissan, V. Guidi, M.K. Kennedy, F.E. Kruis, G. Martinelli, T.G.G. Maffeis, G.T. Owen and S.P. Wilks, Sens. Actuators B **100** 283 (2004)

[18] A. Martí, L. Cuadra and A. Luque, IEEE Trans. Electron Devices **48** 2304 (2001)

[19] M. Mucucci, K. Hess and G.J. Iafrate, J. Appl. Phys. **77** 3267 (1995)

[20] T. Torsti, V. Lindberg, M.J. Puska and B. Hellsing, Phys. Rev B **66** 235420 (2002)

[21] J.B. Krieger, Y. Li and G.F. Iafrate, Phys. Rev. A **45** 101 (1992)

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

[23] A. Savin, C.J. Umrigar and X. Gonze, Chem. Phys. Lett. **288** 391 (1998)

[24] O.V. Grisenko and E.J. Baerends, J. Chem. Phys. **120** 8364 (2004)

[25] O.V. Grisenko and E.J. Baerends, J. Chem. Phys. **117** 9154 (2002)

[26] D.P. Chong, O.V. Grisenko and E.J. Baerends, J. Chem. Phys. **116** 1760 (2002)

[27] O.V. Grisenko, B. Braïda and E.J. Baerends, J. Chem. Phys. **119** 1937 (2003)

[28] A. Görling, Phys. Rev. A **54** 3912 (1996)

[29] J. Kohanoff (2006) Electronic Structure Calculations for Solids and Molecules. 1st ed. Cambridge: Cambridge University Press.

[30] J.P. Perdew, R.G. Parr, M. Levy and J.L. Balduz Jr., Phys. Rev. Lett. **49** 1961 (1982)

[31] T.G.G. Maffeis, G.T. Owen, C. Malagu, G. Martinelli, M.R. Kennedy, F.E. Kruis, S.P. Wilks, Surf. Sci. 550 21 (2004)

[32] C. Wasshuber (1997) About Single-Electron Devices and Circuits. PhD Thesis. Available from http://www.iue.tuwien.ac.at/phd/wasshuber [Accessed on 02/08/2004].

[33] J. Tersoff and D.R. Hamann, Phys. Rev. Lett. **50** 1998 (1983)

[34] A. Sommerfeld and N.H. Frank, Rev. Mod. Phys. **3** 1 (1931)

[35] J.S. Blakemore (1985) Solid State Physics, 2nd ed. Cambridge: Cambridge University Press.

[36] C. Kittel (1996) Introduction to Solid State Physics, 7th ed. New York: John Wiley and Sons.

[37] E. Merzbacher (1970) Quantum Mechanics. 2nd ed. Singapore: John Wiley and Sons.

[38] E. Cartlidge, Physics World **20** 20 (2007)

[39] R.T. Tung, Phys. Rev. Lett. **84** 6078 (2000)

[40] J. Bardeen, Phys. Rev. **71** 717 (1947)

[41] G. Williams, G.S.V.Coles, MRS Bull. **24** 25 (1999)

[42] M.K. Kennedy, F.E. Kruis, H. Fissan, B.R. Metha, S. Stappert and G. Dumpich, J. Appl. Phys. **93** 551 (2003)

[43] J.F. McAleer, P.T. Moseley, J.O.W. Norris, D.E. Williams, J. Chem. Soc., Faraday Trans. 1 **83** 1323 (1987)

[44] N. M. Beekmans, J. Chem. Soc., Faraday Trans. 1 **74** 31 (1978)

[45] C. Xu, J. Tamaki, N. Miura, and N. Yamazoe, Sens. Actuators B **3** 147 (1991)

[46] N.S. Baik, G. Sakai, K. Shimanoe, N. Miura and N. Yamazoe, Sens. Actuators B **65** 97 (2000)

[47] Y. Shimizu and M. Egashira, MRS Bull. **24** (1999) 19

[48] F.J. Arlinghaus, J. Phys. Chem. Solids **35** 931 (1974)

[49] Y.R. Park and K.J. Kim, J. Appl. Phys **94** 6401 (2003)

[50] E. Kreyszig (1999) Advanced Engineering Mathematics. 8th ed. Singapore: John Wiley and Sons.

[51] R.W. Hamming (1986) Numerical Methods for Scientists and Engineers. 2nd ed. Toronto: Dover.

[52] W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling (1992) Numerical Recipies in C: The Art of Scientific Computing. 2nd ed. Cambridge: Cambridge University Press

[53] C.G. Fonstad and R.H. Rediker, J. Appl. Phys. **42** 2911 (1971)

[54] K.J. Button, C.G. Fonstad, W. Dreybrodt, Phys. Rev. B **4** 4539 (1971)

[55] P.K. Clifford and D.T. Tuma, Sens. Actuators **3** 255 (1983)

[56] S.R. Morrison, Sens. Actuators **12** 425 (1987)

[57] J.S. Blakemore, S.S.E. **25** 1067 (1982)

[58] H. Teterycz and B.W. Licznerski, J. Electrochem. Soc. **153** H94 (2006)

[59] J.C. Bernède, N. Barreau, S. Marsillac and L. Assman, Appl. Surf. Sci. **195** 222 (2002)

[60] Th. Kulger, Å. Johansson, I. Dalsegg, U. Gelius and W.R. Salaneck, Synth. Met. **91** 143 (1997)

[61] M. Batzill and U. Diebold, J. Phys.: Condens. Matter **18** L1 (2006)

[62] G.L. Shen, R. Casanova, G. Thornton and I. Colera, J. Phys.: Condens. Matter **3** S291 (1991)

[63] M.S. Soda an P.K. Dubey, J. Phys. D: Appl. Phys **3** 139 (1970)

[64] D. Bohm (1989) Quantum Theory, 1st ed. New York: Dover.

[65] G.B. Arfken and H.J. Weber (2001) Mathematical Methods for Physicists. 5th ed. California: Harcourt Academic Press

[66] R. Burden, J. Faires and A. Reynolds (1981) Numerical Analysis. 2nd ed. Boston: Prindle, Weber and Schmidt.

[67] E. Anderson, Z. Bai, C. Birschof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney and D. Sorensen (1999) LAPACK Users' Guide. Available from http://www.netlib.org/lapack/lug [Accessed on 10/08/2004].

[68] R.B. Lehoucq, D. Sorensen and C. Yang (1997) ARPACK Users' Guide: Solution of Large Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods. Available from http://www.caam.rice.edu/software/ARPACK [Accessed on 01/08/2004]

[69] F.W. Byron and R.W. Fuller (1992) Mathematics of Classical and Quantum Physics. 1st ed. New York: Dover.

[70] M. Weissbluth (1978) Atoms and Molecules, 1st ed. London: Academic Press.

[71] J.C. Flores, J. Phys. A: Math. Gen. **26** 4117 (1993)

[72] M.G. Burt, J.Phys: Condens. Matter **4** 6651 (1992)

[73] G. Woan (2000) The Cambridge Handbook of Physics Formulas. 1st ed. Cambridge: Cambridge University Press.

[74] M. Levy, Proc. Nat. Acad. Sci. USA **76**, 6062 (1979).

[75] R.O. Jones (2006), Introduction to Density Functional Theory and Exchange-Correlation Energy Functionals. Available from http://www.fz-juelich.de/nic-series/volume31 [Accessed on 15/10/2006].

[76] P. Hohenberg and W. Kohn, Phys. Rev. **136** B864 (1964).

[77] W. Kohn, A.D. Becke and P.G. Parr, J. Phys. Chem. **100** 12974 (1996)

[78] N.D. Mermin, Phys. Rev. **137** A1441 (1964)

[79] H. Eschrig and W. E. Pickett, Solid State Commun. **118** 123 (2001).

[80] W. Kohn, A. Savin, and C. A. Ullrich, Int. J. Quantum Chem. **100** 20 (2004).

[81] E. Engel and R.M. Dreizler, J. Comput Chem. **20** 31 (1999)

[82] D.M. Ceperley and B.J. Alder, Phys. Rev. Lett **45** 566 (1980)

[83] J. Muscat, A. Wander and N.M. Harrison, Chem. Phys. Lett. **342** 397 (2001)

[84] B. Tantar and D.M. Ceperley, Phys. Rev. B **39** 5005 (1989)

[85] E. Engel and S.H. Vosko, Phys. Rev. A **47** 2800 (1993)

[86] T. Kotani, Phys. Rev. B **50** 14816 (1994)

[87] T. Kotani, Phys. Rev. Lett. **74** 2989 (1995)

[88] J.C. Slater, Phys. Rev. **81** 385 (1951)

[89] E. Engel and S.H. Vosko, Phys. Rev. B **47** 13164 (1993)

[90] M. Levy and J.P. Perdew, Phys Rev A **32** 2010 (1985).

[91] S.H. Vosko, L. Wilk and M. Nuissar, Can. J. Phys. **58** 1200 (1980)

[92] L.A. Cole and J.P. Perdew, Phys. Rev. A **25** 1265 (1982)

[93] D. Halliday, R. Resnick and J. Walker (1997) Fundamentals of Physics Extended, 5th ed. New York: John Wiley and Sons, Inc.

[94] V. Lantto, P. Romppainen and S. Leppävuori Sens. Actuators **14** 149 (1988)

[95] J. Gullberg (1997) Mathematics: from the birth of numbers. 1st ed. New York: W.W. Norton and Company, Inc.

[96] S. Samson and C.G. Fonstad, J. Appl. Phys. **44** 4618 (1973)

[97] M.C. Carotta, C. Dallara, G. Martinelli and L. Passari, Sens. Actuators B **3** 191 (1991)

[98] S. Weinstein, Physics World, **20** 18 (2007)

[99] J. Naser and H. Ferkel, Nanostructured Mater. **12** 451 (1999)

[100] S. Gasiorowicz (1974) Quantum Physics. 1st ed. New York: John Wiley and Sons, Inc.

[101] R. Turton, (1995) The Quantum Dot. A Journey into the Future of Microelectronics. Oxford: W.H.Freeman and Co.

[102] P. Rinke, A. Qteish, J. Neugebauer, C. Freysoldt and M. Scheffler, 2005 New Journal of Physics **7** 126

[103] L. Hedin 1965 Phys. Rev. **139** A796

[104] M. Kaku (1993) Quantum Field Theory, 1st ed. New York: Oxford University Press.

[105] E.N. Economou (1979) Solid-State Sciences Vol. 7, Green's Functions in Quantum Physics, ed M. Caronda, P. Fulde, H.-J. Queisser 1st ed. Berlin: Springer-Verlag.

[106] C. Friedrich and A. Schindlmayr (2006), Many Body Perturbation theory: the GW approximation. Available from http://www.fz-juelich.de/nic-series/volume31 [Accessed on 15/10/2006].

[107] H.N. Rojas, R.W. Godby and R. J. Needs, Phys. Rev. Lett. **74** 1827 (1995)

[108] M.M. Rieger, L. Steinbeck, I.D.White, H.N. Rojas and R.W. Godby, Comp. Phys. Commun. **117** 211 (1999)

[109] L. Steinbeck, A. Rubio, L. Reining, M. Torrent, I.D. White and R.W. Godby, Comp. Phys. Commun. **125** 105 (2000)

[110] R. Kwapinski, R. Taranko and E. Taranko, Phys. Rev. B **66** 035313 (2002)

[111] L.V. Keldysh, Soviet Physics JETP **20** 1018 (1964)

[112] A. Martin-Rodero, F. Flores and N.H. March. Phys. Rev. B **38** 10047 (1988)

[113] J.R. Cerdá, P.L. de Andres, F. Florez and R. Pérez. Rev B **45** 8721 (1992)

[114] N. Mingo, L. Jurczyszyn, F.J. Garcia-Vidal, R. Saiz-Pardo, P.L. de Andres, F. Flores, S.Y. Wu and W. More. Phys. Rev. B **54** 2225 (1996)

[115] C. Caroli, R. Combescot, P. Nozieres and D. Saint-James. J. Phys. C **5** 21 (1975)

[116] C.B. Duke (1969) Tunnelling in Solids, New York: Academic Press Inc.

[117] E.C. Goldberg, A. Martin-Rodero, R. Monreal and F. Flores. Phys. Rev. B **39** 5684 (1989)

[118] F.J. Garcia-Vidal, A. Martin-Rodero, F. Flores, J. Ortega and R. Pérez. Phys. Rev. B **44** 11412 (1991)

[119] A.T. Johnson, L.P. Kouwenhoven, W. de Jong, N.C. van der Vaart, C.J.P.M. Harmans and C.T. Foxton, Phys. Rev. Lett. **69** 1592 (1992)

[120] J. Bardeen, Phys. Rev. Lett. **6** 57 (1961)

[121] X. Aymerich-Humet, F. Serra-Mestres and J. Millan, J. Appl. Phys. **54** 2850 (1983)

[122] X. Aymerich-Humet, F. Serra-Mestres and J. Millan, S.E.E. **24** 981 (1981)

[123] H. Werner and G. Raymann, Math. Comp. **17** 193 (1963)

[124] J.H. Wilkinson (1965) The Algebraic Eigenvalue Problem. 1st ed. Belfast: Oxford University Press.

[125] G.W. Stewart (1998) Matrix Algorithms Volume 1: Basic Decompositions. 1st ed. Philadelphia: Society for Industrial and Applied Mathematics.

[126] E. Isaacson and H.B. Keller (1966) Analysis of Numerical Methods. 1st ed. New York: John Wiley and Sons, Inc.

Index

C

D

E

F

G

H

I

J

K

L

M

N

O

D

Q

R

S

T

U

V

W