UNIVERSITY OF SOUTHAMPTON

Faculty of Social Sciences
School of Mathematical Sciences

Mathematical Modelling of Dielectric Inclusions and Photoactive Alignment Layers in Liquid Crystal Cells

by

Jordan Ronald Ewan Gill
MMath

A thesis for the degree of
Doctor of Philosophy

March 2021

University of Southampton

Abstract

Faculty of Social Sciences
School of Mathematical Sciences

Doctor of Philosophy

Mathematical Modelling of Dielectric Inclusions and Photoactive Alignment Layers in Liquid Crystal Cells

by Jordan Ronald Ewan Gill

This work concerns two areas. The first is a study of the photoalignment behaviour of thin (approximately 20-50 nm) PAAD films in an effort to determine their molecular alignment upon exposure to polarised light. In conjunction with experimental work carried out by another member of our research group, a complex model is developed to model the transmission of the s- and p-polarised components of light across the experimental system. The model takes into account the mixing of polarisations caused by the photoinduced birefringence of the PAAD film, and couples this with the interference effects caused by the transmission across the multiple different layers of the system. A new method for modelling multi-layer propagation is developed for this problem, which proves to have both the accuracy and stability of the current leading standard method, but with vast improvements in computation time.

The second area of work concerns the multi-scale modelling of doped dielectric systems. The method of homogenisation is utilised to derive an effective permittivity to describe composite systems composed of a host medium and small inclusions, which requires the assumption that the system is periodic at the microscopic scale. This is done for 2D elliptical particles in anisotropic media, and for 3D spheroidal particles in isotropic media. The assumption of periodicity is then relaxed and the applicability of homogenisation is studied when disorder is entered into the system. The results are then compared to traditional effective medium theories used to solve such host/inclusion problems.

Contents

List of Figures **ix**

List of Tables **xv**

Listings **xvii**

Declaration of Authorship **xvii**

Acknowledgements **xix**

1 Introduction **1**

2 Electromagnetic Theory **7**
2.1 Maxwell's Equations . . . 7
2.2 Polarisation and the D Field . . . 8
2.3 Electromagnetic Energy and the Permittivity Tensor . . . 10
2.4 Wave Solutions to Maxwell's Equations . . . 12
2.4.1 Propagation in Isotropic Media . . . 12
2.4.2 Propagation in Anisotropic Media . . . 15
2.4.2.1 O- and E-Wave Solutions . . . 15
2.4.2.2 Field Orientations . . . 17
2.5 Fields at a Material Discontinuity . . . 19
2.5.1 Interface Boundary Conditions . . . 19
2.5.2 S- and P-Polarised Waves . . . 21
2.5.3 Reflection and Refraction in Isotropic Media . . . 22
2.5.4 The Fresnel Amplitude Coefficients in Isotropic Media . . . 24
2.5.5 Transmittance and Reflectance . . . 26
2.5.6 Total Internal Reflection and the Evanescent Wave . . . 27
2.5.7 Fresnel Amplitude Coefficients at an Isotropic/Anisotropic Boundary . . . 28
2.6 Propagation Through Multiple Layers . . . 32
2.6.1 Fabry-Perot Resonator . . . 32
2.6.2 The T-Matrix Formulation . . . 36
2.6.3 The S-Matrix Formulation . . . 38
2.7 Diffraction Theory . . . 39
2.7.1 The Grating Formula . . . 39
2.7.2 Rayleigh-Sommerfeld Diffraction Integrals . . . 42
2.7.3 Fraunhofer and Raman-Nath Diffraction . . . 44

2.7.4 Non-Paraxial Diffraction and the Angular Spectrum 46

3 Modelling Diffraction in Thin PAAD Films 55
3.1 Introduction 55
3.2 Optically Controlled, Dynamic Wave Plates 56
3.2.1 The Nematic Twisted Cell and PAAD Alignment Layer 56
3.2.2 Model and Results 57
3.3 Experimental Photoalignment of Thin PAAD Films 61
3.3.1 The Experimental Setup and Results 61
3.4 New Experimental Results and Model 63
3.5 Ray Method for Multilayer Propagation 64
3.5.1 Transmission Across an Isotropic Layer 64
3.5.2 Transmission Across Multiple Layers - Iterated Ray Method . . . 67
3.5.3 Generalised Multi-Mode Ray Method 70
3.5.4 Transmission Across an Anisotropic Layer 72
3.6 Modelling Isotropic PAAD Layers 74
3.6.1 Transmission Through a Glass Slide 75
3.6.2 Transmission of Isotropic PAAD and Substrate 78
3.7 Modelling PAAD diffraction 80
3.7.1 Surface Relief Gratings in Thick Layers 80
3.7.2 Birefringence Gratings in Thin Layers 82
3.7.2.1 The Modelling Approximations 82
3.7.2.2 Transmission and Reflection of Anisotropic PAAD . . . 83
3.7.2.3 Diffraction from an Isotropic/Anisotropic Grating Layer 87
3.7.2.4 Diffraction from an Aligned PAAD Grating 92
3.7.2.5 Modelling the Experimental System 94
3.7.2.6 Comparison with Experimental Results 97

4 Multiscale Modelling of Doped Liquid Crystal Systems 101
4.1 Introduction 101
4.2 Homogenisation and Effective Medium Theories 102
4.3 Dielectric Particles in Isotropic Medium 105
4.3.1 Setting Up the Problem 105
4.3.2 The Homogenised Equations 106
4.3.3 1D Simplification 110
4.3.4 2D Particles in Medium - Homogenisation vs Finite Elements . . 111
4.3.4.1 The Effective Permittivity 111
4.3.4.2 Accuracy of the Homogenised Solution 113
4.4 Homogenised Electrostatic Problem for Disordered, Spatially Varying Ellipsoidal Particles 116
4.4.1 Varying Particle Geometries 117
4.4.2 Homogenised Equations 118
4.5 Calculating the Effective Permittivity 121
4.5.1 Non-Rotating Particles in Isotropic Medium 121
4.5.1.1 The General Cell Problem 121
4.5.1.2 Simplified First Order Matched Asymptotic Problem . 123
4.5.2 First Order Problem for Rotating Particles in Isotropic Medium . 128

4.5.3 Effective Permittivity for Particles in Isotropic Medium 129
4.5.4 Particles in Anisotropic Medium 132
4.5.4.1 The Canonical Problem 132
4.5.4.2 Metallic Particles in Anisotropic Medium 136
4.5.4.3 Dielectric Particles in Anisotropic Medium 140
4.5.5 Effective Permittivity of Dielectric Particles in Anistropic Medium 143
4.6 3D particles in Isotropic Medium 145
4.6.1 Traditional Effect Medium Theories 145
4.6.2 Spherical Particles 150
4.6.3 Spheroidal Particles 152
4.6.3.1 The Cell Problem - Metallic Prolate Spheroids 152
4.6.3.2 The Cell Problem - Metallic Oblate Spheroids 158
4.6.3.3 The Cell Problem - Dielectric Spheroids 160
4.6.4 3D Effective Permittivity 161
4.6.4.1 The Principal Permittivity Tensor 161
4.6.4.2 Arbitrary Particle Orientation 162
4.6.5 Comparison of Homogenisation with Traditional Theories 166
4.6.6 A Periodic Approximation to the Cell Problem 169
4.6.7 Comparison of Homogenisation with Traditional Theories Cont. 173
4.7 Disordered 2D Particles 176
4.7.1 Homogenisation Vs Finite Elements 176
4.7.2 Comparison of Homogenisation with Traditional Theories 180

5 Conclusions **183**

References **187**

List of Figures

2.1 (a) Diagram showing the orientation of fields associated with the o-wave. (b) Diagram showing the orientation of the fields associated with the e-wave. 18
2.2 A diagram of the basis vectors $\boldsymbol{E}_s$ and $\boldsymbol{E}_p$ for a wave incident with wavevector $\boldsymbol{k}^{(i)}$ on a boundary with normal $\boldsymbol{n}$. 22
2.3 (a) Wavevectors associated with the forward process of reflection and refraction. (b) The reverse process considered in the derivation of the Stokes' relations. 25
2.4 A schematic of the fields arising in a three layer system subject to an incident field E^{in}. 33
2.5 A schematic showing the various quantities used to define irradiance, radiance, and radiant intensity. 48

3.1 Schematic of the bulk alignment of a nematic, twisted cell in its (a) default helical state induced by the surface alignments with no applied field and (b) the reoriented state under an applied voltage. 57
3.2 A schematic of the twist/untwist experimental setup. The blue beam and polariser are used to align the PAAD alignment layer on the near face of the cell. The cell is then probed with the red beam through parallel polarisers for successive twisted and untwisted alignments. 58
3.3 Microscope images of the PAAD aligned cell through multiple twist and untwist states, aligned with a green (a-d) and blue (e-i) beam. 59
3.4 Transmitted intensity data for green (a) and blue (c) beams for repeated twist and untwist steps. (b) and (d) show the asymptotic intensities at each step for the green and blue beams, respectively. 60
3.5 Complex refractive index of the PAAD as measured in [35]. 60
3.6 Schematic diagram of the diffraction efficiency set up. The green pump beam forms an interference pattern on the PAAD, inducing a photoaligned grating. The diffraction efficiency of the grating is measured using the red probe beam. 61
3.7 (A) Schematic of a three layer system, depicting the input field F_0, along with the rays resulting from subsequent reflections and transmissions at the two interfaces. (B) Schematic of a three layer system, depicting the input field, F_0, total reflected field, B_0, and total transmitted field, F_2. $F_1$and B_1 are the sum of all the forward and backward fields inside the centre layer. 65
3.8 Diagram of the geometry needed to calculate the difference in optical path length between two subsequent transmitted or reflected rays. . . . 66

3.9 (a) Schematic representation of the N layer system as a three layer system. The layers 3 to $N-1$ are replaced with an effective boundary, represented by a jagged interface. (b) Schematic representation of the N layer system as a general three layer system. The forwards and backwards fields, F_i and B_i, in any layer i can be evaluated by replacing layers 1 to $i-1$ and $i+1$ to $N-1$ with effective layers. 67
3.10 A comparison of the transmission across a three layer system as a function of angle of incidence, ψ, for (a) the T-matrix method and (b) the S-matrix and Iterated Ray method. The material refractive indices are 2, 1, and 2 (in that order) and the layer thickness is eight wavelengths. The instability of the T-matrix method can be observed in (a), where the transmission can be observed rising sharply around 38°, despite the critical angle for total reflection being around 30°. 69
3.11 (a) A comparison of the amplitudes of the transmission coefficients across a four layer system and the reflection coefficients from the initial three layer subsystem, as a function of the angle of incidence, ψ, as given by the Iterated Ray and S-Matrix formulisms. (b) The angles of the coefficients. The system refractive indices are 2, 1, 2, and 2.5, and the layer thicknesses are 1.25 and 0.5 wavelengths. 70
3.12 Graphs showing the speed increase of the Iterated Ray method over the S-matrix approach when calculating (a) the transmission and (b) reflection. The values are calculated from the average run times provided by the Matlab function `timeit` for 2,750 samples of random structures of varying layer thicknesses and refractive indices for varying numbers of angles of incidence. 71
3.13 Diagram used to find the optical path difference between two emerging rays when there is mode mixing within a layer. The forwards and backwards angles of propagation, θ_f and θ_b, need not be the same, since the angle of the o-rays and forwards and backwards e-rays are all different. 73
3.14 A comparison of the various transmittances, T, and reflectances, R, for s- and p-polarised incident fields to s- and p-polarised output fields, as calculated using the Iterated Ray Method and numerically with Comsol. The system refractive indices are $n_{in} = 1$, $n_e = 5$, $n_o = 1.5$, and $n_{out} = 2$. The layer thickness is two wavelengths and the optical axis is at an angle of $\pi/3$ with the y axis, with its projection in the xz plane (plane of incidence) equal to $\pi/5$. 75
3.15 A comparison of the phase, in radians, of the Fresnel coefficients for the transmission and reflection of s- and p-polarised incident fields to s- and p-polarised output fields. The system refractive indices are $n_{in} = 1$, $n_e = 5$, $n_o = 1.5$, and $n_{out} = 2$. The layer thickness is two wavelengths and the optical axis is at an angle of $\pi/3$ with the y axis, with its projection in the xz plane (plane of incidence) equal to $\pi/5$. 76
3.16 Graphs showing the experimental (blue) and fit (orange) transmission data for a HeNe 632.8nm laser through a glass slide. 76
3.17 Graphs showing the experimental and fitted transmission curves for a (left) s-polarised and (right) p-polarised 632.8nm beam incident onto the PAAD/ITO/glass system. 79

3.18 Normalised transmissions as a function of the angle of incidence, ψ, for the s- and p-polarisations of the first diffracted orders for an aligned PAAD grating. 82
3.19 A schematic of the approximate grating geometry and the key modelling assumption, that the transmission through each thin section of isotropic or anisotropic PAAD can be approximated by the effective transmission coefficients t_{iso} or t_{ani} for an infinitely extending layer. 83
3.20 Diagram depicting the angles used to describe the optical axis of the PAAD. 84
3.21 Schematic of the PAAD grating diffraction problem. The incident field, E_{in}, has its plane of incidence in the xz plane. The grating is treated as an effective layer (represented by a jagged line at $z = 0$) between the first and third layers. The emerging field, $E_e = E_t|_{z=0^+}$, is given by a square wave approximation applied to the isotropic and anistropic layer transmission coefficients, and gives rise to the transmitted field, E_t, comprised of the various diffracted orders described by their effective transmition coefficients $t^{(i)}$. 87
3.22 A comparison of the diffracted field as calculated by the Angular Spectrum and Rayleigh-Sommerfeld methods at normal incidence. (a) Shows the norm of the electric field while (b) shows the magnitude of the individual orders, with the black dashed lines marking the evanescent cut-off. The system parameters are $f = 3/4$, $\Lambda = 2.5\lambda$, and $z = 0.224\lambda$. 91
3.23 A comparison of the diffracted field as calculated by the Angular Spectrum and Rayleigh-Sommerfeld methods at a 60° incidence angle. (a) Shows the norm of the electric field while (b) shows the magnitude of the individual orders, with the black dashed lines marking the evanescent cut-off. The system parameters are $f = 3/4$, $\Lambda = 2.5\lambda$, and $z = 0.224\lambda$. 91
3.24 A schematic of the model of the full experimental system. The system is represented as three layers, air, ITO, and air, separated by effective boundaries (jagged lines) that account for the effects of the PAAD and glass layers. We neglect diffracted orders higher than the first, and the left section shows the fundamental order (blue) crossing the effective PAAD boundary and entering the ITO layer after diffracting into the negative (green), fundamental, and positive (purple) orders (with both s- and p-polarisations of each order making up the 6 modes of the model). The second effective boundary accounting for the glass layer is a purely isotropic subsystem, so any mode reflects back as the same mode, shown in the centre section. We ignore higher orders and repeat diffraction effects, so a backwards fundamental mode reflects and diffracts to fundamental, positive and negative modes, while a backwards diffracted mode can only reflect back to the same mode. This is shown in the right section. 93
3.25 (a) The experimental diffraction efficiencies and (b) the best fit results from the theoretical model. The optical axis is at $\theta = 90°$, $\phi = 22°$, all other system parameters are according to the isotropic fit results. 98
3.26 The theoretical diffraction efficiencies for the optical axis at $\theta = 90°$, $\phi = 22°$, $L_{ITO} = 100$ nm, and all other system parameters according to the isotropic fit results. 98

4.1 Periodic approximation of the microscopic geometry of the doped system. 106

4.2 The unit cell of the 1D problem. 110
4.3 The components of ε_{eff} as a function of particle angle for ellipsoidal particles with major and minor semiaxes of 0.25 and 0.15 and effective permittivity 2 suspended in an isotropic medium of permittivity 1. 112
4.4 (a) Effective permittivity as a function of particle permittivity. $\varepsilon_m = 1$, particle major and minor semiaxes 0.25 and 0.15, and $\theta = 0$. (b) Saturation permittivity ($\varepsilon_p = 100$) as a function of area fraction, with $\varepsilon_m = 1$ and particle aspect ratio set at 0.25:0.15. 113
4.5 Anisotropy in ε_{eff} as a function of particle aspect ratio with the particle major semiaxis at 0.25. $\varepsilon_m = 1$, $\varepsilon_p = 100$, $\theta = 0$ 113
4.6 The homogenised potential subject to a sinusoidal voltage modulation $\phi(x,1) = 2\sin(2\pi x)$ at the top plate. Calculated for particles at 60° with major and minor semiaxes equal to 0.25 and 0.15, $\varepsilon_p = 2$, and $\varepsilon_m = 1$. (a) Front view (b) Side (yz) view. 114
4.7 The microscopic potential, as calculated by Comsol, subject to a sinusoidal voltage modulation $\phi(x,1) = 3\sin(2\pi x)$ at the top plate. Calculated for a 5x5 array of particles at 60° with major and minor semi-axes equal to 0.25 and 0.15, $\varepsilon_p = 2$, and$\varepsilon_m = 1$. (a) Front view (b) Side (yz) view.115
4.8 (a) Comparison of the macroscopic and microscopic potential subject to a sinusoidal voltage modulation $\phi(x,1) = 3\sin(2\pi x)$ at the top plate. Calculated for a microscopic geometry of a 20 by 20 array of particles at 60° with major and minor semi-axes equal to 0.25 and 0.15, $\varepsilon_p = 2$, and $\varepsilon_m = 1$. (b) Difference in the macroscopic and microscopic solutions as a function of the microscopic array size, calculated with the infinity norm applied to the entire system. 115
4.9 Depiction of the elliptic coordinate system [58]. The curves of constant μ trace out ellipses with decreasing eccentricity, while the curves of constant ν trace out hyperbolas. F is the focal length of the coordinate system, with a larger F giving more eccentric ellipses. The unit vectors $\hat{\boldsymbol{e}}_\mu$ and $\hat{\boldsymbol{e}}_\nu$ are normal and tangential to the ellipses, respectively. 124
4.10 Analytic solution to the cell problem for non-rotated particles in isotropic medium. 127
4.11 Comparison of the analytic solution for a non-rotated particle in isotropic medium with the finite elements solution provided by Comsol. 127
4.12 The analytic solution to the first order cell problem for a particle at an angle of $\pi/8$, $\varepsilon_m = 1$, and $\varepsilon_p = 2$. 129
4.13 Norm of the difference between the analytic solution and Comsol solution for a particle at an angle of $\pi/8$, $\varepsilon_m = 1$, and $\varepsilon_p = 2$. 129
4.14 A diagram of the particle shape in the $\overline{\boldsymbol{y}}$ coordinate system. The major and minor semi-axes, $\boldsymbol{v}^{(a)}$ and $\boldsymbol{v}^{(b)}$, have lengths $\overline{a}$ and $\overline{b}$, and the particle angle is parameterised by θ. 138
4.15 Analytic solution for a metallic particle in an anisotropic medium with axis at $\pi/3$ to the y_1 axis. $a = 0.075$, $b = 0.05$, and $\varepsilon_m = \mathrm{R}(\pi/3)\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\mathrm{R}^T(\pi/3)$ .139
4.16 Norm of the difference between the analytic solution and the finite elements solution from Comsol. Solution is for a metallic particle in an anisotropic medium with axis at $\pi/3$ to the y_1 axis. $a = 0.075$, $b = 0.05$, and $\varepsilon_m = \mathrm{R}(\pi/3)\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\mathrm{R}^T(\pi/3)$. 140

4.17 Analytic solution for a dielectric particle in an anisotropic medium with axis at $\pi/3$ to the y_1 axis. $a = 0.075$, $b = 0.05$, and $\varepsilon_m = \mathrm{R}(\pi/3)\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\mathrm{R}^T(\pi/3)$. 142
4.18 Norm of the difference between the analytic solution and the finite elements solution from Comsol. Solution is for a dielectric particle in an anisotropic medium with axis at $\pi/3$ to the y_1 axis. $a = 0.075$, $b = 0.05$, and $\varepsilon_m = \mathrm{R}(\pi/3)\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\mathrm{R}^T(\pi/3)$ 142
4.19 (a) Diagram of the particle geometries (green ellipses) and the optical axis of the host medium (blue lines). (b) The percentage error between the homogenised and Comsol solutions for the electric potential. 145
4.20 Diagram depicting the prolate spheroidal coordinate system (b), formed by rotating the elliptic coordinate system (a) about the z axis. 152
4.21 The x-component of $\boldsymbol{\gamma}$ for prolate spheroidal, metallic particles aligned along the z axis, and a comparison with the numerical solution from Comsol, shown in the xy plane. The system is described by $\varepsilon_m = 1$, $e = 0.95$, and a volume fraction of 0.02. 157
4.22 The x-component of $\boldsymbol{\gamma}$ for prolate spheroidal, metallic particles aligned along the z axis, and a comparison with the numerical solution from Comsol, shown in the xz plane. The system is described by $\varepsilon_m = 1$, $e = 0.95$, and a volume fraction of 0.02. 157
4.23 The z-component of $\boldsymbol{\gamma}$ for prolate spheroidal, metallic particles aligned along the z axis, and a comparison with the numerical solution from Comsol, shown in the xz plane. The system is described by $\varepsilon_m = 1$, $e = 0.95$, and a volume fraction of 0.02. 157
4.24 Diagram depicting the oblate spheroidal coordinate system (b), formed by rotating the elliptic coordinate system (a) about the z axis. 158
4.25 The normalised error in γ_1 for metallic particles between the analytic, low volume fraction approximation and the numerical solution from Comsol, shown in the xy and xz planes. The system is described by $\varepsilon_m = 1$, $e = 0.95$, and a volume fraction of 0.02. 160
4.26 Diagram of the angles used to parameterise the major/optical axis, $\boldsymbol{a}$, of the spheroids. 163
4.27 Graph showing the behaviour of the depolarisation factors, L_z, with the particle eccentricity, e. 164
4.28 A comparison of the various effective medium theories for spherical particles in an isotropic medium 166
4.29 A comparison of the various permittivities as a function of volume fraction for prolate particles. The eccentricity is fixed at 0.8, along with $\varepsilon_p = 10$ and $\varepsilon_m = 1$. 168
4.30 A comparison of the various permittivities as a function of eccentricity for prolate particles. The volume fraction is fixed at 0.02, along with $\varepsilon_p = 10$ and $\varepsilon_m = 1$. 169
4.31 A comparison of the various permittivities as a function of eccentricity for prolate particles. The major semi-axis is fixed at 0.15, along with $\varepsilon_p = 10$ and $\varepsilon_m = 1$. 169

4.32 The normalised error between the linearly shifted γ_1 solution and the periodic solution from Comsol. As in figure 4.21, the system is described by $\varepsilon_p = 10$, $\varepsilon_m = 1$, $e = 0.95$, and a volume fraction of 0.02. 172
4.33 A comparison of the various effective permittivities for prolate particles in the principal frame of the effective medium (particles major axis along z), with e =0.8, $\varepsilon_p = 10$, and $\varepsilon_m = 1$. 174
4.34 A comparison of the various permittivities as a function of eccentricity for prolate particles. The volume fraction is fixed at 0.02, along with $\varepsilon_p = 10$ and $\varepsilon_m = 1$. 174
4.35 A comparison of the various permittivities as a function of eccentricity for prolate particles. The major semi-axis is fixed at 0.15, along with $\varepsilon_p = 10$ and $\varepsilon_m = 1$. 174
4.36 A comparison of the various effective permittivities for oblate particles in the principal frame of the effective medium (particles major axis along z), with e =0.8, $\varepsilon_p = 10$, and $\varepsilon_m = 1$. 175
4.37 A comparison of the various permittivities as a function of eccentricity for oblate particles. The volume fraction is fixed at 0.02, along with $\varepsilon_p = 10$ and $\varepsilon_m = 1$. 176
4.38 The percentage error between the homogenised potential and the microscopic potential for (a) disordered and (b) random particles. The spatial dependence of the system parameters is the same as in figure (4.19), though the particles are not on an ordered lattice this time around. . . . 177
4.39 Graph showing the average of the maximum normalised error between the homogenised and microscopic solutions for the potential as a function of N, the number of particles, for (a) disordered and (b) random distributions. The error bars show one standard deviation. 178
4.40 The average of the maximum normalised error between the homogenised and microscopic potentials as a function of volume fraction. The number of particles is fixed at 64. 179
4.41 Graphs showing the average, normalised error between the homogenised and experimental permittivities for N (a) disordered and (b) random, circular particles in an isotropic medium. The error bars show one standard deviation. 180
4.42 Graphs showing a comparison of the various effective permittivities for (a) disordered, and (b) random, circular particles in an isotropic medium. The error bars show one standard deviation in the experimental permittivity as calculated for N particles. 181

List of Tables

3.1 Expected transmitted intensity measured between parallel polarisers for a 6.5 µm thick, E7 LC cell in the twisted state 59
3.2 Experimental results for the measured diffraction efficiencies $\eta_{\perp}$ and $\eta_{\parallel}$, corresponding to probe polarisations perpendicular and parallel to the plane of incidence. $\Delta n_{\perp}$ and $\Delta n_{\parallel}$ are the calculated birefringence seen by the two polarisations [35]. 62

Declaration of Authorship

I declare that this thesis and the work presented in it is my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission

Signed:..... Date:.....

Acknowledgements

A thank you to EPSRC for funding this project and providing me the opportunity to study the subject I love. Also a thank you to Elena Perivolari for her experimental work on PAAD diffraction, and to my supervisors Malgosia Kaczmarek, Keith Daly, and, in particular, Giampaolo D'Alessandro, for his endless time and patience helping me through the last four years.

Chapter 1

Introduction

Liquid crystals are an intermediate state of matter which lies between a fluid and a solid crystal. Whilst able to flow like a liquid, they also exhibit optical properties similar to that of crystalline solids. These optical properties owe to the liquid crystal's molecular shape and ability (under the right conditions) to form a local alignment. In an aligned state, the liquid crystal is a birefringent material, acting on light as a solid crystal would. However, where the alignment and position are fixed in a solid crystal, a liquid crystal can flow and the alignment and optical axis can be controlled. This gives us the ability to create tunable systems which act on light as we desire. Liquid crystals have thus found uses in many devices in which the selective control of light transmission is desirable, such as wavelength specific filtering, modulation of the amplitude and phase of incident light, or even altering the direction of propagation [2].

In all of these devices it is through the careful control of the liquid crystal alignment that we are able to manipulate light in such a variety of ways. Since the advent of the technology, two original ideas have persisted; the predetermination of the surface alignments and the application of external fields to control the bulk alignment. What has changed and vastly improved is the level of precision with which we can control these factors. What began as binary on/off control of light transmission in very large pixels (calculator displays for instance) has now evolved into the ability to transmit variable intensities through millions of different coloured pixels many times per second, thus allowing full colour imaging such as in HDTVs and laptop screens [59]. Although having been vastly improved upon since their initial invention, devices such as these, where the active control is provided by an external field, suffer from some clear disadvantages. First, there is the requirement that continuous power must be applied to maintain an alignment. Second, it may not be practical, or even possible, to attach electrodes and apply an external voltage in some cases. Finally, it must be remembered that the surface alignment remains fixed. This limits the possible

configurations of the bulk alignment that can be induced by deforming the "default" alignment prescribed by the surface.

The alignment layers are commonly made by the mechanical rubbing of a suitable material, which, through surface interactions, induces a macroscopic alignment of the liquid crystal at the surfaces. Typically, this allows for an alignment only in one orientation, and with a pretilt angle of only a few degrees [22]. Through the elastic interactions of the liquid crystal molecules, the surface alignment induces a bulk alignment with the preferred optical properties. As a result of this, finding new ways of determining more accurate and complex alignments has been an area of great interest. The creation of microstructures in the alignment layers, for example, has been shown to allow much greater control over cell alignment, and hence the behaviour of the interacting light. A combination of alignment materials has been shown to produce alignment layers with controllable anchoring energies for both the polar and azimuthal angles [61], while the stacking of polymer structures on surface substrates has been shown to control pre-tilt angle over a virtually complete 90° range [18].

Even more interesting are the systems in which the alignment is tunable after the construction of the cell. Carbon nanotubes in an alignment layer, for instance, can be electrically controlled and used as optical diffusers or beam shapers [26]. It has also been shown that stacked polymer alignment layers can be tuned after construction of the cell, achieved by adding compounds to the liquid crystal that interact with the polymers and alter the alignment [17]. Of these tunable alignment layers, photosensitive alignment layers are particularly prominent and have been the topic of extensive research [19].

In particular, the use of photochromic molecules in alignment layers has long been demonstrated to offer reversible realignment of liquid crystal systems [9, 60]. This is achieved by exposing the photosensitive material to an electric field, causing a change in its chemical structure and orientation. The reorientation, in turn, causes the nearby liquid crystal molecules to reorient with it, providing the means to prescribe complex surface alignments for a liquid crystal cell both reversibly, and without the need for an applied field to the bulk of the cell. In many cases, once exposed to the appropriate light source and the alignment is set, the alignment will remain fixed with no need for continuous power. One particular class of materials, known as photoaligning azobenzene dyes (PAADs), have long demonstrated themselves to be capable of forming high resolution patterns [46, 45], enabling their use for polarisation gratings and spiral phase wave-plates, and to be effective in the alignment of liquid crystals [20, 60]. They also require low exposure energies, have an absorption spectrum extended past 400 nm, and their deposition methods onto substrates are very simple [40].

The utility of PAAD materials for complex liquid systems has made them an area of particular interest in our research group. One project has focused on the creation of optically controlled, dynamic half-wave plates. These are based on a twisted, nematic liquid crystal cell, and controlled using the photo-aligning properties of PAAD [43]. The results of these experiments presented some intriguing questions as to the exact nature and dynamics of the combined PAAD and liquid crystal photoalignment. In an effort to better understand this, a study of the thin PAAD layers in isolation was undertaken. This work considered the diffraction efficiency of photo-recorded refractive index gratings in standalone thin PAAD films [35]. Although there is a vast amount of research concerning the photoalignment of PAAD already ([11, 4, 49], to name but a few), this primarily concerns thicker samples than those used in the liquid crystal photoalignment and diffraction efficiency experiments. Under a periodic illumination, thicker samples usually exhibit surface relief which contributes to the formation of gratings [27, 14]. The thin samples in the diffraction efficiency experiments were found to have no surface relief, and thus the grating must be purely due to the periodic variation of the refractive index. However, under the assumption that the PAAD molecules reorient in the usual manner, the experimental results cannot be explained.

Thus we come to the first main section of work in this thesis, presented in part two. New experimental work was carried out by another member of the group, obtaining diffraction efficiency data over a wide range of incident angles, and for both polarisations of input beam. My work outlined in this thesis is the developement of a new model linking the alignment and birefringence of the aligned PAAD to the expected diffraction efficiencies for comparison with the experimental data.

This problem involves a variety of different topics. Firstly, it requires a study of propagation in layered media. The experimental system is made up from multiple layers of different materials, leading to resonance effects in the transmission that need to be accounted for. The process of fitting the theoretical model to the experimental results requires a substantial number of calculations that, owing to the number of unknowns in the experimental system, is far too computationally expensive and time consuming. To combat this, a new method for multi-layer propagation, known as the Iterated Ray method, is developed and implemented. It is found to have the accuracy and stability of the leading industry standard method, the S-matrix formulation, though proves to offer substantial increases in computational speed. The general method behind Iterated Ray formulation is intuitively extendable to systems with multiple coupled modes. This is used to model the transmission of light across a PAAD layer in its aligned anisotropic state where, like any crystal, liquid or solid, it supports two distinct modes of propagation. These are known as the ordinary and extraordinary modes, and their determination is another topic discussed. Finally, the diffraction process itself needs to be accounted for, resulting from the periodic

modulation of the field transmitted through the part aligned, part unaligned PAAD layer. Owing to the large angles of diffraction associated with the system, and the wide range of incident angles, a non-paraxial scalar diffraction theory is coupled to the rest of model.

The second section of work, presented in part three, concerns the modelling of doped liquid crystal systems, an area of research studying the addition of foreign materials to produce alterations to the properties of the liquid crystal. Examples include altering the boundary alignment [17], optical control of the clearing temperature [30], changes in the elastic and dielectric constants [44], and even the introduction of memory effects [1], to name but a few. The ability to not only control the alignment of the liquid crystal, but also its properties and how it behaves, is clearly an area of great interest. It gives us the ability to create even more complex systems to do things otherwise not possible, or to make existing systems easier or more efficient to operate; such as tuning a liquid crystal's responsiveness to external fields, for example.

Forming mathematical models of such doped systems is clearly an important part of understanding, constructing, and improving them. This is not a simple problem, however, as they often possess vastly differing length scales between their macroscopic and microscopic structures. This means that direct calculation of solutions is incredibly computationally expensive, or even infeasible. This is not, in actuality, a problem unique to doped liquid crystal systems, and the work and methods presented in this thesis are more of a study in multiscale modelling of anisotropic media. While studying liquid crystals we are essentially solving Laplace's equation on a multiscale geometry, but, owing to the fact that so many physical systems are governed by similar equations, it is easy to see that this work is analogous or easily extendable to many other problems such as diffusion and heat conduction.

Part three presents my work using the method of homogenisation to approximate the behaviour of these multiscale systems. Under the assumption of periodicity at the microscopic scale, the effects of the particle inclusions can be modelled as a small perturbation about a leading order field that varies only on the macroscopic scale. The perturbation is described by the microscopic behaviour of the system, which need only be considered once when solving the so called cell problem for a representative unit cell. The solution to cell problem then couples to the leading order field to give the final homogenised solution, described by an effective permittivity that describes the average behaviour of the composite system as if it were a single homogeneous material.

In this work, the homogenised equations for spatially varying dielectric particles in medium are first devoloped, and analytical approximations to the cell problem for low volume fractions of particles are then considered. First, the homogenised description of 2D elliptical particles in an anisotropic medium is found, followed by 3D spheroidal

particles in an isotropic medium. The final work in this thesis then relaxes the assumption of periodicity to better approximate real world systems, and the applicability of homogenisation for disorded particles is studied. Throughout, a comparison of the analytical and numerical solutions to the homogenised problem are made, along with a comparison to traditional theories used extensively in physics and engineering. All of the work in this thesis relies heavily on the theory of electromagnetism, and so we begin in part one with the necessary theory used to build the later models and understand the physical systems.

Chapter 2

Electromagnetic Theory

2.1 Maxwell's Equations

To model the PAAD diffraction grating and the doped liquid crystal systems, we need an understanding of how light propagates and interacts with matter, and we therefore now begin with the fundamental concepts that will be used throughout this work. The foundation of electromagnetism lies in Maxwell's equations which describe the electromagnetic fields arising from the presence of electric charges. In vacuum, these fields are comprised of two fundamental fields, the electric field, $\boldsymbol{E}$, and the magnetic field, $\boldsymbol{H}$, that are coupled together. However, in the presence of matter, we must also consider additional fields arising from the interaction of these electromagnetic fields with the electrically charged particles from which the material is made (discussed in the next section). These fields are the displacement field, $\boldsymbol{D}$, the magnetic induction, $\boldsymbol{B}$, and the electric current density $\boldsymbol{J}$. In SI units Maxwell's equations in matter are then [5, chapter 1]

$$\boldsymbol{\nabla} \cdot \boldsymbol{D} = \rho_f \tag{2.1a}$$

$$\boldsymbol{\nabla} \cdot \boldsymbol{B} = 0 \tag{2.1b}$$

$$\boldsymbol{\nabla} \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t} \tag{2.1c}$$

$$\boldsymbol{\nabla} \times \boldsymbol{H} = \boldsymbol{J} + \frac{\partial \boldsymbol{D}}{\partial t}, \tag{2.1d}$$

where ρ_f is the free charge density in the material and $\boldsymbol{J}$ the current density. To close the system of equations we must also add the constitutive relations

$$\boldsymbol{D} = \varepsilon \boldsymbol{E} \tag{2.2a}$$

$$\boldsymbol{B} = \mu \boldsymbol{H} \tag{2.2b}$$

$$\boldsymbol{J} = \sigma \boldsymbol{E}, \tag{2.2c}$$

where ε is the permittivity, μ the permeability, and σ the conductivity of the material. The materials involved in the grating system are all dielectric, meaning that they are non-conductive ($\sigma = 0$) and there are no currents. They are also non-magnetic and so their permeability is just equal to μ_0, the permeability of free space. The permittivity, from the point of view of an electromagnetic field, is all that distinguishes the different materials in the system from one another. For isotropic materials such as air and glass, the properties of the material are uniform in all directions throughout the medium and the permittivity is just a scalar. However, for birefringent materials such as PAAD, the permittivity seen by an electric field differs depending on its orientation. This is reflected in the permittivity becoming a rank two tensor, with its various components describing the permittivity in the different cartesian directions. The equations that form the basis of this work are then

$$\boldsymbol{\nabla} \cdot \boldsymbol{D} = 0 \tag{2.3a}$$

$$\boldsymbol{\nabla} \cdot \boldsymbol{H} = 0 \tag{2.3b}$$

$$\boldsymbol{\nabla} \times \boldsymbol{E} = -\mu_0 \frac{\partial \boldsymbol{H}}{\partial t} \tag{2.3c}$$

$$\boldsymbol{\nabla} \times \boldsymbol{H} = \frac{\partial \boldsymbol{D}}{\partial t}, \tag{2.3d}$$

We will now discuss the nature of the Displacement field, $\boldsymbol{D}$**,** and its relation (2.2a) with the electric field.

2.2 Polarisation and the D Field

The molecules from which different materials are made are themselves made from charged particles. These molecules and their constituent particles all act upon one another, either attracting or repelling, in such a way that a stable equilibrium is reached. This equilibrium can be disturbed if an external electric field is applied to the material. This is due to the field exerting a force on the charges, resulting in them being displaced them from their equilibrium positions. In the case of dielectrics, the charges associated with each molecule are bound and unable to move too far from their equilibrium positions, with the positive charges shifted somewhat in the direction of the electric field, and the negative charges in the direction opposite. Given a volume of material, one side will gain an excess of negative charge while the opposite side will gain an excess of positive charge, resulting in a net electric field.

This field is known as the polarisation and is given by [21]

$$\boldsymbol{P} = \varepsilon_0 \chi_e \boldsymbol{E}, \tag{2.4}$$

assuming that the material responds linearly to the external electric field. In the case of metals, charges are able to move freely, and so are able to reach an equillibrium position that gives a polarisation equal and opposite to the incident field, resulting in no field inside the metal.

The quantity χ_e is the electric susceptibility of the material. It describes the response of a dielectric to an electric field and is a rank two tensor. The components of this tensor are, in general, different from one another, and each represent the size of the field induced by an external field applied in the various cartesian directions. This is due to the fact that the constituent molecules of a material are asymmetric. The dipole moment (separation of opposite charges) of an individual molecule is given by

$$\boldsymbol{p} = \varepsilon_0 \alpha \boldsymbol{E}_{loc}, \tag{2.5}$$

where α is the polarisability tensor of the molecule and $\boldsymbol{E}_{loc}$ is the local, microscopic field seen by the molecule ($\boldsymbol{E}$ is the macroscopic field, defined as the average over a volume that is large with respect to the molecular scale). The total polarisation field is the sum of the dipoles from each of the N molecules,

$$\boldsymbol{P} = \sum_{i=1}^{N} \boldsymbol{p}_i. \tag{2.6}$$

If the molecules are randomly oriented, and a large enough volume is considered, this sum will always average to the same direction as the external field, $\boldsymbol{E}$ [13]. This means that χ is simply a scalar multiple of the identity and can be considered a scalar as in the previous section. If the molecules all share the same alignment, then the direction dependence of the polarisation of the volume will behave just as the polarisability of the individual molecules. If there is some other fixed order to the arrangement of molecules, i.e., the material is a crystal, the polarisation need not be the same as the molecular polarisability but it will still possess a directional dependence. The susceptibility χ is then, in general, represented by a tensor.

With this understanding of the collective material response to an applied field, we can now find the total field present in the system. It is the sum of the applied field and the materials response. This total field is the $\boldsymbol{D}$ field in the previous section, and we have that

$$\begin{aligned} \boldsymbol{D} &= \varepsilon_0 \boldsymbol{E} + \boldsymbol{P} \\ &= \varepsilon_0 \left(\mathcal{I} + \chi_e \right) \boldsymbol{E} \\ &= \varepsilon \boldsymbol{E}. \end{aligned} \tag{2.7}$$

This defines the permittivity

$$\varepsilon = \varepsilon_0 \varepsilon_r \tag{2.8}$$

and relative permittivity

$$\varepsilon_r = \mathcal{I} + \chi_e \tag{2.9}$$

of the material, which gives a mathematical description of the macroscopic behaviour of a material from the microscopic properties of its constituent molecules.

2.3 Electromagnetic Energy and the Permittivity Tensor

The principle of conservation of energy is an important aspect of modeling many aspects of electromagnetic fields. It is therefore important to understand and be able to describe the energy of an electromagnetic field, and its propagation. We begin by taking the scalar product of equation (2.1c) with $\boldsymbol{H}$ and the product of equation (2.1d) with $\boldsymbol{E}$, the difference of which gives

$$\boldsymbol{\nabla} \cdot (\boldsymbol{E} \times \boldsymbol{H}) + \boldsymbol{E} \cdot \frac{\partial \boldsymbol{D}}{\partial t} + \boldsymbol{H} \cdot \frac{\partial \boldsymbol{B}}{\partial t} + \boldsymbol{J} \cdot \boldsymbol{E} = 0 \tag{2.10}$$

applying the vector identity $\boldsymbol{\nabla} \cdot (\boldsymbol{A} \times \boldsymbol{B}) = \boldsymbol{A} \cdot \boldsymbol{\nabla} \times \boldsymbol{B} - \boldsymbol{B} \cdot \boldsymbol{\nabla} \times \boldsymbol{A}$. Integrating over an arbitrary volume and applying the Divergence Theorem

$$\int \left(\boldsymbol{E} \cdot \frac{\partial \boldsymbol{D}}{\partial t} + \boldsymbol{H} \cdot \frac{\partial \boldsymbol{B}}{\partial t} \right) dV + \int \boldsymbol{J} \cdot \boldsymbol{E} dV + \int (\boldsymbol{E} \times \boldsymbol{H}) \cdot \boldsymbol{n} dS = 0. \tag{2.11}$$

It can be shown [5] that the second term represents the sum of energy dissipation (in conducting media) and any mechanical work done by the field inside the volume, this term is zero in this work.

Looking at the third term, the units of $\boldsymbol{E}$ are Vm^{-1} whilst the units of $\boldsymbol{H}$ are Am^{-1}. Their product then has units of Wm^{-2} and this surface integral represents the total energy flux across the boundary of the volume per unit time. We can then define the Poynting vector

$$\boldsymbol{S} = \frac{1}{2} \boldsymbol{E} \times \boldsymbol{H} \tag{2.12}$$

which gives the instantaneous energy flux per area normal to the direction of $\boldsymbol{S}$. Sometimes the average energy flux per time, $\langle \boldsymbol{S} \rangle$, is of more interest. In this case, writing $\boldsymbol{E} = \tilde{\boldsymbol{E}} e^{-i\omega t}$ and $\boldsymbol{H} = \tilde{\boldsymbol{H}} e^{-i\omega t}$ and integrating $\boldsymbol{S}$ over a full period $T = 2\pi/\omega$ is gives

$$\langle \boldsymbol{S} \rangle = \frac{1}{2} \mathrm{Re} \left(\boldsymbol{E} \times \boldsymbol{H}^* \right), \tag{2.13}$$

where * denotes the complex conjugate.

We now look at the first term of equation (2.11). The units of the $\boldsymbol{D}$ and $\boldsymbol{B}$ fields are Cm^{-2} and $\mathrm{Nm}^{-1}\mathrm{A}^{-1}$, respectively. Both of the terms in the first integral then have units of Wm^{-3} and the volume integral has units of W. We see, then, that equation (2.11) must be an energy law describing the conservation of energy within the volume, and that the first term describes the rate of change of the total electric and magnetic energy in the volume. This means we can write the first integral as

$$\frac{\partial}{\partial t}\int w dV, \tag{2.14}$$

where w is the total energy density of the volume. Writing this as the sum of the electric and magnetic energy densities, w_e and w_m, and comparing to (2.11), we see

$$w_e = \frac{1}{2}\boldsymbol{E}\cdot\boldsymbol{D} \tag{2.15}$$

and

$$w_m = \frac{1}{2}\boldsymbol{H}\cdot\boldsymbol{B}, \tag{2.16}$$

subject to the condition that ε and μ are symmetric. The symmetry requirement is neccessary since

$$\begin{aligned}\boldsymbol{E}\cdot\frac{\partial \boldsymbol{D}}{\partial t} &= \frac{\partial}{\partial t}\frac{1}{2}\boldsymbol{E}\cdot\boldsymbol{D}\\ &= \frac{1}{2}\left(E_i\varepsilon_{ij}\frac{\partial E_j}{\partial t} + \frac{\partial E_i}{\partial t}\varepsilon_{ij}E_j\right)\end{aligned} \tag{2.17}$$

if and only if $\varepsilon_{ij} = \varepsilon_{ji}$.

Given this symmetry and assuming that ε is real (as we shall see in the next section, a complex permittivity/refractive index means that the material is absorbing) the Principal Axis Theorem tells us that this matrix can always be diagonalised, that is, there is a coordinate system in which

$$\begin{aligned}\varepsilon &= \varepsilon_0\begin{pmatrix}\varepsilon_x & 0 & 0\\ 0 & \varepsilon_y & 0\\ 0 & 0 & \varepsilon_z\end{pmatrix}\\ &= \varepsilon_0\begin{pmatrix}n_x^2 & 0 & 0\\ 0 & n_y^2 & 0\\ 0 & 0 & n_z^2\end{pmatrix}.\end{aligned} \tag{2.18}$$

The n_i are the refractive indices seen by an electric field parallel to the $\hat{\boldsymbol{e}}_i$ basis vector. In the case of uniaxial crystals (such as PAAD in its aligned state), two of these refractive indices are equal. This means that there is one special direction, known as the optical axis, for which an electric field that is parallel to the optical axis will experience what is known as the extraordinary refractive index. For electric fields

orthogonal to the optical axis, the so called ordinary refractive index is seen. Choosing coordinates in which the optical axis lies along the z axis we have

$$\varepsilon = \varepsilon_0 \begin{pmatrix} n_o^2 & 0 & 0 \\ 0 & n_o^2 & 0 \\ 0 & 0 & n_e^2 \end{pmatrix}. \tag{2.19}$$

These coordinates are known as the principal frame of the crystal. It should also be noted that this is not unique under rotations around the optical (z) axis, since the new x and y directions will remain orthogonal to the optical axis and still see n_o.

2.4 Wave Solutions to Maxwell's Equations

2.4.1 Propagation in Isotropic Media

We have seen how an electromagnetic field interacts with matter and how there is not only an energy stored in such a field, but that this energy also propagates. Looking at the energy law (2.11) we can clearly see that, if there is a non-zero flux and a movement of energy, there must also be a change in the fields themselves. We now seek solutions to Maxwell's equations to understand how light propagates in an infinite, homogeneous medium. First, we consider the case where ε is a scalar and the medium is isotropic. In this case equations (2.3a) and (2.3d) become simple equations for the electric field, $\boldsymbol{E}$, rather than $\boldsymbol{D}$. By taking the curl of equation (2.3c) and applying the vector identity $\boldsymbol{\nabla} \times \boldsymbol{\nabla} \times \boldsymbol{v} = \boldsymbol{\nabla}(\boldsymbol{\nabla} \cdot \boldsymbol{v}) - \nabla^2 \boldsymbol{v}$ we find

$$\boldsymbol{\nabla}(\boldsymbol{\nabla} \cdot \boldsymbol{E}) - \nabla^2 \boldsymbol{E} = -\mu_0 \frac{\partial}{\partial t} \boldsymbol{\nabla} \times \boldsymbol{H}. \tag{2.20}$$

Equation (2.3a) then tells us $\boldsymbol{\nabla} \cdot \boldsymbol{E} = 0$ (since ε is a constant scalar) and so by substituting in equation (2.3d) we arrive at

$$\frac{\partial^2 \boldsymbol{E}}{\partial t^2} - \frac{1}{\mu_0 \varepsilon} \boldsymbol{\nabla}^2 \boldsymbol{E} = 0. \tag{2.21}$$

By first taking the curl of equation (2.3d) it can be similarly obtained that

$$\frac{\partial^2 \boldsymbol{H}}{\partial t^2} - \frac{1}{\mu_0 \varepsilon} \boldsymbol{\nabla}^2 \boldsymbol{H} = 0. \tag{2.22}$$

These are both wave equations that describe the propagation of the electric and magnetic fields as waves traveling at a speed of

$$\sqrt{\mu_0 \varepsilon}^{-1} = c \sqrt{\varepsilon_r}^{-1} = \frac{c}{n}, \tag{2.23}$$

where $c = \sqrt{\mu_0 \varepsilon_0}^{-1}$ is the speed of light in vacuum, and $n = \sqrt{\varepsilon_r}$ is the refractive index of the material. We can separate the spatial and temporal parts of the solution by assuming $\boldsymbol{E} = \tilde{\boldsymbol{E}} e^{-i\omega t}$, giving

$$\boldsymbol{\nabla}^2 \tilde{\boldsymbol{E}} + \frac{n^2 \omega^2}{c^2} \tilde{\boldsymbol{E}} = 0 \tag{2.24}$$

for the spatial part of the electric field. The factor ω/c is equal to $k_0 = 2\pi/\lambda_0$, the wavenumber of light in vacuum. We define the wavenumber of light in a medium with refractive index n as

$$k = k_0 n \tag{2.25}$$

and it follows that the spatial dependence of our solution is governed by

$$\boldsymbol{\nabla}^2 \tilde{\boldsymbol{E}} + k^2 \tilde{\boldsymbol{E}} = 0. \tag{2.26}$$

This gives three separate, scalar equations, one for each of the components of $\tilde{\boldsymbol{E}}$, that are eigenvalue problems for the Laplacian. The independence of the different components means we need only consider fields in which the spatial dependence for all components is the same (up to a scale factor), i.e,

$$\tilde{\boldsymbol{E}} = \boldsymbol{E}_0 \psi(\boldsymbol{r}), \tag{2.27}$$

where $\boldsymbol{E}_0$ is a constant vector known as the polarisation of the field. We can construct any field we like from the sum of such fields and, provided they each satisfy Maxwell's equations, the sum will too (since Maxwell's equations are linear). The simplest solutions to (2.26) are given by the complex exponentials as

$$\begin{aligned} \psi(\boldsymbol{r}) &= e^{i\left(k_x x + k_y y + k_z z\right)} \\ &= e^{i\boldsymbol{k}\cdot\boldsymbol{r}}, \end{aligned} \tag{2.28}$$

subject to the dispersion relation

$$k_x^2 + k_y^2 + k_z^2 = k^2 \tag{2.29}$$

arising as a consequence of separating variables. We see, then, that there is an infinite number of solutions given by the different combinations of k_i that satisfy this relation. What is more, this infinite family of exponentials form an orthonormal basis of continuous, differentiable functions satisfying Maxwell's equations. Coupled with the linearity of Maxwell's equations, this means that for a given physical system, we can consider the behaviour of these basis solutions, known as plane waves, to construct the behaviour of any given field. This is the principle of Fourier theory used extensively in later sections of this work.

We look now at some properties of the solutions in equation (2.28). By writing the position vector $\boldsymbol{r}$ in terms of its components parallel and perpendicular to $\boldsymbol{k}$, we see that

$$\psi(\boldsymbol{r}) = e^{i\boldsymbol{k}\cdot\left[\left(\hat{\boldsymbol{k}}\cdot\boldsymbol{r}\right)\hat{\boldsymbol{k}}+\left(\hat{\boldsymbol{k}}_\perp\cdot\boldsymbol{r}\right)\hat{\boldsymbol{k}}_\perp\right]} \tag{2.30}$$

$$= e^{ikr\cos\theta}, \tag{2.31}$$

where θ is the angle between $\boldsymbol{k}$ and $\boldsymbol{r}$. The full solution is then

$$E_i = e^{i(kr\cos\theta-\omega t)} \tag{2.32}$$

and represents a wave in space, with the surfaces of constant phase propagating in the direction of $\boldsymbol{k}$, known as the wavevector, and at a speed of $\omega/k = c/n$, as before. We note here that the factor $k = k_0 n$ may be complex (since n may be), in which case

$$E_i = e^{ik_0(\mathrm{Re}(n)r\cos\theta-\omega t)}e^{-\mathrm{Im}(n)k_0 r\cos\theta} \tag{2.33}$$

and the wave decays as it propagates. In this way we see that a complex permittivity can be used to account for absorption of energy by a material.

We find another important property by substituting equation (2.28) into equation (2.3a). This gives

$$k_i\tilde{E}_i = \boldsymbol{k}\cdot\tilde{\boldsymbol{E}}_i = 0, \tag{2.34}$$

which tells us that the field and its wavevector are orthogonal, i.e., that light travels as a transverse wave.

Turning our attention to the magnetic field we see from equations (2.3c), (2.27), and (2.28) that

$$\boldsymbol{k}\times\boldsymbol{E} = \mu_0\omega\boldsymbol{H}. \tag{2.35}$$

This tells us that the magnetic field oscillates in phase with the electric field and is perpendicular to both the electric field and the direction of propagation. Taking the absolute value of both sides, rearranging and using the wave speed definition (2.23) we obtain

$$Z = \frac{|\boldsymbol{E}|}{|\boldsymbol{H}|} = \sqrt{\frac{\mu_0}{\varepsilon}}. \tag{2.36}$$

This is known as the wave impedance and gives the relative magnitudes of the electric and magnetic fields. Now using equations (2.12) and (2.13) we can find the instantaneous and time averaged energy flux associated with the plane wave as

$$\boldsymbol{S} = \frac{\boldsymbol{E}\cdot\boldsymbol{E}}{2Z}\hat{\boldsymbol{k}}. \tag{2.37}$$

and

$$\begin{aligned}\langle \boldsymbol{S} \rangle &= \frac{|\boldsymbol{E}|^2}{2Z} \mathrm{Re}\left(\hat{\boldsymbol{E}} \times \hat{\boldsymbol{H}}*\right) \\ &= \frac{|\boldsymbol{E}|^2}{2Z} \hat{\boldsymbol{k}}. \end{aligned} \tag{2.38}$$

There are no other restrictions on the fields. For any given direction $\boldsymbol{k}$, the fields $\boldsymbol{E}$ and $\boldsymbol{H}$ exist as an orthogonal pair of waves at any orientation in the plane perpendicular to $\boldsymbol{k}$, and propagating at speed c/n. This simplicity, however, is due to the assumption that the medium is isotropic and ε is a scalar. As we shall see in the next section, things become much more complicated inside birefringent media such as PAAD.

2.4.2 Propagation in Anisotropic Media

2.4.2.1 O- and E-Wave Solutions

We now look for solutions to Maxwell's equation inside an infinite, homogeneous, uniaxial material [21]. Once again we combine equations 2.3c and 2.3d to get

$$\boldsymbol{\nabla} \times \boldsymbol{\nabla} \times \boldsymbol{E} = -\mu_0 \frac{\partial^2 \boldsymbol{D}}{\partial t^2} = -\mu_0 \varepsilon \frac{\partial^2 \boldsymbol{E}}{\partial t^2}, \tag{2.39}$$

remembering that this time ε is a tensor. Owing to the results of the previous section, we make the ansatz that

$$\boldsymbol{D} = \varepsilon \boldsymbol{E} = \varepsilon \hat{\boldsymbol{E}} e^{i(\boldsymbol{k}\cdot\boldsymbol{r} - \omega t)} \tag{2.40}$$

and we find that

$$-\boldsymbol{k} \times \boldsymbol{k} \times \boldsymbol{E} = \frac{k_0^2}{\varepsilon_0} \boldsymbol{D} = k_0^2 \varepsilon_r \boldsymbol{E}. \tag{2.41}$$

This is still a wave equation, though now the components of $\boldsymbol{E}$ are coupled together on the right hand side by the multiplication with ε, making the problem much more complicated. We now face a choice of whether to solve for $\boldsymbol{E}$, and have a complicated right hand side, or to solve for $\boldsymbol{D}$, and have a complicated left hand side. We choose to study the nature of the $\boldsymbol{D}$ field. This is due to the important property given by immediately by the first Maxwell equation (2.3a), namely, that

$$\boldsymbol{k} \cdot \boldsymbol{D} = 0 \tag{2.42}$$

and that the $\boldsymbol{D}$ field is orthogonal to the direction of propagation ($\boldsymbol{E}$ is, in general, not parallel to $\boldsymbol{D}$ or orthogonal to $\boldsymbol{k}$).

We rewrite equation (2.41) in the form

$$\left[\left(\boldsymbol{k}\otimes\boldsymbol{k}-k^2\mathcal{I}\right)\varepsilon_r^{-1}+k_0^2\mathcal{I}\right]\boldsymbol{D}=0, \tag{2.43}$$

where $\mathcal{I}$ is the identity matrix. In this way, we see that this is just an eigenvalue problem. The solution to this system of equations first lies in finding $\boldsymbol{k}$ such that the determinant of the term in square brackets is zero. For simplicity, we will work in the principal frame of the crystal where the optical axis is aligned with the z axis and ε_r is diagonal, greatly simplifying calculation. The determinant is then

$$\begin{vmatrix} \left(k_x^2-k^2\right)n_o^{-2}+k_0^2 & k_xk_yn_o^{-2} & k_xk_zn_e^{-2} \\ k_xk_yn_o^{-2} & \left(k_y^2-k^2\right)n_o^{-2}+k_0^2 & k_yk_zn_e^{-2} \\ k_xk_zn_o^{-2} & k_yk_zn_o^{-2} & \left(k_z^2-k^2\right)n_e^{-2}+k_0^2 \end{vmatrix} \tag{2.44}$$

which, upon setting equal to zero and simplifying, yields the equation

$$\left(k_0^2-k^2n_o^{-2}\right)\left(k_0^2-\left(k_x^2+k_y^2\right)n_e^{-2}-k_z^2n_o^{-2}\right)=0. \tag{2.45}$$

This gives two possible solutions, either

$$k^2=k_0^2n_o^2 \tag{2.46}$$

or

$$\frac{k_x^2+k_y^2}{n_e^2}+\frac{k_z^2}{n_o^2}=k_0^2. \tag{2.47}$$

The first solution says that, regardless of the direction of propagation, there is a wave which propagates with speed c/n_o. This is known as the ordinary wave, or o-wave. The second solution corresponds to the extraordinary wave (or e-wave) whose speed depends on its direction of propagation. We can describe the wavevector by its angle θ with the z axis and polar angle ϕ in the xy plane by

$$k_x=k\sin\theta\cos\phi \tag{2.48}$$

$$k_y=k\sin\theta\sin\phi \tag{2.49}$$

$$k_z=k\cos\theta. \tag{2.50}$$

Substituting this into (2.47) we find

$$k=\frac{n_en_o}{\sqrt{n_e^2\cos^2\theta+n_o^2\sin^2\theta}}k_0 \tag{2.51}$$

$$=n_{\text{eff}}\left(\theta\right)k_0, \tag{2.52}$$

which defines the effective refractive index n_{eff}. The extraordinary wave, then, travels at a speed of $c/n_{\text{eff}}\left(\theta\right)$, which changes depending on the angle made between the

wavevector and optical axis of the crystal.

2.4.2.2 Field Orientations

We now return to the wave equation (2.41) to determine the direction of the $\boldsymbol{D}$ field associated with each of these waves. As mentioned in the previous section, the diagonal form of the permittivity tensor in its principal frame is maintained under rotations around the optical axis. We use this to rotate our coordinates such that the wavevector lies in the xz plane and $k_y = 0$; the left hand side becomes

$$-\boldsymbol{k}\times\boldsymbol{k}\times\left(\varepsilon^{-1}\boldsymbol{D}\right) = \frac{-1}{\varepsilon_0}\begin{pmatrix} -k_z^2 n_o^{-2} D_x + k_x k_z n_e^{-2} D_z \\ -k^2 n_o^{-2} D_y \\ k_x k_z n_o^{-2} D_x - k_x^2 n_e^{-2} D_z \end{pmatrix}. \tag{2.53}$$

By equation (2.42)

$$k_x D_x = -k_z D_z \tag{2.54}$$

and this can be rewritten as

$$-\boldsymbol{k}\times\boldsymbol{k}\times\left(\varepsilon^{-1}\boldsymbol{D}\right) = \frac{1}{\varepsilon_0}\begin{pmatrix} \left(k_x^2 n_e^{-2} + k_z^2 n_o^{-2}\right) D_x \\ k^2 n_o^{-2} D_y \\ \left(k_x^2 n_e^{-2} + k_z n_o^{-2}\right) D_z \end{pmatrix}. \tag{2.55}$$

We first consider the o-wave, where substitution of (2.46) and rearranging gives

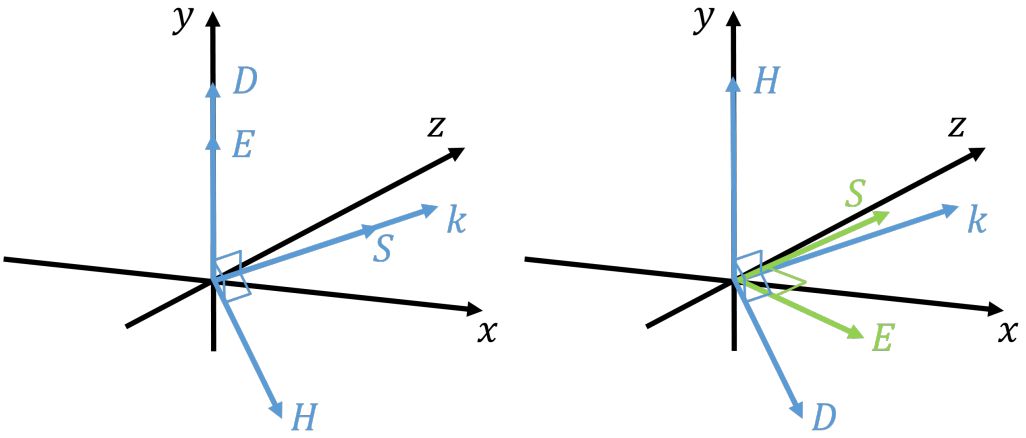
$$\begin{aligned} 0 &= \left(k_x^2 n_e^{-2} + k_z n_o^{-2} - k^2 n_o^{-2}\right) D_{x,z} \\ &= k_x^2 \left(n_e^{-2} - n_o^{-2}\right) D_{x,z}. \end{aligned} \tag{2.56}$$

This presents two possibilities for the o-wave. First, if $k_x = 0$ and the direction of propagation is along the optical axis, this relation is always satisfied and $\boldsymbol{D}_o$ can lie anywhere in the xy plane (remembering that (2.42) must still hold). The other option is that $k_x \neq 0$ and so we must have that $D_x = D_z = 0$ and $\boldsymbol{D}_o$ is parallel to the y axis, perpendicular to both $\boldsymbol{k}$ and the optical axis. In either instance, multiplying by ε^{-1} simply rescales $\boldsymbol{D}_o$ and we find that $\boldsymbol{E}_o$ and $\boldsymbol{D}_o$ are parallel. From equation (2.3d) we see that $\boldsymbol{H}_o$ is perpendicular to both $\boldsymbol{k}$ and $\boldsymbol{E}_o/\boldsymbol{D}_o$, lying in the xz plane. The orientations of the various fields are shown in figure 2.1a.

We now consider the e-wave. Substituting equation (2.47) and $k_y = 0$ into (2.55) gives

$$k^2 n_o^{-2} D_y = D_y. \tag{2.57}$$

We then must have that $D_y = 0$ and $\boldsymbol{D}_e$ lies in the plane of incidence, perpendicular to both $\boldsymbol{k}$ and $\boldsymbol{D}_o$. This time, since $\varepsilon_{11}^{-1} \neq \varepsilon_{33}^{-1}$, $\boldsymbol{E}_e$ and $\boldsymbol{D}_e$ are not parallel. Both are, again,



(A) (B)

FIGURE 2.1: (a) Diagram showing the orientation of fields associated with the o-wave. (b) Diagram showing the orientation of the fields associated with the e-wave.

perpendicular to $\boldsymbol{H}_e$ (which points in the y direction) by equations (2.3c) and (2.3d). This is depicted in figure (2.1b).

Another interesting property of the extraordinary wave is that, unlike the o-wave (and propagation in isotropic media), the direction of energy transfer, $\boldsymbol{S}$, is not in the direction of wave propagation, $\boldsymbol{k}$. By definition of the Poynting vector, equation (2.12), this direction is perpendicular to $\boldsymbol{H}$, and so lies in the plane of $\boldsymbol{k}$ and the optical axis. It is also perpendicular to $\boldsymbol{E}$, which is not perpendicular to $\boldsymbol{k}$, and so $\boldsymbol{S}$ must make an oblique angle with $\boldsymbol{k}$ in the plane. Denoting by ϕ the angle between $\boldsymbol{S}$ and the optical axis,

$$\tan\phi = \frac{S_x}{S_z} = -\frac{E_z}{E_x}, \tag{2.58}$$

using $\boldsymbol{S} \cdot \boldsymbol{E} = 0$. In a similar way,

$$\tan\theta = -\frac{D_z}{D_x} = -\frac{n_e^2}{n_o^2}\frac{E_z}{E_x} \tag{2.59}$$

and we find that

$$\tan\phi = \frac{n_o^2}{n_e^2}\tan\theta. \tag{2.60}$$

All the properties until now have all been derived assuming that the wave exists inside an infinite homogeneous medium. When we consider systems made up from different materials, there will be discontinuities in the material properties appearing in Maxwell's equations. This will, in turn, cause a discontinuity in the fields described by the equations. We will now consider the behaviour of the fields at such a discontinuity and derive relations to relate the fields to either side.

2.5 Fields at a Material Discontinuity

2.5.1 Interface Boundary Conditions

There are four conditions to derive, one from each of Maxwell's equations. They are found by considering the integral form of Maxwell's equations [5]:

$$\int \boldsymbol{D} \cdot d\boldsymbol{S} = \int \rho_f dV \tag{2.61}$$

$$\int \boldsymbol{B} \cdot d\boldsymbol{S} = 0 \tag{2.62}$$

$$\oint \boldsymbol{E} \cdot d\boldsymbol{l} = -\frac{d}{dt} \int \boldsymbol{B} \cdot d\boldsymbol{S} \tag{2.63}$$

$$\oint \boldsymbol{H} \cdot d\boldsymbol{l} = \int \boldsymbol{J} + \frac{d}{dt} \boldsymbol{D} \cdot d\boldsymbol{S}, \tag{2.64}$$

valid for any closed volume V bounded by S, and any closed loop l bounding a surface S. First we consider (2.62), and take the volume of integration (bounded by S) to be a cylinder of height δh and cross sectional area δA. The cross sections lie parallel to the boundary with the ends at $\pm\frac{\delta h}{2}$ to either side of the boundary. In the limit of $\delta h \to 0$ the portion of the integral over the sides of the cylinder vanish, leaving

$$\int \boldsymbol{n} \cdot \left(\boldsymbol{B}^+ - \boldsymbol{B}^-\right) \delta A = 0, \tag{2.65}$$

where $\boldsymbol{B}^+$ and $\boldsymbol{B}^-$ denote the fields on either side of the boundary and $\boldsymbol{n}$ is the normal to material boundary. This is true for arbitrary δA and so must hold pointwise, i.e.,

$$\boldsymbol{n} \cdot \left(\boldsymbol{B}^+ - \boldsymbol{B}^-\right) = 0. \tag{2.66}$$

This is our first boundary condition which says that the normal component of $\boldsymbol{B}$ must be continuous across a boundary.

Still considering the same volume of integration, we can apply a similar argument to the first Maxwell equation (2.61). Things are a little more complicated this time, however, since as δh and $V \to 0$, the charge density ρ_f becomes undefined. We avoid this problem as follows. Letting the total charge be Q, we have

$$\rho_f = \frac{dQ}{dV}, \tag{2.67}$$

and we can define the surface charge density by

$$\sigma_f = \frac{dQ}{dS}. \tag{2.68}$$

It follows that $\rho_f dV = \sigma_f dS$ and equation (2.61) can be written as

$$\int \boldsymbol{D} \cdot d\boldsymbol{S} = \int \sigma_f dS. \tag{2.69}$$

Now as we let $\delta h \to 0$, we obtain our second boundary condition

$$\boldsymbol{n} \cdot \left(\boldsymbol{D}^+ - \boldsymbol{D}^-\right) = \sigma_f. \tag{2.70}$$

This says that there is a jump in the normal component of the $\boldsymbol{D}$ field equal to the surface charge density. In the systems modelled in this work there are no surface charges present, and so we also just have that the normal component of $\boldsymbol{D}$ is continuous.

For the third and fourth Maxwell equations, (2.63) and (2.64), we consider a surface of integration that is a rectangle with sides of height δh, and top and bottom edges of width δw. The sides are perpendicular to the boundary, and the top and bottom edges lie at $\pm\frac{\delta h}{2}$ above and below the boundary. In the limit of $\delta h \to 0$ the surface integrals also go to zero, and we are just left with the line integrals parallel to the boundary. Calling the normal to the rectangular surface $\boldsymbol{n}_R$, the remaining line integrals are parameterised by their tangent vector $\boldsymbol{t} = \boldsymbol{n}_R \times \boldsymbol{n}$. This gives, for example,

$$0 = \int \left(\boldsymbol{E}^+ - \boldsymbol{E}^-\right) \cdot d\boldsymbol{t} \tag{2.71}$$

$$= \int \boldsymbol{n}_R \cdot \left[\boldsymbol{n} \times \left(\boldsymbol{E}^+ - \boldsymbol{E}^-\right)\right] dl, \tag{2.72}$$

where $\boldsymbol{E}^+$ and $\boldsymbol{E}^-$ denote the fields on either side of the boundary. This must hold for all $\boldsymbol{n}_R$ (rotations of the rectangle around $\boldsymbol{n}$), and so

$$\boldsymbol{n} \times \left(\boldsymbol{E}^+ - \boldsymbol{E}^-\right) = 0. \tag{2.73}$$

This is our third boundary condition, the tangential components of the electric field must be continuous. Applying the same integration to equation (2.64) we find, analogously to the boundary condition for $\boldsymbol{D}$, that

$$\boldsymbol{n} \times \left(\boldsymbol{H}^+ - \boldsymbol{H}^-\right) = \boldsymbol{J}_s, \tag{2.74}$$

where $\boldsymbol{J}_s$ is the surface current density. This final boundary condition says that the tangential components of $\boldsymbol{H}$ have a jump equal to the surface current density. There are no currents in the systems in this work and so we have the condition that the tangential components of $\boldsymbol{H}$ are continuous.

2.5.2 S- and P-Polarised Waves

We have seen how in an anisotropic medium, for a given direction of propagation, only o- and e-waves can propagate, restricting the number of orientations of the waves polarisation to two. However, for propagation in an isotropic medium, the orientation of the $\boldsymbol{D}$ and $\boldsymbol{E}$ fields can be anywhere in the plane perpendicular to their wavevector. These possible orientations form a 2D vector space which can be described by a basis of two vectors. If we can determine the behaviour of these basis vectors at an interface, we can study any incident wave by decomposing onto this basis and summing the behaviour of the components.

To define the basis we consider a plane wave, described by its wavevector $\boldsymbol{k}^{(i)}$, that is incident upon a boundary, described by its normal $\boldsymbol{n}$, between two media. The plane that is spanned by $\boldsymbol{k}^{(i)}$ and $\boldsymbol{n}$ is known as the plane of incidence, and can be described by its unit normal

$$\boldsymbol{N} = \hat{\boldsymbol{k}}^{(i)} \times \boldsymbol{n}. \tag{2.75}$$

From this we now define two orthogonal polarisations of wave, known as the s- and p-polarisations, that will form our basis. The electric field of the s-polarised wave satisfies

$$\boldsymbol{E}_s \times \boldsymbol{N} = 0, \tag{2.76}$$

i.e., it is normal to the plane of incidence and parallel to the material boundary. The p-polarised wave satisfies

$$\boldsymbol{E}_p \cdot \boldsymbol{N} = 0, \tag{2.77}$$

i.e., it is in the plane of incidence and, in general, forms an oblique angle with the material boundary. We also define the angle made between $\boldsymbol{k}^{(i)}$ and the surface normal, known as the angle of incidence θ, which satisfies

$$\hat{\boldsymbol{k}}^{(i)} \cdot \boldsymbol{n} = \cos\theta. \tag{2.78}$$

These are all depicted in figure 2.2.

There are two important things to note. The first is that the definition of both polarisations remains satisfied under a change of sign. Careful attention then needs to be paid as to which orientation is being used since a reversal of sign equates to a phase shift of π. Second, at normal incidence the two definitions of polarisation are not unique and coincide with one another. We therefore assert that, at normal incidence, the s- and p-polarisations are such that they agree with the above definitions in the limit that the angle of incidence goes to zero.

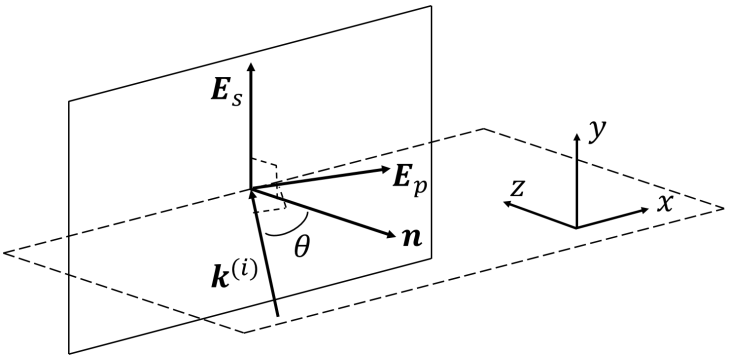


FIGURE 2.2: A diagram of the basis vectors $\boldsymbol{E}_s$ and $\boldsymbol{E}_p$ for a wave incident with wavevector $\boldsymbol{k}^{(i)}$ on a boundary with normal $\boldsymbol{n}$.

2.5.3 Reflection and Refraction in Isotropic Media

We are now in a position to describe the outgoing waves that result from an incoming wave, described by wavevector $\boldsymbol{k}^{(i)}$, that is incident upon a boundary between two isotropic media. We define our coordinate system as in figure (2.2). The $z = 0$ plane is taken to lie at the material boundary and we rotate our coordinates such that the plane of incidence lies in the xz plane. We also assume that the incident wave is s-polarised and that it is the only field in the region $z < 0$.

The incident wave is described by

$$\boldsymbol{E}^{(i)} = \left(0, E_y^{(i)}, 0\right) \tag{2.79}$$

and, for the tangential components of $\boldsymbol{E}$ to be continuous, we must have a transmitted field, where $z > 0$, given by

$$\boldsymbol{E}^{(t)} = \left(0, E_y^{(i)}, E_z^{(t)}\right). \tag{2.80}$$

We will show that the field on the incident side of the boundary, $z < 0$, must also contain a reflected wave.

Assuming only an incident and transmitted field, we must have that the phase change of $\boldsymbol{E}^{(t)}$ in the x and y directions is the same as in $\boldsymbol{E}^{(i)}$ (for continuity of the tangential components everywhere on the boundary). To satisfy Maxwell's equations and the dispersion relation (2.29), we must then have that

$$\boldsymbol{E}^{(t)} = \hat{\boldsymbol{E}}^{(t)} e^{i\left(k_x^{(i)} x + \sqrt{k^{(t)^2} - k_x^{(i)^2}} z\right)}, \tag{2.81}$$

and we see that the transmitted field is also a plane wave. Note that this wave must be forward propagating because it can only take energy away from the boundary.

The z component of $\boldsymbol{E}^{(t)}$ is found by applying the continuity of the normal component of $\boldsymbol{D}$. We are dealing with isotropic media, where $\boldsymbol{E}$ and $\boldsymbol{D}$ are parallel, and it follows

that $E_z^{(t)} = 0$ and both $\boldsymbol{E}$ fields are the same. By equation (2.13) the ratio of the average energy flux into and away from the boundary at any point is then $Z_2/Z_1 = n_1/n_2 \neq 1$ (where n_i is the refractive index of medium i). By the conservation of energy this cannot be possible and necessitates the existence of another wave. The only allowed wave that can be transmitted into the second medium is given by equation (2.81), and so, by necessity, we must have a reflected wave,

$$\boldsymbol{E}^{(r)} = \hat{\boldsymbol{E}}^{(r)} e^{i\left(k_x^{(i)} x - k_z^{(i)} z\right)}, \tag{2.82}$$

to account for this energy imbalance.

We see that for the waves propagating away from the boundary the component of the wavevectors that are tangential to the boundary must be the same as for the incident wave. Only the normal component changes on account of the change in wavenumber and the dispersion relation (2.29). All waves must propagate in the plane of incidence and so we can describe the waves in terms of the angles their wavevectors make with the boundary normal, i.e.,

$$\boldsymbol{k}^{(i)} = k^{(i)} \left(-\sin\theta_i, 0, \cos\theta_i\right), \tag{2.83}$$

$$\boldsymbol{k}^{(r)} = k^{(i)} \left(\sin\theta_r, 0, -\cos\theta_r\right), \tag{2.84}$$

$$\boldsymbol{k}^{(t)} = k^{(i)} \left(-\sin\theta_t, 0, \cos\theta_t\right). \tag{2.85}$$

It follows that for the reflected wave

$$\theta_r = -\theta_i, \tag{2.86}$$

whilst for the transmitted wave

$$n_1 \sin\theta_i = n_2 \sin\theta_t, \tag{2.87}$$

which is known as Snell's Law. We see that, as a consequence of the wavelength changing between media, the direction of propagation must also change across a boundary so that the phase of the fields can be continuous, in a process known as refraction.

Given the directions of propagation, we can now determine the orientation of the outgoing electric fields. If we suppose that the transmitted and reflected fields have non-zero x and z components, we immediately obtain

$$E_x^{(r)} = E_x^{(t)} \tag{2.88}$$

and

$$\frac{E_z^{(t)}}{E_z^{(r)}} = \frac{n_2^2}{n_1^2} \tag{2.89}$$

from our continuity boundary conditions. Multiplying (2.88) by $k_x^{(r)}k_x^{(t)}$ and using that $\boldsymbol{k}\cdot\boldsymbol{E}=0$ we get

$$\frac{E_z^{(t)}}{E_z^{(r)}}=\frac{k_z^{(r)}k_x^{(t)}}{k_z^{(t)}k_x^{(r)}}. \tag{2.90}$$

Clearly these equations cannot agree in general and so both the transmitted and reflected wave can only be polarised in the $\pm y$ direction only, that is, they are polarised along the same axis as the incident wave. For the case of an incident p-polarised wave the same results are obtained by following the same steps considering the $\boldsymbol{H}$ field, instead.

2.5.4 The Fresnel Amplitude Coefficients in Isotropic Media

For a plane wave incident on a boundary we now know the orientation and direction of propagation of the outgoing waves. We can now use this to find the amplitude of the outgoing waves in relation to the amplitude of the incoming. We seek to write

$$\left|\boldsymbol{E}^{(r)}\right|=r\left|\boldsymbol{E}^{(i)}\right| \tag{2.91}$$

and

$$\left|\boldsymbol{E}^{(t)}\right|=t\left|\boldsymbol{E}^{(i)}\right|. \tag{2.92}$$

The coefficients r and t are known as the Fresnel amplitude coefficients. We first consider an incoming s-polarised wave. The $\boldsymbol{H}$ fields are in the plane of incidence perpendicular to $\boldsymbol{k}$, giving

$$\boldsymbol{H}=\frac{|\boldsymbol{E}|}{Z}\left(\hat{k}_z,0,-\hat{k}_x\right). \tag{2.93}$$

The continuity of the tangential component of $\boldsymbol{E}$ and $\boldsymbol{H}$ gives

$$1+r_s=t_s \tag{2.94}$$

and

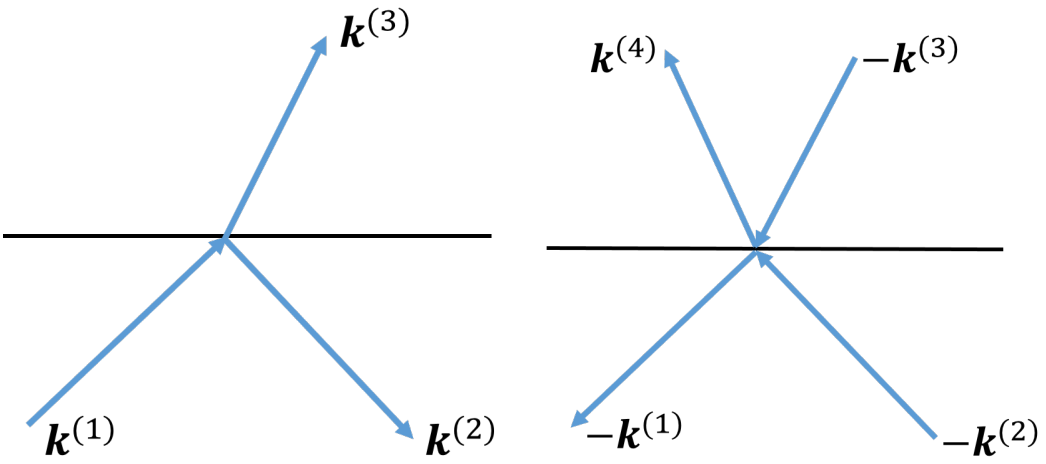
$$n_1\left(\hat{k}_z^{(i)}-r_s\hat{k}_z^{(r)}\right)=n_2t_s\hat{k}_z^{(t)}, \tag{2.95}$$

respectively. These simultaneous equations can be solved to give

$$r_s=\frac{q_i-q_t}{q_i+q_t}=\frac{n_1\cos\theta_i-n_2\cos\theta_t}{n_1\cos\theta_i+n_2\cos\theta_t}, \tag{2.96a}$$

$$t_s=\frac{2q_i}{q_i+q_t}=\frac{2n_1\cos\theta_i}{n_1\cos\theta_i+n_2\cos\theta_t}. \tag{2.96b}$$

The q_α terms are the normal component of the wavevector $\boldsymbol{k}^{(\alpha)}$, and this definition will be used consistently throughout this work. Repeating the same analysis for the



(A) (B)

FIGURE 2.3: (a) Wavevectors associated with the forward process of reflection and refraction. (b) The reverse process considered in the derivation of the Stokes' relations.

p-polarised wave gives

$$r_p = \frac{n_2^2 q_i - n_1^2 q_t}{n_2^2 q_2^{(i)} + n_1^2 q^{(t)}} = \frac{n_2 \cos\theta_i - n_1 \cos\theta_t}{n_2 \cos\theta_i + n_1 \cos\theta_t}, \tag{2.96c}$$

$$t_p = \frac{2n_1 n_2 q_i}{n_2^2 q_i + n_1^2 q_t} = \frac{2n_1 \cos\theta_i}{n_2 \cos\theta_i + n_1 \cos\theta_t}. \tag{2.96d}$$

These relations assume that the positive orientation of the p-polarisation always points to the right of the wavevector as you are looking along it. That is, as in figure 2.2,

$$\boldsymbol{E}_p = \hat{\boldsymbol{k}} \times \hat{\boldsymbol{e}}_y, \tag{2.97}$$

and the z component of $\boldsymbol{E}_p$ always points towards the same material before and after reflection and transmission. Some authors define the p-polarisation such that its x component always have the same sign, which has the effect of reversing the sign of r_p (as defined in (2.96c)). We choose the former convention so that in later sections where we consider forwards and backwards waves in multi-layered systems, we have a consistent definition of positive orientation, thus avoiding accidental π phase shifts.

In the case that there is no absorption, the process is time reversible and we can derive the so called Stokes' relations between between the coefficients associated with a boundary. For the forward process we consider an incident wave along $\boldsymbol{k}^{(1)}$, a reflected wave along $\boldsymbol{k}^{(2)}$, and a transmitted wave along $\boldsymbol{k}^{(3)}$, shown in figure (2.3a). For the backward process we would then have two incident waves, along $-\boldsymbol{k}^{(2)}$ and $-\boldsymbol{k}^{(3)}$, two reflected waves, along $-\boldsymbol{k}^{(1)}$ and $\boldsymbol{k}^{(4)}$, and two transmitted waves, also along $-\boldsymbol{k}^{(1)}$ and $\boldsymbol{k}^{(4)}$, all shown in figure (2.3b). Now, for the process to be reversible, we need the backward process to agree with the time-reversal of the forward process. This means that the sum of the waves travelling along $-\boldsymbol{k}^{(1)}$ in the backward process

must be equal to E_0, the magnitude of the incident wave in the forward process, that is,

$$r_{12}\left(r_{12}E_0\right) + t_{21}\left(t_{12}E_0\right) = E_0. \tag{2.98}$$

The coefficients t_{ij}/r_{ij} are the Fresnel transmission/reflection coefficients going from material i to j. We also require that the sum of the waves travelling along $\boldsymbol{k}^{(4)}$add to zero, giving

$$t_{12}\left(r_{12}E_0\right) + r_{21}\left(t_{12}E_0\right) = 0. \tag{2.99}$$

Putting these together we have the Stokes' relations

$$t_{12}t_{21} + r_{12}^2 = 1, \tag{2.100}$$

$$r_{12} = -r_{21}. \tag{2.101}$$

This can also be verified by direct substitution of the formulae for the Fresnel coefficients.

2.5.5 Transmittance and Reflectance

We can also relate the energy of the outgoing waves to the energy of the incoming waves. The average power across the boundary is found by taking the product of the Poynting vector and the boundary normal. By equation (2.38) the power of each plane wave is

$$P^{(i)} = \frac{n_1\left|\boldsymbol{E}^{(i)}\right|^2}{2}\hat{k}_z^{(i)}, \tag{2.102}$$

$$P^{(r)} = \frac{n_1\left|\boldsymbol{E}^{(r)}\right|^2}{2}\hat{k}_z^{(r)}, \tag{2.103}$$

$$P^{(t)} = \frac{n_2\left|\boldsymbol{E}^{(t)}\right|^2}{2}\hat{k}_z^{(t)}. \tag{2.104}$$

We then find the ratio of the transmitted power to incident power, known as the transmittance T, to be

$$T = \frac{q_2}{q_1}\left|t\right|^2 = \frac{n_2\cos\theta_t}{n_1\cos\theta_i}\left|t\right|^2, \tag{2.105}$$

making use of equations (2.83), (2.85), and (2.92). We can also find the ratio of reflected power to incident power, known as the reflectance, R, to be

$$R = \left|r\right|^2, \tag{2.106}$$

making use of equations (2.83), (2.84), and (2.91). By direct substitution of equations (2.96a) to (2.96d) we find

$$T + R = 1 \tag{2.107}$$

and, as expected, energy is conserved across the boundary.

2.5.6 Total Internal Reflection and the Evanescent Wave

We now return to Snell's law, equation (2.87), governing the angle of refraction. In the case that $n_2 < n_1$ there is a critical angle

$$\theta_{\text{crit}} = \sin^{-1}\left(\frac{n_2}{n_1}\right) \tag{2.108}$$

such that for $\theta > \theta_{\text{crit}}$

$$\sin\theta_t > 1. \tag{2.109}$$

The solution to this relation is that the direction of propagation θ_t is complex and, clearly, this cannot represent a plane wave. In fact, if we consider the normal component of the wavevector, we see that

$$k_z = k_2\cos\theta_t = k_2\sqrt{1 - \frac{n_1}{n_2}\sin\theta_i} = ik_2\sqrt{\frac{n_1}{n_2}\sin\theta_i - 1} \tag{2.110}$$

becomes imaginary. The transmitted wave is then

$$\boldsymbol{E}^{(t)} = \hat{\boldsymbol{E}}^{(t)} e^{i(k_2\sin\theta_t x - \omega t)} e^{-k_2\sqrt{\frac{n_1}{n_2}\sin\theta_i - 1}z}, \tag{2.111}$$

and we see that the wave, known as an evanescent wave, decays with distance from the boundary.

Looking at the reflection coefficient, from equations (2.96a) and (2.96c), we see that r is of the form

$$r = \frac{X - iY}{X + iY}. \tag{2.112}$$

It follows $R = |r|^2 = 1$ and all of the (time averaged) power is reflected. In this instance, we have what is known as total internal reflection.

This does not mean that no energy crosses the boundary, though. In fact, r lies on the unit circle and represents a phase shift between the incident and reflected wave. There is, then, an instantaneous imbalance of power between the incoming and reflected waves, and the evanescent wave serves as an overflow tank passing energy back and forth across the boundary to accommodate this[36]. We can see this by considering

that, since $\boldsymbol{E}$ and $\boldsymbol{H}$ are perpendicular to $\boldsymbol{k}$,

$$\tilde{E}_x^{(t)} = -\frac{k_z^{(t)}}{k_x^{(i)}}\tilde{E}_z^{(t)} \tag{2.113}$$

for a p-polarised wave and

$$\tilde{H}_x^{(t)} = -\frac{k_z^{(t)}}{k_x^{(i)}}\tilde{H}_z^{(t)} \tag{2.114}$$

for an s-polarised. The factor $k_z^{(t)}/k_x^{(i)}$ is purely imaginary and the in-plane components of $E^{(t)}$ (p-polarised) or $H^{(t)}$ (s-polarised) oscillate 90° out of phase, tracing an ellipse. In both cases the Poynting vector similarly oscillates, with the direction of energy flux across the boundary switching directions when $E^{(t)}$ or $H^{(t)}$ reach their minimum or maximum. We should note that it is this back and forth normal to the boundary that, in analogy with a mexican wave, gives the *appearance* of a wave travelling along the boundary (as in equation (2.111)), which is why these are often referred to as surface waves. These waves cannot actually propagate in the x direction, or else they would have a component of $\boldsymbol{E}$ or $\boldsymbol{H}$ parallel to $\boldsymbol{k}$ and would propagate with a speed of $c/n_1 \sin\theta_i$, rather than c/n_2.

The fact that the evanescent wave can carry energy across and away from the boundary is important for the process known as frustrated total internal reflection. Suppose we add a third layer to the system with refractive index $n_3 > n_2$ such that $n_1 \sin\theta_1/n_3 \leq 1$, i.e., the normal component of the wavevector is real in this medium. If the middle layer, in which the evanescent wave exists, is thin enough, the wave will not have completely decayed upon reaching the third layer. The usual boundary conditions will then necessitate a propagating wave in the third medium, and some of the incident energy will be transported across the system instead of being reflected. This notion will be important in section 2.6.

2.5.7 Fresnel Amplitude Coefficients at an Isotropic/Anisotropic Boundary

So far, we have only considered a boundary between two isotropic materials. We will now consider the cases in which a wave is incident upon a uniaxial material from an isotropic material [31]. Once again we take the material boundary to be the $z = 0$ plane and the plane of incidence to be the xz plane. Phase matching at the boundary means that all wavevectors must have the same x component, denoted by K, and zero y component. All waves must also have the same time dependence, $e^{-i\omega t}$, and we omit this term, though it is implicitly understood to be present in all fields. As a result, we can describe any wave present in the system as

$$\boldsymbol{E}^{\alpha} = \hat{\boldsymbol{E}}^{\alpha} A_{\alpha} e^{i(Kx + q_{\alpha} z)}, \tag{2.115}$$

where $\hat{\boldsymbol{E}}^\alpha$ is the polarisation, A_α is the amplitude, $q_\alpha = \sqrt{k_0^2 n_\alpha^2 - K^2}$ is the normal component, and n_α is the refractive index seen by the α wave. This encompasses the forwards and backwards s- and p-waves in the isotropic medium, and the o- and e-waves in the anistropic medium. By equation (2.3c) the corresponding magnetic fields are

$$\boldsymbol{H}^\alpha = \frac{1}{i\mu_0\omega} \begin{pmatrix} -\partial_z E_y^\alpha \\ \partial_z E_x^\alpha - iKE_z^\alpha \\ \partial_x E_y^\alpha \end{pmatrix} e^{i(Kx+q_\alpha z)}. \tag{2.116}$$

For our boundary conditions we use the continuity of the tangential components of $\boldsymbol{E}$ and $\boldsymbol{H}$, implying continuity of E_x, E_y, $\partial_z E_y$, and $\partial_z E_x - iKE_z$.

Since the incident wave can be expressed in terms of its s- and p-components, we will now consider separately how these two polarisations are reflected to s- and p-waves and how they transmit to the o- and e-waves. Beginning with an s-polarised wave, the z dependence of the incident field is given by

$$\boldsymbol{E}^{in} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} e^{iq_{in}}, \tag{2.117}$$

the reflected field as

$$\boldsymbol{E}^r = r_{ss} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} e^{-iq_{in}} - \frac{r_{sp}}{k_{in}} \begin{pmatrix} q_{in} \\ 0 \\ K \end{pmatrix} e^{-iq_{in}}, \tag{2.118}$$

and the transmitted field as

$$\boldsymbol{E}^t = t_{so}\hat{\boldsymbol{E}}^o e^{iq_o z} + t_{se}\hat{\boldsymbol{E}}^e e^{iq_e z}. \tag{2.119}$$

This defines the amplitude coefficients r_{ss}, r_{sp}, t_{so}, and t_{se} as the reflection and transmission of an s-wave to an s-, p-, o-, and e-wave, respectively. Applying the four continuity continuity conditions to the fields $\boldsymbol{E}^{in} + \boldsymbol{E}^r$ and $\boldsymbol{E}^t$ and then solving the four simultaneous equations then gives

$$r_{ss} = \left[(q_{in} - q_e) A_o E_y^e - (q_{in} - q_o) A_e E_y^o \right] D_{in}^{-1}, \tag{2.120a}$$

$$r_{sp} = -2k_{in} \left(A_o E_x^e - A_e E_x^o \right) D_{in}^{-1}, \tag{2.120b}$$

$$t_{so} = -2q_{in} A_e D_{in}^{-1}, \tag{2.120c}$$

$$t_{se} = 2q_{in} A_o D_{in}^{-1}, \tag{2.120d}$$

where the A_α terms are defined by

$$A_\alpha = \left(k_{in}^2 q_{in}^{-1} + q_\alpha \right) E_x^\alpha - KE_z^\alpha \tag{2.120e}$$

and the denominator D_{in} by

$$D_{in} = (q_{in} + q_e)\, A_o E_y^e - (q_{in} + q_o)\, A_e E_y^o. \tag{2.120f}$$

For an incident p-wave we define the z dependence of the incident field as

$$\boldsymbol{E}^{in} = \frac{1}{k_{in}} \begin{pmatrix} q_{in} \\ 0 \\ -K \end{pmatrix} e^{iq_{in}}, \tag{2.121}$$

the reflected field as

$$\boldsymbol{E}^{r} = r_{ps} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} e^{-iq_{in}} - \frac{r_{pp}}{k_{in}} \begin{pmatrix} q_{in} \\ 0 \\ K \end{pmatrix} e^{-iq_{in}}, \tag{2.122}$$

and the transmitted field as

$$\boldsymbol{E}^{t} = t_{po} \hat{\boldsymbol{E}}^o e^{iq_o z} + t_{pe} \hat{\boldsymbol{E}}^e e^{iq_e z}. \tag{2.123}$$

As before, applying the continuity conditions and solving the system of equations gives

$$r_{pp} = -\frac{2k_{in}}{q_{in}} \left[(q_{in} + q_e)\, E_y^e E_x^o - (q_{in} + q_o)\, E_y^o E_x^e \right] D^{-1} + 1 \tag{2.124a}$$

$$r_{ps} = 2k_{in}\, (q_e - q_o)\, E_y^o E_y^e D^{-1} \tag{2.124b}$$

$$t_{po} = 2k_{in}\, (q_{in} + q_e)\, E_y^e D^{-1} \tag{2.124c}$$

$$t_{pe} = -2k_{in}\, (q_{in} + q_o)\, E_y^o D^{-1}. \tag{2.124d}$$

It should be noted that these equation differ slightly from their derivation in [31]. Firstly, all terms in this work are determined in terms of the wavevector components K and q, where they are originally determined in [31], in part, by an angle of propagation. The orientation of the coordinate axes and in which direction this angle is measured are ambiguous, but the problems arising from this can be avoided by writing the equations as above. This work also has a change of sign to [31] in the equations for r_{pp} and r_{sp}. The author of [31] defines the p-polarisation of the reflected wave as pointing to the left of the wavevector, contrary to the convention in this work, introduced in equations (2.96d) and (2.96c) for isotropic media. The r_{pp} and r_{sp} coefficients have thus be redefined such that the positive orientation of a p-wave is consistent, regardless of direction of travel, avoiding troubles later when propagation in multilayered systems is considered.

We now consider the case of an o- or e-wave incident upon the boundary between a crystal and an isotropic material [54]. We differentiate between the forward and backward travelling waves with the superscripts $+$ and $-$, respectively. In this way, the field $\boldsymbol{E}^{o+}$ is the electric field of the forward travelling (incident) o-wave, while q_{e-} is the normal component of the backward travelling (reflected) e-wave wavevector, etc. Considering first an incident o-wave we then have the z dependence of the incident field as

$$\boldsymbol{E}^{in} = \hat{\boldsymbol{E}}^{o+} e^{iq_{o+}}, \tag{2.125}$$

the reflected field as

$$\boldsymbol{E}^{r} = r_{oo}\hat{\boldsymbol{E}}^{o-} e^{iq_{o-}} + r_{oe}\hat{\boldsymbol{E}}^{e-} e^{iq_{e-}}, \tag{2.126}$$

and the transmitted field as

$$\boldsymbol{E}^{t} = \frac{t_{op}}{k_{out}} \begin{pmatrix} q_{out} \\ 0 \\ -K \end{pmatrix} e^{iq_{out}z} + t_{os} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} e^{iq_{out}z}. \tag{2.127}$$

The positive p-polarisation is again defined as lying to the right of the wavevector to agree with all previous coefficient definitions. Applying the boundary conditions and solving the system of equations then gives that

$$r_{oo} = \left[B_{e-} \left(q_{out} - q_{o+} \right) E_y^{o+} - B_{o+} \left(q_{out} - q_{e-} \right) E_y^{e-} \right] D_{out}^{-1}, \tag{2.128a}$$

$$r_{oe} = \left[B_{o+} \left(q_{out} - q_{o-} \right) E_y^{e-} - B_{o-} \left(q_{out} - q_{o+} \right) E_y^{o+} \right] D_{out}^{-1}, \tag{2.128b}$$

$$\begin{aligned} t_{os} = \Big[& B_{e-} \left(q_{o-} - q_{o+} \right) E_y^{o+} E_y^{o-} + B_{o-} \left(q_{o+} - q_{e-} \right) E_y^{o+} E_y^{e-} \\ & + B_{o+} \left(q_{e-} - q_{o-} \right) E_y^{o-} E_y^{e-} \Big] D^{-1}, \end{aligned} \tag{2.128c}$$

$$\begin{aligned} t_{op} = k_{out} q_{out}^{-1} \Big\{ & B_{e-} \left[\left(q_{out} - q_{o+} \right) E_x^{o-} E_y^{o+} - \left(q_{out} - q_{o+} \right) E_x^{o+} E_y^{o-} \right] \\ & + B_{o-} \left[\left(q_{out} - q_{e-} \right) E_x^{o+} E_y^{e-} - \left(q_{out} - q_{o+} \right) E_x^{e-} E_y^{o+} \right] \\ & + B_{o+} \left[\left(q_{out} - q_{o-} \right) E_x^{e-} E_y^{o-} - \left(q_{out} - q_{e-} \right) E_x^{o-} E_y^{e-} \right] \Big\} D_{out}^{-1} \end{aligned} \tag{2.128d}$$

where the denominator is given by

$$D_{out} = B_{o-} \left(q_{out} - q_{e-} \right) E_y^{e-} - B_{e-} \left(q_{out} - q_{o-} \right) E_y^{o-} \tag{2.128e}$$

and the B_α terms are given by

$$B_\alpha = \left(k_{out}^2 q_{in} - q_\alpha \right) E_x^\alpha + K E_z^\alpha. \tag{2.128f}$$

Repeating the same procedure for an incident e-wave we have the incident field as

$$\boldsymbol{E}^{in} = \hat{\boldsymbol{E}}^{e+} e^{iq_{e+}}, \tag{2.129}$$

the reflected field as

$$\boldsymbol{E}^r = r_{eo}\hat{\boldsymbol{E}}^{o-}e^{iq_{o-}} + r_{ee}\hat{\boldsymbol{E}}^{e-}e^{iq_{e-}}, \tag{2.130}$$

and the transmitted field as

$$\boldsymbol{E}^t = \frac{t_{ep}}{k_{out}}\begin{pmatrix} q_{out} \\ 0 \\ -K \end{pmatrix} e^{iq_{out}z} + t_{es}\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} e^{iq_{out}z}. \tag{2.131}$$

The Fresnel coefficients are then

$$r_{eo} = \left[B_{e-}\left(q_{out} - q_{e+}\right)E_y^{e+} - B_{e+}\left(q_{out} - q_{e-}\right)E_y^{e-}\right]D_{out}^{-1}, \tag{2.132a}$$

$$r_{ee} = \left[B_{e+}\left(q_{out} - q_{o-}\right)E_y^{o-} - B_{o-}\left(q_{out} - q_{e+}\right)E_y^{e+}\right]D_{out}^{-1}, \tag{2.132b}$$

$$\begin{aligned} t_{es} = \Big[&B_{o-}\left(q_{e+} - q_{e-}\right)E_y^{e+}E_y^{e-} + B_{e-}\left(q_{o-} - q_{e+}\right)E_y^{e+}E_y^{o-} \\ &+B_{e+}\left(q_{e-} - q_{o-}\right)E_y^{o-}E_y^{e-}\Big]D^{-1}, \end{aligned} \tag{2.132c}$$

$$\begin{aligned} t_{ep} = k_{out}q_{out}^{-1}\Big\{&B_{o-}\left[\left(q_{out} - q_{e-}\right)E_x^{e+}E_y^{e-} - \left(q_{out} - q_{e+}\right)E_x^{e-}E_y^{e+}\right] \\ &+ B_{e-}\left[\left(q_{out} - q_{e+}\right)E_x^{o-}E_y^{e+} - \left(q_{out} - q_{o-}\right)E_x^{e+}E_y^{o-}\right] \\ &+B_{e+}\left[\left(q_{out} - q_{o-}\right)E_x^{e-}E_y^{o-} - \left(q_{out} - q_{e-}\right)E_x^{o-}E_y^{e-}\right]\Big\}D_{out}^{-1}. \end{aligned} \tag{2.132d}$$

2.6 Propagation Through Multiple Layers

2.6.1 Fabry-Perot Resonator

Until now we have only considered the case of two semi-infinite layers, with the transmitted and reflected waves propagating to infinity. We now consider the so called Fabry-Perot resonator/cavity, a three layer system of isotropic materials in which a wave is incident upon a layer of finite thickness (but infinite width), with the media before and after the layer extending to infinity. This means that we are now looking to determine the fields that are reflected and transmitted from a layer, and not just a single boundary.

We have already seen that when a field is incident upon a material boundary, there must be both a transmitted and reflected wave. Since the layer has two boundaries, there must be two fields in the centre layer, one traveling towards the third material (like the incident wave) and one back towards the incident layer. The fields of the system are depicted in figure (2.4). Phase matching at the boundary, as we already have seen, necessitates the continuity of the tangential component of the wavevector at a boundary. Since it must be continuous from medium one to medium two, and then from medium two to medium three, we can conclude that it must be continuous

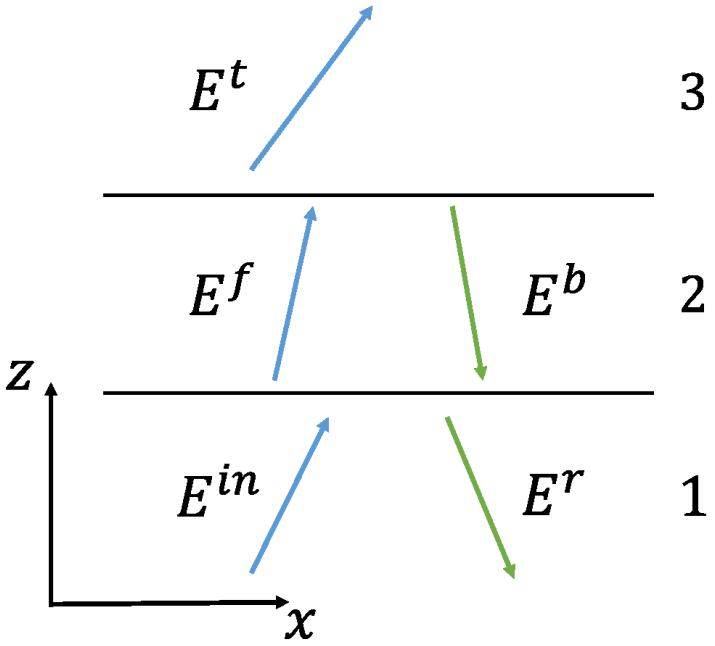


FIGURE 2.4: A schematic of the fields arising in a three layer system subject to an incident field E^{in}.

throughout all layers. This also means that the magnitude of the normal components of the forward and backward wavevectors in each layer are the same. Take the plane of incidence, then, to be the xz plane, and the boundaries to lie at $z = 0$ and $z = L$ (the thickness of the layer), all waves have the same x dependence, zero y dependence, and we describe the z dependence of the field in each layer as

$$\boldsymbol{E}^{(1)} = \boldsymbol{E}^{in} e^{iq_1 z} + \boldsymbol{E}^{r} e^{-iq_1 z}, \tag{2.133a}$$

$$\boldsymbol{E}^{(2)} = \boldsymbol{E}^{f} e^{iq_2 z} + \boldsymbol{E}^{b} e^{-iq_2 z}, \tag{2.133b}$$

$$\boldsymbol{E}^{(3)} = \boldsymbol{E}^{t} e^{iq_3 z}, \tag{2.133c}$$

where $\boldsymbol{E}^i$ and q_i are the field and normal component of the wavevector in layer i. If we consider an incident s-polarised wave, we know from the results of section (2.5.3) that all of the other fields are also s-polarised. The boundary conditions we need to apply are then the continuity of E_y and $\partial_z E_y$ at each of the boundaries. This gives us the system of equations

$$E_y^{in} + E_y^{r} = E_y^{f} + E_y^{b} \tag{2.134a}$$

$$q_1 \left(E_y^{in} - E_y^{r} \right) = q_2 \left(E_y^{f} - E_y^{b} \right) \tag{2.134b}$$

$$e^{iq_2 L} E_y^{f} + e^{-iq_2 L} E_y^{b} = E_y^{t} \tag{2.134c}$$

$$q_2 \left(e^{iq_2 L} E_y^{f} - e^{-iq_2 L} E_y^{b} \right) = q_3 E_y^{t}. \tag{2.134d}$$

If we eliminate E_y^t we find that

$$\begin{aligned} E_y^b &= \frac{q_2 - q_3}{q_2 + q_3} e^{2iq_2L} E_y^b \\ &= r_{23}^s e^{2iq_2L} E_y^f \end{aligned} \tag{2.135}$$

(by equation (2.96a)), where the superscript s and subscripts ij denote that it is the Fresnel coefficient for an s-polarised wave going from layer i to layer j. Intuitively, we see that the backward field is simply given by propagating the forward field by $z = L$, reflecting from the second boundary, and propagating the field back by another $z = L$. We can now eliminate the fields in the layer to write

$$E_y^r = \frac{r_{12}^s + r_{23}^s e^{2iq_2L}}{1 + r_{12}^s r_{23}^s e^{2iq_2L}} E_y^{in}. \tag{2.136}$$

This allows us to define the effective reflection coefficient of the layer for an s-polarised wave as,

$$r_{13}^s = \frac{E_y^r}{E_y^{in}} = \frac{r_{12}^s + r_{23}^s e^{2iq_2L}}{1 + r_{12}^s r_{23}^s e^{2iq_2L}}, \tag{2.137}$$

We can similarly define an effective transmission coefficient. Eliminating E_y^r from equations (2.134a) and (2.134b) gives

$$\begin{aligned} E_y^f &= \frac{2q_1}{q_1 + q_2} E_y^{in} - \frac{q_1 - q_2}{q_1 + q_2} E_y^b \\ &= t_{12}^s E_y^{in} + r_{21}^s E_y^b, \end{aligned} \tag{2.138}$$

making use of equations (2.96b),(2.96a), and (2.101). Intuitively, again, we see that the forward field in the layer is equal to the transmission of the incident field plus the reflection of the backward field. By using equation (2.135) we can actually write this entirely in terms of the incoming field,

$$E_y^f = \frac{t_{12}^s}{1 + r_{12}^s r_{23}^s e^{2iq_2L}} E_y^{in}. \tag{2.139}$$

Finally, by eliminating $e^{-iq_2L} E_y^b$ from equations (2.134c) and (2.134d), we can use this to write

$$\begin{aligned} E_y^{out} &= \frac{2q_2}{q_2 + q_3} \frac{t_{12}^s e^{iq_2L}}{1 + r_{12}^s r_{23}^s e^{2iq_2L}} E_y^{in} \\ &= \frac{t_{12}^s t_{23}^s e^{iq_2L}}{1 + r_{12}^s r_{23}^s e^{2iq_2L}} E_y^{in}, \end{aligned} \tag{2.140}$$

(again using equation (2.96d)) defining the effective transmission coefficient as

$$t_{13}^s = \frac{E_y^{out}}{E_y^{in}} = \frac{t_{12}^s t_{23}^s e^{iq_2 L}}{1 + r_{12}^s r_{23}^s e^{2iq_2 L}}. \tag{2.141}$$

If, instead, we were to consider a p-polarised wave incident upon the layer, we can repeat virtually the same analysis by replacing $\boldsymbol{E}$ with $\boldsymbol{H}$. Equation (2.3d) allows us to write

$$\boldsymbol{E} = \frac{-i}{\omega\varepsilon} \begin{pmatrix} \partial_z H_y \\ \partial_z H_x - iKH_z \\ \partial_x H_y \end{pmatrix} e^{i(Kx + q_\alpha z)} \tag{2.142}$$

and so, for an incident p-polarised wave where only H_y is non-zero, our boundary conditions are the continuity of H_y and $\varepsilon^{-1}\partial_z H_y$. This generates a system of equations with almost the same form as (2.134a) to (2.134d). The result, as you would expect, is that

$$r_{13}^p = \frac{|\boldsymbol{E}^r|}{|\boldsymbol{E}^{in}|} = \frac{H_y^r}{H_y^{in}} = \frac{r_{12}^p + r_{23}^p e^{2iq_2 L}}{1 + r_{12}^p r_{23}^p e^{2iq_2 L}} \tag{2.143}$$

and

$$t_{13}^p = \frac{|\boldsymbol{E}^t|}{|\boldsymbol{E}^{in}|} = \frac{n_1}{n_3}\frac{H_y^t}{H_y^{in}} = \frac{t_{12}^p t_{23}^p e^{iq_2 L}}{1 + r_{12}^p r_{23}^p e^{2iq_2 L}}, \tag{2.144}$$

in analogy with the s-polarised case.

An important feature of these solutions is the existence of resonances inside the layer. The transmission (and reflection) of the layer does not just depend on the reflection and transmission coefficients at the two boundaries, the phase change in crossing the layer is also very important. Looking at the transmittance of the system

$$T = \frac{q_3}{q_1}|t_{13}|^2 = \frac{T_{12}T_{23}}{(1 + |r_{12}r_{23}|)^2 - 4r_{12}r_{23}\sin^2(q_2 L)}, \tag{2.145}$$

we see oscillations in the transmitted field depending on the phase shift $q_2 L$. In the limit that $L \to 0$ and we have no layer, direct substitution of the Fresnel coefficients (2.96a) to (2.96d) gives that

$$t_{13}^s = \frac{2q_1}{q_1 + q_3} \tag{2.146}$$

and

$$t_{13}^p = \frac{2n_1 n_3 q_1}{n_3^2 q_1 + n_1^2 q_3}. \tag{2.147}$$

As you would expect, these are exactly the coefficients describing the transmission across an ordinary boundary between medium 1 and 3. More interestingly, however, are the cases of the resonances where (for non-zero L) the phase shift is a multiple of π. The exponential term in the denominators of (2.137), (2.141), (2.143), and (2.144) is again equal to one, and we find that the incident wave transmits as if the layer were

not there (up to a π phase shift in the case that $q_2 L$ is an odd multiple of π). In the case that medium 1 and 3 are the same, this means that none of the light is reflected whatsoever.

2.6.2 The T-Matrix Formulation

The problem for the three layer system, as worked out in the previous section, is already quite algebraically intensive. If we were to add more layers to the system, the problem would very quickly become too complicated. We can, however, make the problem far simpler to solve numerically by rewriting these problems in matrix form. This is known as the transfer matrix, or T-matrix, approach [5].

As an example, we consider the same three layer system again. We can rewrite the system of equations (2.134a) to (2.134d) as

$$\begin{pmatrix} 1 & 1 \\ q_1 & -q_1 \end{pmatrix} \begin{pmatrix} E_y^{in} \\ E_y^{r} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ q_2 & -q_2 \end{pmatrix} \begin{pmatrix} E_y^{f} \\ E_y^{b} \end{pmatrix} \tag{2.148}$$

and

$$\begin{pmatrix} 1 & 1 \\ q_2 & -q_2 \end{pmatrix} \begin{pmatrix} e^{iq_2 L} & 0 \\ 0 & e^{iq_2 L} \end{pmatrix} \begin{pmatrix} E_y^{f} \\ E_y^{b} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ q_3 & -q_3 \end{pmatrix} \begin{pmatrix} E_y^{t} \\ 0 \end{pmatrix}. \tag{2.149}$$

Simple matrix operations then allow us to write an equation of the form

$$\begin{pmatrix} E_y^{t} \\ 0 \end{pmatrix} = \mathrm{M} \begin{pmatrix} E_y^{in} \\ E_y^{r} \end{pmatrix}. \tag{2.150}$$

From this we first deduce that the reflection of the layer is given by $-\mathrm{M}_{21}\mathrm{M}_{22}^{-1}$, and then that the transmission of the layer is given by $\mathrm{M}_{11} - \mathrm{M}_{12}\mathrm{M}_{21}\mathrm{M}_{22}^{-1}$ (where M_{22}^{-1} denotes the inverse of M_{22}, not the fourth element of M^{-1}).

In general, if we have $N+1$ layers (incident layer is 0 and output layer is N), we can relate the forward and backward fields, F and B, in layers n and $n+1$ by

$$\begin{pmatrix} 1 & 1 \\ q_n & -q_n \end{pmatrix} \begin{pmatrix} e^{iq_n} & 0 \\ 0 & e^{-iq_n} \end{pmatrix} \begin{pmatrix} F^{(n)} \\ B^{(n)} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ q_{n+1} & -q_{n+1} \end{pmatrix} \begin{pmatrix} F^{(n+1)} \\ B^{(n+1)} \end{pmatrix}. \tag{2.151}$$

Equivalently, we have the relationship

$$\begin{pmatrix} F^{(n+1)} \\ B^{(n+1)} \end{pmatrix} = \mathrm{T}^{(n)}\mathrm{P}^{(n)} \begin{pmatrix} F^{(n)} \\ B^{(n)} \end{pmatrix} \tag{2.152}$$

where

$$\mathrm{T}^{(n)} = \begin{pmatrix} 1 & 1 \\ q_{n+1} & -q_{n+1} \end{pmatrix}^{-1} \begin{pmatrix} 1 & 1 \\ q_n & -q_n \end{pmatrix} \tag{2.153}$$

accounts for the transmission and reflection at the n'th boundary and

$$\mathrm{P}^{(n)} = \begin{pmatrix} e^{iq_nL_n} & 0 \\ 0 & e^{-iq_nL_n} \end{pmatrix} \tag{2.154}$$

accounts for the phase change across the n'th layer, with thickness L_n. We can then associate a matrix

$$\mathrm{M}^{(n)} = \mathrm{T}^{(n)}\mathrm{P}^{(n)} \tag{2.155}$$

with each layer and the entire system can be described by

$$\mathrm{M} = \left(\prod_{i=1}^{N} \mathrm{M}^{(n)}\right) \begin{pmatrix} 1 & 1 \\ q_1 & -q_1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 1 \\ q_0 & -q_0 \end{pmatrix}. \tag{2.156}$$

From the relationship

$$\begin{pmatrix} F_y^{(N)} \\ B_y^{(N)} \end{pmatrix} = \mathrm{M} \begin{pmatrix} F_y^{(0)} \\ B_y^{(0)} \end{pmatrix} \tag{2.157}$$

and setting $B_y^{(N)} = 0$, we deduce that the reflection and transmission of the system are, as before,

$$r_{0N}^s = -\mathrm{M}_{21}\mathrm{M}_{22}^{-1} \tag{2.158}$$

$$t_{0N}^s = \mathrm{M}_{11} - \mathrm{M}_{12}\mathrm{M}_{21}\mathrm{M}_{22}^{-1} \tag{2.159}$$

when travelling from layer 0 to layer N. Setting $F_y^{(0)} = 0$ we find that the reflection and transmission when travelling from layer N to layer 0 are

$$r_{N0}^s = \mathrm{M}_{12}\mathrm{M}_{22}^{-1} \tag{2.160}$$

$$t_{N0}^s = \mathrm{M}_{22}^{-1}. \tag{2.161}$$

Just as in the algebraic method in the previous section, an analogous method holds for an incident p-polarised wave. The p-polarised case can be calculated simply by changing the q_n's for $q_n\varepsilon_n^{-1}$ in the $\mathrm{T}^{(n)}$ matrices, thus relating the $\boldsymbol{H}$ fields on either side of the layered system. By equation (2.36) the reflection and transmission coefficients are then

$$r_{0N}^p = -\mathrm{M}_{21}\mathrm{M}_{22}^{-1} \tag{2.162}$$

$$t_{0N}^p = n_0 n_N^{-1}\left(\mathrm{M}_{11} - \mathrm{M}_{12}\mathrm{M}_{21}\mathrm{M}_{22}^{-1}\right) \tag{2.163}$$

$$r_{N0}^p = \mathrm{M}_{12}\mathrm{M}_{22}^{-1} \tag{2.164}$$

$$t^p_{N0} = n_N n_0^{-1} \mathrm{M}_{22}^{-1}. \tag{2.165}$$

2.6.3 The S-Matrix Formulation

While the transfer matrix is theoretically correct, it can run into accuracy problems during its numerical computation. In the event of evanescent waves forming inside a layer, the normal component of the wavevector q_n becomes imaginary. We see, then, from equation (2.154), that while the forward mode is exponentially decaying, the backward is actually growing exponentially. By relating the fields in neighbouring layers in this way, we are not only growing the backward field exponentially, but also any numerical errors. This instability in the T-matrix formulation can lead to enormous errors in the final solution. This problem, however, can avoided by using the scattering matrix (or S-matrix) approach [28].

In this method, rather than relating the fields on one side of the system to the fields on the other, the S-matrix relates the outgoing fields to the incoming. Using the notation of the previous section, we have the system of equations

$$\begin{pmatrix} F^{(N)} \\ B^{(0)} \end{pmatrix} = \mathrm{S}^{(N)} \begin{pmatrix} F^{(0)} \\ B^{(N)} \end{pmatrix}. \tag{2.166}$$

The full S-matrix, $\mathrm{S}^{(N)}$, is built inductively. First, we assume that we know $\mathrm{S}^{(n)}$, the matrix describing the layer 0 to layer n subsystem in the manner of (2.166). We then relate layer $n+1$ to layer n using (2.152) to write

$$\begin{pmatrix} F^{(n)} \\ B^{(n)} \end{pmatrix} = \mathrm{L}^{(n)} \begin{pmatrix} F^{(n+1)} \\ B^{(n+1)} \end{pmatrix}, \tag{2.167}$$

where the layer matrices $\mathrm{L}^{(n)}$ are the inverse of the matrices $\mathrm{M}^{(n)}$. Eliminating the fields $F^{(n)}$ and $B^{(n)}$ then gives

$$\begin{aligned} \begin{pmatrix} F^{(n+1)} \\ B^{(0)} \end{pmatrix} &= \begin{pmatrix} -\mathrm{S}^{(n)}_{22}\mathrm{L}^{(n)}_{21} & 1 \\ \mathrm{L}^{(n)}_{11} - \mathrm{S}^{(n)}_{12}\mathrm{L}^{(n)}_{21} & 0 \end{pmatrix}^{-1} \begin{pmatrix} \mathrm{S}^{(n)}_{21} & \mathrm{S}^{(n)}_{22}\mathrm{L}^{(n)}_{22} \\ \mathrm{S}^{(n)}_{11} & \mathrm{S}^{(n)}_{12}\mathrm{L}^{(n)}_{22} - \mathrm{L}^{(n)}_{12} \end{pmatrix} \begin{pmatrix} F^{(0)} \\ B^{(n+1)} \end{pmatrix} \\ &= \mathrm{S}^{(n+1)} \begin{pmatrix} F^{(0)} \\ B^{(n+1)} \end{pmatrix}. \end{aligned} \tag{2.168}$$

All of the S-matrices are then fully defined by setting $\mathrm{S}^{(0)}$ as the identity and iteratively calculating

$$\mathrm{S}^{(n+1)}_{11} = \mathrm{M}^{(n)}_{11} \left(1 - \mathrm{M}^{(n)}_{11}\mathrm{S}^{(n)}_{12}\mathrm{L}^{(n)}_{21}\right)^{-1} \mathrm{S}^{(n)}_{11}, \tag{2.169a}$$

$$\mathrm{S}^{(n+1)}_{12} = \mathrm{M}^{(n)}_{11} \left(1 - \mathrm{M}^{(n)}_{11}\mathrm{S}^{(n)}_{12}\mathrm{L}^{(n)}_{21}\right)^{-1} \left(\mathrm{S}^{(n)}_{12}\mathrm{L}^{(n)}_{22} - \mathrm{L}^{(n)}_{12}\right), \tag{2.169b}$$

$$\mathrm{S}_{21}^{(n+1)} = \mathrm{S}_{22}^{(n)}\mathrm{L}_{21}^{(n)}\mathrm{S}_{11}^{(n+1)} + \mathrm{S}_{21}^{(n)}, \tag{2.169c}$$
$$\mathrm{S}_{22}^{(n+1)} = \mathrm{S}_{22}^{(n)}\mathrm{L}_{21}^{(n)}\mathrm{S}_{11}^{(n+1)} + \mathrm{S}_{22}^{(n)}\mathrm{L}_{22}^{(n)}. \tag{2.169d}$$

If we suppose that the fields in the n'th layer are evanescent, we can see from equation (2.152) that the elements $\mathrm{M}_{i1}^{(n)}$ are decaying whilst the $\mathrm{M}_{i2}^{(n)}$ are growing. Conversely, the $\mathrm{L}_{i1}^{(n)}$ are growing whilst the $\mathrm{L}_{i2}^{(n)}$ are decaying. From this it can be seen that all four elements of $\mathrm{S}^{(n+1)}$are at most $O\left(1\right)$, with none of them growing exponentially, and it is the calculation of the propagating modes that dominates numerically, not the growing evanescent waves and any associated errors in their calculation.

Under the S-matrix formulation the transmission and reflection coefficients of the system are much simpler to find. By equation (2.166) we simply have

$$t_{0N}^{s} = \mathrm{S}_{11}^{(N)}, \tag{2.170a}$$
$$r_{0N}^{s} = \mathrm{S}_{21}^{(N)}, \tag{2.170b}$$
$$t_{N0}^{s} = \mathrm{S}_{22}^{(N)}, \tag{2.170c}$$
$$r_{N0}^{s} = \mathrm{S}_{12}^{(N)}, \tag{2.170d}$$

and

$$t_{0N}^{p} = n_0 n_N^{-1}\mathrm{S}_{11}^{(N)}, \tag{2.170e}$$
$$r_{0N}^{p} = \mathrm{S}_{21}^{(N)}, \tag{2.170f}$$
$$t_{N0}^{p} = n_N n_0^{-1}\mathrm{S}_{22}^{(N)}, \tag{2.170g}$$
$$r_{N0}^{p} = \mathrm{S}_{12}^{(N)}, \tag{2.170h}$$

making sure to use

$$\mathrm{M}^{(n)} = \begin{cases} \begin{pmatrix} 1 & 1 \\ q_{n+1} & -q_{n+1} \end{pmatrix}^{-1} \begin{pmatrix} 1 & 1 \\ q_n & -q_n \end{pmatrix} & \text{s-polarised} \\ \begin{pmatrix} 1 & 1 \\ \varepsilon_{n+1}^{-1}q_{n+1} & -\varepsilon_{n+1}^{-1}q_{n+1} \end{pmatrix}^{-1} \begin{pmatrix} 1 & 1 \\ \varepsilon_n^{-1}q_n & -\varepsilon_n^{-1}q_n \end{pmatrix} & \text{p-polarised.} \end{cases} \tag{2.171}$$

2.7 Diffraction Theory

2.7.1 The Grating Formula

So far, we have only considered the case in which a wave is incident upon one homogenous material from another. If this is not the case, different parts of the wave

must transmit and reflect differently, depending on the parameters of the material at the point they are incident. The resulting waves interfere with one another in a process known as diffraction, resulting in the formation of various different patterns. The nature of the patterns depend upon the properties of the incident light and the material and geometric properties of the media involved. In this section we will consider diffraction gratings, periodic structures that induce a similarly periodic modulation of the incident beam to produce diffraction effects. These gratings can be periodic in multiple directions, though in this work we only consider gratings with one direction of periodicity, and that this direction lies in the boundary plane.

In the presence of periodic media, Bloch's theorem tells us that the general solution to the spatial part of the wave equation (2.26) is of the form [57]

$$\tilde{E}_G(\boldsymbol{r}) = e^{i\boldsymbol{k}\cdot\boldsymbol{r}} u(\boldsymbol{r}), \tag{2.172}$$

where $u(\boldsymbol{r})$ has the same periodicity as the medium, and $\boldsymbol{k}$ is some wavevector. Letting $\boldsymbol{\Lambda}$ be the vector of translational symmetry of the medium, it follows that

$$\begin{aligned} \tilde{E}_G(\boldsymbol{r}+\boldsymbol{\Lambda}) &= e^{i\boldsymbol{k}\cdot(\boldsymbol{r}+\boldsymbol{\Lambda})} u(\boldsymbol{r}+\boldsymbol{\Lambda}) \\ &= e^{i\boldsymbol{k}\cdot\boldsymbol{\Lambda}} \tilde{E}_G(\boldsymbol{r}) \end{aligned} \tag{2.173}$$

by definition of $\tilde{E}_G$ and u. The wave, then, is not actually periodic like the medium, as you may expect. It is, instead, only periodic up to a phase change when changing position by $\boldsymbol{\Lambda}$.

Now, we assume that the surface of the grating is the $z = 0$ plane, and also that $\boldsymbol{\Lambda}$ is contained in this plane. We decompose the Bloch wave wavevector as

$$\boldsymbol{k} = \boldsymbol{k}_{\parallel} + \boldsymbol{k}_{\perp} + \boldsymbol{k}_z, \tag{2.174}$$

where $\boldsymbol{k}_z$ is the z component, and $\boldsymbol{k}_{\parallel}$ and $\boldsymbol{k}_{\perp}$ are the decomposition of the xy components into vectors that are parallel and perpendicular to $\boldsymbol{\Lambda}$. If we now consider $\tilde{E}'_G(\boldsymbol{r}) = \tilde{E}_G(\boldsymbol{r})\, e^{-i\boldsymbol{k}_{\parallel}\cdot\boldsymbol{r}}$, we see that

$$\begin{aligned} \tilde{E}'_G(\boldsymbol{r}+\boldsymbol{\Lambda}) &= \tilde{E}'_G(\boldsymbol{r}+\boldsymbol{\Lambda})\, e^{-i\boldsymbol{k}_{\parallel}\cdot(\boldsymbol{r}+\boldsymbol{\Lambda})} \\ &= \tilde{E}_G(\boldsymbol{r})\, e^{i\boldsymbol{k}\cdot\boldsymbol{\Lambda}} e^{-i\boldsymbol{k}_{\parallel}\cdot\boldsymbol{r}} \\ &= \tilde{E}'_G(\boldsymbol{r}), \end{aligned} \tag{2.175}$$

and $\tilde{E}'_G$ *is* periodic in the direction of $\boldsymbol{\Lambda}$. The second equality comes from (2.173), and the third equality follows from the fact that $\boldsymbol{k}\cdot\boldsymbol{\Lambda} = \left(\boldsymbol{k}_{\parallel} + \boldsymbol{k}_{\perp} + \boldsymbol{k}_z\right)\cdot\boldsymbol{\Lambda} = \boldsymbol{k}_{\parallel}\cdot\boldsymbol{\Lambda}$.

We use this as an opportunity to introduce Fourier theory for wave optics. As discussed in section 2.4.1, we can decompose any field as a sum of plane waves (owing to the orthogonality of the complex exponentials), which can greatly simplify

the calculation and understanding of the behaviour of a system. This is particularly powerful in the case of diffraction from periodic media, where the resulting periodic functions have a discrete Fourier series. $\tilde{E}_G'$, for example, is periodic in $\mathbf{\Lambda}$, and so we know we can write its dependence in the direction of $\mathbf{\Lambda}$ as the sum of plane waves who are themselves periodic in $\mathbf{\Lambda}$. We also know that the dependence of the Bloch wave in the direction of $\boldsymbol{k}_\perp$ is fixed by phase matching with the wave incident upon the grating, leaving only the z dependence unknown. We put this together to find

$$\tilde{E}_G(\boldsymbol{r})\, e^{-i\boldsymbol{k}_\parallel\cdot\boldsymbol{r}} = \sum_{m=-\infty}^{\infty} \tilde{E}_m^G(z)\, e^{i(\boldsymbol{k}_\perp + m\boldsymbol{k}_\Lambda)\cdot\boldsymbol{r}}, \tag{2.176}$$

where $\boldsymbol{k}_\Lambda = k_\Lambda \hat{\mathbf{\Lambda}}$ is the grating wavevector, and $k_\Lambda = 2\pi\Lambda^{-1}$ is the grating wavenumber. Upon leaving the grating, phase matching at the second boundary tells that the transmitted field, $\tilde{E}_T$, is of the form

$$\begin{aligned}\tilde{E}_T(\boldsymbol{r}) &= \sum_{m=-\infty}^{\infty} \tilde{E}_m^T(z)\, e^{i\left(\boldsymbol{k}_\perp + \boldsymbol{k}_\parallel + m\boldsymbol{k}_\Lambda\right)\cdot\boldsymbol{r}} \\ &= \sum_{m=-\infty}^{\infty} \tilde{E}_m^T(z)\, e^{i\left(\boldsymbol{k}_\perp + \boldsymbol{k}_\parallel^{(m)}\right)\cdot\boldsymbol{r}}.\end{aligned} \tag{2.177}$$

It must also satisfy the wave equation (2.26), which leads to

$$\frac{d^2\tilde{E}_m^T}{dz^2} + \left[k^2 - \left|\boldsymbol{k}_\perp + \boldsymbol{k}_\parallel^{(m)}\right|^2\right]\tilde{E}_m = 0 \tag{2.178}$$

for each m. Each of these equations has two plane wave solutions given by

$$\tilde{E}_m^T = e^{\pm i k_z^{(m)}\cdot\boldsymbol{r}}, \tag{2.179}$$

where $\boldsymbol{k}_z^{(m)}$ points in the z direction and

$$\left|\boldsymbol{k}_z^{(m)}\right| = \sqrt{k^2 - \left|\boldsymbol{k}_\perp + \boldsymbol{k}_\parallel^{(m)}\right|^2}; \tag{2.180}$$

only the forward propagating solution is applicable, however. It follows that that the general solution for the transmitted field is

$$\tilde{E}^T(\boldsymbol{r}) = \sum_{m=-\infty}^{\infty} \tilde{E}_m e^{i\left(\boldsymbol{k}_\perp + \boldsymbol{k}_\parallel^{(m)} + \boldsymbol{k}_z^{(m)}\right)\cdot\boldsymbol{r}}, \tag{2.181}$$

where the $\tilde{E}_m$ are constants. We see that the grating results in a set of emerging waves, known as the diffracted orders, with the m'th order characterised by the relation

$$\boldsymbol{k}_\parallel^{(m)} = \boldsymbol{k}_\parallel + m\boldsymbol{k}_\Lambda. \tag{2.182}$$

Since we must have phase matching between the incident wave and the Bloch wave

(2.172), $\boldsymbol{k}_{\parallel}$ and $\boldsymbol{k}_{\perp}$ are simply the tangential component of the incident wavevector projected onto $\boldsymbol{\Lambda}$ and $\hat{\boldsymbol{e}}_z \times \boldsymbol{\Lambda}$, respectively.

Thus, through the use of the Fourier transform, we have been able to show that any periodic medium, without knowing any other information than its periodicity, will cause an incident electric field to be diffracted into multiple plane waves. Each wave will always have a tangential wavevector component that is equal to the sum of that of the incident wave and a multiple of the grating wavevector. This does not, however, fully solve the diffraction problem. While we may know the directions of the various emerging waves, we, as of yet, know little about their intensities.

We can, however, make the immediate observation that only some of the diffracted orders can actually propagate. An immediate consequence of (2.182) is that, for large enough $|m|$, the m'th diffracted wave must be evanescent, and only finite a number of orders can propagate. The indices of the propagating modes are then bounded by

$$m_{max} = \left\lfloor \frac{\sqrt{k^2 - k_{\perp}^2} - k_{\parallel}}{k_{\Lambda}} \right\rfloor \tag{2.183a}$$

and

$$m_{min} = \left\lceil -\frac{\sqrt{k^2 - k_{\perp}^2} + k_{\parallel}}{k_{\Lambda}} \right\rceil . \tag{2.183b}$$

2.7.2 Rayleigh-Sommerfeld Diffraction Integrals

In an effort to model and understand the intensities of various diffracted orders produced by a grating, we will first discuss the so called Rayleigh-Sommerfeld diffraction integrals. Under some assumptions on the field leaving the grating, these integrals give an exact description of the diffracted field. While solving the full integrals is in no-way trivial and needs to be completed numerically, they provide a basis for a range of other methods which are far easier to solve after some approximations and simplifications. The full numeric solutions also provide a worthwhile comparison to check the accuracy of other methods.

We start with the Green identity

$$\int \left(U\nabla^2 V - V\nabla^2 U \right) dV = \int \left(U\boldsymbol{n} \cdot \boldsymbol{\nabla} V - V\boldsymbol{n} \cdot \boldsymbol{\nabla} U \right) dS \tag{2.184}$$

for any functions U and V that are continuous and differentiable in a volume V and on its boundary S (described by its inward normal $\boldsymbol{n}$). We will apply this to the spatial part of the electric field, $\tilde{E}$, and G, the Green's function for the Helmholtz equation, satisfying

$$\left(\nabla^2 + k^2 \right) G = \delta \left(\boldsymbol{r} - \boldsymbol{r}' \right) . \tag{2.185}$$

Now, we can see that G is not continuous at the point $\boldsymbol{r}'$. This means that we cannot directly apply the Green identity to any volume containing $\boldsymbol{r}'$. Instead, we consider integrating over a volume V', formed by removing from V a small sphere of radius ε centered on $\boldsymbol{r}'$, denoted by S_ε. The integral over V can then found by taking the limit that $\varepsilon \to 0$. Both $\tilde{E}$ and G satisfy the homogeneous Helmholtz equation in V' and so, from (2.184), we find that

$$\int \tilde{E}\boldsymbol{n}\cdot\boldsymbol{\nabla}G - G\boldsymbol{n}\cdot\boldsymbol{\nabla}\tilde{E}dS + \lim_{\varepsilon\to 0}\int \tilde{E}\boldsymbol{n}\cdot\boldsymbol{\nabla}G - G\boldsymbol{n}\cdot\boldsymbol{\nabla}\tilde{E}dS_\varepsilon = 0. \tag{2.186}$$

Now, to the choice of G. For a point source at $\boldsymbol{r}'$ the solution in free space must be radially symmetric. Thus, solving equation (2.185) in spherical coordinates gives the free space impulse response

$$G_f\left(\boldsymbol{r}-\boldsymbol{r}'\right) = \frac{e^{ik|\boldsymbol{r}-\boldsymbol{r}'|}}{|\boldsymbol{r}-\boldsymbol{r}'|}, \tag{2.187}$$

which is also known as a spherical wave. However, our diffraction problem lies only in the half-space bounded by the plane of the grating, and so we must modify G_f so that it still satisfies (2.185), whilst also allowing us to form a well defined boundary value problem. There are two choices. Defining $\tilde{\boldsymbol{r}}'$ as the reflection of $\boldsymbol{r}'$ across the boundary, the first choice is

$$G_D = G_f\left(\boldsymbol{r}-\boldsymbol{r}'\right) - G_f\left(\boldsymbol{r}-\tilde{\boldsymbol{r}}'\right), \tag{2.188}$$

which is zero at the boundary, and the second choice is

$$G_N = G_f\left(\boldsymbol{r}-\boldsymbol{r}'\right) + G_f\left(\boldsymbol{r}-\tilde{\boldsymbol{r}}'\right), \tag{2.189}$$

which has zero normal derivative at the boundary. As we will see soon, G_D gives a Dirichlet problem for $\tilde{E}$, while G_N gives a Neumann problem.

We return to (2.186). For $\boldsymbol{R} = \boldsymbol{r}$ or $\boldsymbol{r}'$, the normal derivative is given by

$$\boldsymbol{n}\cdot\boldsymbol{\nabla}G_f\left(\boldsymbol{r}-\boldsymbol{R}\right) = \boldsymbol{n}\cdot\hat{\boldsymbol{R}}\left(ik - \frac{1}{|\boldsymbol{r}-\boldsymbol{R}|}\right)\frac{e^{ik|\boldsymbol{r}-\boldsymbol{R}|}}{|\boldsymbol{r}-\boldsymbol{R}|} \tag{2.190}$$

and we find that the only contribution to the second integral comes from $G_f\left(\boldsymbol{r}-\boldsymbol{r}'\right)$ and is equal to $4\pi\tilde{E}\left(\boldsymbol{r}'\right)$. Taking the surface S to bound the intersection of the grating boundary and an infinitely large sphere centered on $\boldsymbol{r}'$, the integral over the spherical portion of the surface will be zero (as it is infinitely far away from the soure at $\boldsymbol{r}'$) and we are left only with the intergal over the grating boundary. This gives the first and second Rayleigh integrals

$$\tilde{E}\left(\boldsymbol{r}'\right) = -\frac{1}{4\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\tilde{E}\boldsymbol{n}\cdot\boldsymbol{\nabla}G_D dxdy \tag{2.191}$$

and

$$\tilde{E}\left(\boldsymbol{r}'\right) = \frac{1}{4\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G_N \boldsymbol{n}\cdot\boldsymbol{\nabla}\tilde{E}dxdy, \tag{2.192}$$

where A is boundary of the grating. If we now add the assumption that the grating transmission can be given as function of the incident field, that is,

$$\tilde{E}|_{z=0} = t_{GD}\left(x,y\right)\tilde{E}^{in}|_{z=0}, \tag{2.193}$$

the first and second Rayleigh-Sommerfeld diffraction integrals are given by

$$\begin{aligned}\tilde{E}\left(\boldsymbol{r}'\right) &= -\frac{1}{4\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} t_G\left(x,y\right)\tilde{E}^{in}\frac{\partial G_D}{\partial z}dxdy \\ &= -\frac{z}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} t_G\left(x,y\right)\tilde{E}^{in}\left(ik-\frac{1}{R}\right)\frac{e^{ikR}}{R^2}dxdy\end{aligned} \tag{2.194a}$$

and

$$\begin{aligned}\tilde{E}\left(\boldsymbol{r}'\right) &= \frac{1}{4\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G_N\frac{\partial}{\partial z}\left[t_G\left(x,y\right)\tilde{E}^{in}\right]dxdy \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} t_G\left(x,y\right)\frac{e^{ikR}}{R}\frac{\partial\tilde{E}^{in}}{\partial z}dxdy,\end{aligned} \tag{2.194b}$$

where $R = \sqrt{\left(x'-x\right)^2+\left(y'-y\right)^2+z^2}$. Whilst the Rayleigh diffraction integrals are an exact solution of the boundary value problem, the determination of the transmitted field at $z=0$ is not a trivial problem, and the Rayleigh-Sommerfeld integrals are only as accurate as the assumption in (2.193).

2.7.3 Fraunhofer and Raman-Nath Diffraction

As already mentioned, the Rayleigh-Sommerfeld integrals are, in general, very difficult to solve. However, under certain conditions we can make approximations within the integral that make it a far easier problem to solve. We will consider the problem of calculating the diffracted field at a plane that is parallel to the plane of the emerging field and separated by a distance z.

We define the variable

$$\rho = \sqrt{\left(x'-x\right)^2+\left(y'-y^2\right)} \tag{2.195}$$

which gives

$$R^2 = \rho^2 + z^2. \tag{2.196}$$

If we assume that ρ is sufficiently small compared to z

$$R \approx z\left(1+\frac{\rho^2}{2z^2}\right) \tag{2.197}$$

and

$$\frac{1}{R} \approx \frac{1}{z}\left(1 - \frac{\rho^2}{2z^2}\right). \tag{2.198}$$

This is known as the paraxial approximation; we consider only the points $\boldsymbol{r}'$ that make a small angle (z/R) with the normal (z) axis to the grating. Keeping terms that are at most $O(\rho/z)$ the first Rayleigh-Sommerfeld integral becomes

$$\tilde{E}(\boldsymbol{r}') \approx -\frac{e^{ikz}}{2\pi z}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} t_G(x,y)\,\tilde{E}^{in}\left(ik - \frac{1}{z}\right)e^{\frac{ik}{2z}\rho^2}dxdy. \tag{2.199}$$

If we also assume that we are in the far field, that is, $z \gg \lambda$, this can be further simplified to

$$\tilde{E}(\boldsymbol{r}') \approx \frac{-ie^{ik\left(z+\frac{x'^2+y'^2}{2z}\right)}}{\lambda z}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} t_G(x,y)\,\tilde{E}^{in}e^{-\frac{ik}{z}(x'x+y'y)}dxdy, \tag{2.200}$$

which is known as the Fraunhofer diffraction integral. We recognise that this integral is the Fourier transform of the emerging field $t_G(x,y)\,\tilde{E}^{in}$ with respect to the frequencies

$$f_x = \frac{x'}{\lambda z} \tag{2.201}$$

and

$$f_y = \frac{y'}{\lambda z}. \tag{2.202}$$

This can greatly simplify the calculation of the diffraction pattern in the paraxial/far field regime. The pattern that can be seen and measured is the intensity, or irrandiance, which is just

$$\left|\tilde{E}(\boldsymbol{r}')\right|^2 = \left|\mathcal{F}\left\{\tilde{E}(\boldsymbol{r}')\right\}\right|^2. \tag{2.203}$$

In this way, the plethora of existing Fourier theory can often make very complicated diffraction calculations relatively simple and easy.

A particular example of this is the Raman-Nath diffraction formula. This concerns diffraction from a medium with a sinusoidally varying refactive index or, equivalently, any structure that introduces a sinusoidal modulation of the transmitted phase. We once again let the grating lie in the $z = 0$ plane and then write the refractive index as

$$n(x) = n_{avg} + \Delta n\cos\left(\frac{2\pi x}{\Lambda}\right). \tag{2.204}$$

If we first consider the case of normal incidence, the modulation of the emerging field can be described by

$$\tilde{E}^{out} \propto e^{ik_0 n(x)L}, \tag{2.205}$$

where L is the thickness of the grating. Using the Jacobi-Anger identity

$$e^{iz\cos\theta} = \sum_{m=-\infty}^{\infty} i^m J_m(z)\, e^{im\theta} \tag{2.206}$$

we find

$$\tilde{E}^{out} = e^{ik_0 n_{avg} L} \sum_{m=-\infty}^{\infty} i^m J_m(k_0 \Delta n L)\, e^{\frac{2\pi i m x}{\Lambda}}, \tag{2.207}$$

where J_n is the n'th order Bessel function of the first kind. The Fourier transform is then

$$\mathcal{F}\left\{\tilde{E}^{out}\right\} = e^{ik_0 n_{avg} L} \sum_{m=-\infty}^{\infty} i^m J_m(k_0 \Delta n L)\, \delta\left(f_x - \frac{m}{\Lambda}\right). \tag{2.208}$$

The $\delta\left(f_x - m\Lambda^{-1}\right)$ we have, in essence, seen before, they are just the grating equation (2.182), giving the different diffracted orders whose tangential wavevector components are integer multiples of the grating wavevector. Now, however, we also have the magnitude as well as the direction of the orders. In this case the power into the $m'th$ order is simply given by

$$P_m = J_m(k_0 \Delta n L)^2. \tag{2.209}$$

2.7.4 Non-Paraxial Diffraction and the Angular Spectrum

Whilst the Fraunhofer formulation is very simple and powerful it does suffer from serious limitations on the systems in which it is applicable. In situations such as diffraction from a large (with respect to wavelength) aperture, we are taking the Fourier transform of a function with wide support and, as we know from standard theory, the resulting transform is narrow. To give some intuition as to what this means, consider the Fraunhofer integral (2.200) under a change of variables

$$\hat{x} = \frac{x}{\lambda}, \quad \hat{y} = \frac{y}{\lambda}, \quad \hat{z} = \frac{z}{\lambda}. \tag{2.210}$$

We then have a transform of the form

$$\mathcal{F}\{E\} \propto \lambda^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E e^{-2\pi i(\alpha\hat{x}+\beta\hat{y})} d\hat{x} d\hat{y}, \tag{2.211}$$

where

$$\alpha = \frac{\hat{x}}{\hat{z}}, \quad \beta = \frac{\hat{y}}{\hat{z}}, \tag{2.212}$$

are the direction cosines that the wavevector of the plane wave $e^{2\pi i(\alpha x+\beta y+\gamma z)}$ makes with the x and y axes. Thus, in this way, the Fourier transform decomposes the emerging field into its plane wave components and gives what is called its angular spectrum. In the case of the large aperture this means the diffracted field is composed

of plane waves propagating at small angles, i.e., we are in the paraxial regime and the Fraunhofer formula is a perfect choice.

In general, however, this will not be the case. As we know from our grating equation (2.182), when the grating space Λ becomes smaller and the grating wavevector $\boldsymbol{k}_\Lambda \propto \Lambda^{-1}$ becomes larger, the wavevectors of the diffracted orders become more spread out. The diffracted field then becomes further and further form the paraxial regime and the Fraunhofer formula will quickly lose accuracy.

There is also the case to consider where the incident beam is at an angle. Consider the Raman-Nath example from the previous section. If we consider a wave incident at an angle α_0, we would introduce a linear phase variation of $e^{ik\alpha_0 x}$ across both the incident, and hence emerging, fields. By the standard shift theorem of Fourier theory, the Fourier transform, and hence the irradiance pattern, would also be shifted by α_0, but there would be no change in the power of each order. This linear shift invariance can be very useful if α_0 is small and we remain in the paraxial regime. It is one of the reasons the Fraunhofer approach is often so simple and effective. However, for large α_0 this prediction is in direct contradiction with experimental observations, where the relative intensities of the orders are attenuated as their angle of propagation becomes steeper.

This is all problematic. Later on, when dealing with diffraction from PAAD gratings, we will need to consider gratings with a spacing of only a few wavelengths and a large range of incident angles. We will in no way be close to the paraxial regime and a more suitable method for modelling the diffraction will be necessary. We consider instead the method of [16], which shows that for non-paraxial systems it is actually the diffracted radiance, and not the irradiance, that is found by the Fourier transform of the emerging field. We make some modifications to this method and its derivation so that it is valid for infinite periodic structures.

Irradiance is defined as

$$E_e = \frac{\partial P}{\partial A_c}, \tag{2.213}$$

the power (P) per area of collecting surface (A_c). Radiance is defined as

$$L = \frac{\partial^2 P}{\partial \omega_c \partial A_s \cos\theta_s}, \tag{2.214}$$

the power per solid angle of collecting surface $(\partial\omega_c)$ per projected source area $(\partial A_s \cos\theta_s)$ visible from the collector, where ∂A_s is the area of the source and θ_s is the angle from the source normal. Figure (2.5) shows a schematic depicting the various quantities. We also define the radiant intensity

$$I = \frac{\partial P}{\partial \omega_c}, \tag{2.215}$$

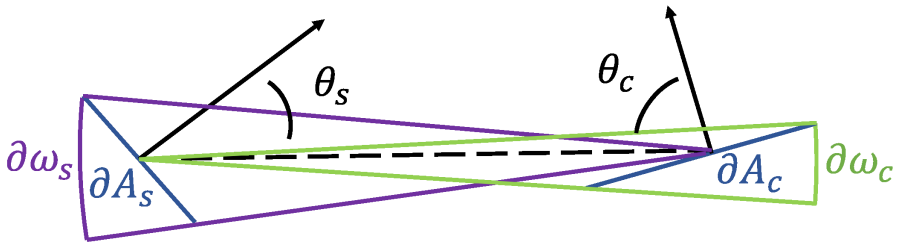


FIGURE 2.5: A schematic showing the various quantities used to define irradiance, radiance, and radiant intensity.

the power per solid angle of collecting surface. The radiant intensity is associated with a point source while the radiance is the analogue for an extended source, which can be thought of as a continuum of point sources. When dealing with an infinitely extending grating, the notion of projected area loses meaning, instead it makes sense to deal with the average radiance

$$\hat{L} = \frac{\partial^2 P}{\partial\omega_c \partial \cos\theta_s}. \tag{2.216}$$

Using (2.215) we see immediately that, for an infinitesimal portion of the grating ∂A_s, the total power into a hemisphere H centred on ∂A_s is given by

$$P = \int_H I \partial\omega_c \tag{2.217}$$

$$= \int_0^{\pi} \int_{-\pi/2}^{\pi/2} I(\theta, \phi) \sin\phi d\phi d\theta. \tag{2.218}$$

The direction cosines (with the x, y, and z axes) of any point on the sphere are given by

$$\alpha = \sin\phi\cos\theta, \tag{2.219}$$

$$\beta = \sin\phi\sin\theta, \tag{2.220}$$

$$\gamma = \cos\phi, \tag{2.221}$$

and we find

$$P = \int_{-1}^{1} \int_{-\sqrt{1-\alpha^2}}^{\sqrt{1-\alpha^2}} \frac{I(\alpha, \beta)}{\gamma} d\beta d\alpha. \tag{2.222}$$

Now, for a finite energy signal, i.e, a non-periodic function, we can equate this with the total power in the emerging field via Parseval's theorem and obtain the results of [16], namely, that

$$\left| \mathcal{F}\left\{ t_G(\hat{x}, \hat{y}) \tilde{E}^{in}(\hat{x}, \hat{y}) \right\} \right|^2 = \frac{\lambda^2}{A_s} L(\alpha, \beta). \tag{2.223}$$

This shows that it is the radiance that is given by the Fourier transform of the emerging field, and not the irradiance as used in the Fraunhofer approximation. We

can then say that the total power is given by

$$\int_{-1}^{1}\int_{-\sqrt{1-\alpha^2}}^{\sqrt{1-\alpha^2}}\left|\mathcal{F}\left\{t_G\left(\hat{x},\hat{y}\right)\tilde{E}^{in}\left(\hat{x},\hat{y}\right)\right\}\right|^2 d\beta d\alpha = \int_{-1}^{1}\int_{-\sqrt{1-\alpha^2}}^{\sqrt{1-\alpha^2}}\frac{\lambda^2}{A_s}L\left(\alpha,\beta\right)d\beta d\alpha, \quad (2.224)$$

which must be the same as the total power in (2.222), and we conclude that

$$I\left(\alpha,\beta\right) = \frac{\lambda^2}{A_s}\gamma L\left(\alpha,\beta\right). \quad (2.225)$$

It is the radiant intensity, then, that gives the power in the plane wave $e^{2\pi i(\alpha\tilde{x}+\beta\tilde{y})}$, and for the m'th diffracted order characterised by its propagation cosines α_m and β_m, the diffraction efficieny is $I\left(\alpha_m,\beta_m\right)$. The factor of γ that is missed in the Fraunhofer approximation is exactly what is responsible for the attenuation of power in non-paraxial diffraction. All of the standard Fourier methods and results can easily be applied to non-paraxial systems if we simply simply include this cosine factor.

We now consider the case of an infinite, periodic grating, and, for now, assume a normally incident field. The signal of the emerging field is also infinite and periodic, and is therefore not a finite energy signal. It is not square integrable, nor is its Fourier transform, because it consists of a number of $\delta-$functions. We must instead consider the average power of the signal. We can do so with the autocorrelation, defined by

$$\rho_{\tau\xi}\left(g\right) = \frac{1}{ST}\int_0^T\int_0^S g^*\left(t,s\right)g\left(t+\tau,s+\xi\right)dtds \quad (2.226)$$

for a periodic signal g with periods T and S. The Wiener-Khintchin theorem states that [48] S_{f_t,f_s}, the spectral power density of a signal, is related to $\rho_{\tau\xi}$ by

$$S_{f_t,f_s}\left(g\right) = \mathcal{F}\left\{\rho_{\tau\xi}\left(g\right)\right\}, \quad (2.227)$$

where f_t and f_s are the spatial frequencies contained within g. It gives the power of each frequency in a signal, and in the case of a finite energy signal it is exactly the modulus squared of the Fourier transform. Now, we need to make an important distinction. The power, as defined in (2.226), is the amplitude squared of each frequency component in a given signal. Applying this to the emerging electric field, the autocorrelation has units of V^2 and the power spectral density has units of $\mathrm{V}^2\omega_s^{-1}$. This is not the same as the electromagnetic power. By considering the discussion of the electromagnetic field energy and the units of the $\boldsymbol{E}$ and $\boldsymbol{H}$ fields in section 2.3, we see that a better choice is to consider the cross-correlation, defined by

$$C_{\tau\xi}\left(g,h\right) = \frac{1}{ST}\int_0^T\int_0^S g^*\left(t,s\right)h\left(t+\tau,s+\xi\right)dtds, \quad (2.228)$$

of the electric and magnetic fields, which would have units of electromagnetic power (W). Since the electric field is composed of a series of plane waves, each with a

corresponding plane wave in the spectrum of the magnetic field (related by the impedance (2.36)), the electromagnetic power spectral density can then be written as

$$\begin{aligned} S^{\varepsilon}_{f_x,f_x}(E) &= \frac{1}{2Z}\mathcal{F}\left\{\rho_{xy}(E)\right\} \\ &= \frac{1}{2Z}S_{f_x,f_y}(E). \end{aligned} \tag{2.229}$$

We return to the total power diffracted into our hemisphere from ∂A_s. This must be the same as the total electromagnetic power contained in the emerging field, and so

$$P = \frac{1}{2Z}\int_{-1}^{1}\int_{-\sqrt{1-\alpha^2}}^{\sqrt{1-\alpha^2}} S_{\alpha\beta}\left(\tilde{t}_G(\hat{x},\hat{y})\,\tilde{E}^{in}(\hat{x},\hat{y})\right)d\beta d\alpha. \tag{2.230}$$

Here we have again normalised our spatial variables with respect to λ so that we are decomposing the autocorrelation into frequencies corresponding to the direction cosines α and β. The limits of integration are restricted to the unit circle in $\alpha\beta$ space since, for now, we will consider the case in which there are no evanescent orders. In general, though, the limits are infinite. Comparing with (2.222), we see that

$$I(\alpha,\beta) = \frac{\gamma}{2Z}S_{\alpha\beta}\left(\tilde{t}_G(\hat{x},\hat{y})\,\tilde{E}^{in}(\hat{x},\hat{y})\right). \tag{2.231}$$

Finally, from equation (2.216), we see that the radiant intensity I can also be obtained by integrating the average radiance over the cosine of the angle at which the radiation leaves the source. Noting that $\cos\theta_s$ is γ, and that the average radiance is constant over the grating, we find

$$\hat{L}(\alpha,\beta) = \frac{1}{2Z}S_{\alpha\beta}\left(\tilde{t}_G(\hat{x},\hat{y})\,\tilde{E}^{in}(\hat{x},\hat{y})\right) \tag{2.232}$$

and, similarly to the finite energy signal case, we have that

$$I(\alpha,\beta) = \gamma\hat{L}(\alpha,\beta). \tag{2.233}$$

Analagously to the non-periodic case, we find the diffraction efficiency of the m'th order is given by $I(\alpha,\beta)$, the product of its average radiance and the cosine of its direction of propagation to the grating normal.

To give this result a footing in more familiar terms, consider a 1D grating in the x direction. The diffracted field can be written as

$$\underline{E} = \left(\sum_{m=-\infty}^{\infty} E_m e^{2\pi i\alpha_m\tilde{x}}\right)\underline{\hat{E}}, \tag{2.234}$$

$$\alpha_m = \alpha_0 + \frac{m}{\tilde{\Lambda}}, \tag{2.235}$$

where $\underline{\hat{E}}$ is the polarisation of the electric field, Λ is the grating spacing, and α_0 is direction cosine of the incident field. The magnetic field is then given by

$$\underline{H} = \frac{1}{Z}\left(\sum_{m=-\infty}^{\infty} E_m e^{2\pi i \alpha_m \tilde{x}} \underline{\hat{H}}_m\right), \tag{2.236}$$

where

$$\underline{\hat{H}}_m = \underline{k}_m \times \underline{\hat{E}}, \tag{2.237}$$

and n is the refractive index. We now consider the average power emerging from the grating by

$$P = \frac{1}{\Lambda}\int_0^{\Lambda} \langle \underline{S} \rangle \cdot \underline{n} dx, \tag{2.238}$$

which, after non-dimensionalising, can be written as

$$P = \frac{1}{2Z\tilde{\Lambda}}\int_0^{\tilde{\Lambda}} \mathrm{Re}\left\{\left(\sum_{m=-\infty}^{\infty} E_m e^{2\pi i \alpha_m \tilde{x}}\right)\underline{\hat{E}} \times \left(\sum_{l=-\infty}^{\infty} E_l^* e^{-2\pi i \alpha_m \tilde{x}} \underline{\hat{H}}_l{}^*\right)\right\} \cdot \underline{n} d\tilde{x}. \tag{2.239}$$

The general term is

$$P_{nm} = \frac{1}{2Z\tilde{\Lambda}}\int_0^{\tilde{\Lambda}} \mathrm{Re}\left\{\tilde{E}_m \tilde{E}_l^* e^{2\pi i(\alpha_m - \alpha_l)\tilde{x}} \underline{\hat{k}}_l^*\right\} \cdot \underline{n} d\tilde{x}, \tag{2.240}$$

and we must complete the integral

$$\begin{aligned}\int_0^{\tilde{\Lambda}} e^{2\pi i(\alpha_m - \alpha_l)\tilde{x}} d\tilde{x} &= \int_0^{\tilde{\Lambda}} e^{2\pi i(m-l)\frac{\tilde{x}}{\tilde{\Lambda}}} d\tilde{x} \\ &= \frac{\tilde{\Lambda}}{2\pi(m-l)}\left[e^{2\pi i(m-l)\frac{\tilde{x}}{\tilde{\Lambda}}}\right]_0^{\tilde{\Lambda}} \\ &= \tilde{\Lambda}\delta(m-l).\end{aligned} \tag{2.241}$$

Putting all the terms together, we are left with (assuming no evanescent modes)

$$P = \frac{1}{2Z}\sum_{m=-\infty}^{\infty} |E_m|^2 \gamma_m. \tag{2.242}$$

This is just the equation

$$\begin{aligned}P &= \frac{1}{2Z}\int_{-1}^{1} S_\alpha\left(\tilde{E}\right)\gamma d\alpha \\ &= \int_{-1}^{1} \hat{L}(\alpha)\gamma d\alpha, \\ &= \int_{-1}^{1} I(\alpha) d\alpha.\end{aligned} \tag{2.243}$$

We see, then, that the radiance $\hat{L}(\alpha)$ is just $|\langle \mathbf{S}_\alpha \rangle|$, the magnitude of the time averaged Poynting vector associated with the plane wave component $e^{2\pi i \alpha \tilde{x}}$, and that the radiant

intensity/diffraction efficiency is just the flux $\langle \boldsymbol{S}_\alpha \rangle \cdot \boldsymbol{n}$.

We now consider that the field incident upon the grating is at an oblique angle α_0, and the emerging field is consequently shifted by a factor of $e^{2\pi i \alpha_0 \tilde{x}}$. From the definition (2.226) we see that

$$\begin{aligned}
\rho_X \left(\tilde{E} e^{2\pi i \alpha_0 \tilde{x}} \right) &= \frac{1}{\tilde{\Lambda}} \int_0^{\tilde{\Lambda}} \tilde{E}^* \left(\tilde{x} \right) e^{-2\pi i \alpha_0 \tilde{x}} \tilde{E} \left(\tilde{x} + X \right) e^{2\pi i \alpha_0 (\tilde{x} + X)} d\tilde{x} \\
&= e^{2\pi i \alpha_0 X} \frac{1}{\tilde{\Lambda}} \int_0^{\tilde{\Lambda}} \tilde{E}^* \left(\tilde{x} \right) \tilde{E} \left(\tilde{x} + X \right) d\tilde{x} \\
&= e^{2\pi i \alpha_0 X} \rho_X \left(\tilde{E} \right)
\end{aligned} \tag{2.244}$$

Taking the Fourier transform we find the radiance of the shifted field, $\check{L}\left(\alpha\right)$, is

$$\check{L}\left(\alpha\right) = \frac{1}{2Z} S_\alpha \left(\tilde{E} e^{2\pi i \alpha_0 \tilde{x}} \right) \tag{2.245}$$

$$= \frac{1}{2Z} S_\alpha \left(\tilde{E} \right) * \delta \left(\alpha - \alpha_0 \right) \tag{2.246}$$

$$= \hat{L} \left(\alpha - \alpha_0 \right), \tag{2.247}$$

where $\hat{L}$ is the radiance of the original un-shifted field. The radiance is thus a linear shift invariant quantity. It is only the distance between the diffracted orders and the fundamental order in direction cosine space that affects their radiance. These distances are determined by the geometry of the grating and, as such, the radiance is a fundamental property of a grating that does not change. On the other hand, the intensities of the orders do change, being attenuated according to the normal part of their wavevectors, as in (2.231).

Lastly, we consider the case where evanescent orders are present, and that not all of the energy in the field can propagate away from the grating. There must be a redistribution of energy into the propagating orders so that energy does not accumulate in the grating. In [16] the authors assert that the radiance should be renormalised according to

$$\hat{L}' = K \hat{L}, \tag{2.248}$$

where

$$K = \begin{cases} \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{L} d\alpha d\beta}{\int_{-1}^{1} \int_{-\sqrt{1-\alpha^2}}^{\sqrt{1-\alpha^2}} \hat{L} d\beta d\alpha} & \alpha^2 + \beta^2 \leq 1 \\ 0 & \text{otherwise.} \end{cases} \tag{2.249}$$

In this way, all the energy in the emerging field is sent into the propagating orders. The integrals (2.222) and (2.230) are then over the same limits and we are able to equate their integrands, as in (2.231). However, the effect on the radiance, which is linear shift invariant, can only be a constant rescaling of the entire spectrum. Rescaling the radiance by a function of α, β would have the effect of convolving the autocorrelation of the emerging field with the inverse Fourier transform of the rescaling function.

Whilst the inverse of the autocorrelation is not unique up to phase shifts of the different frequencies, changing the amplitudes would completely change the form of the emerging field, implying a physical change of the grating. Even if the above rescaling was applied to the entire spectrum, it is still clear to see that it is incorrect. In the case of total internal reflection at an isotropic boundary, the transmitted energy would just be set to zero. This cannot be the case since we know the evanescent order exists specifically to ensure that energy is conserved instantaneously.

By the earlier discussion, a natural solution to this problem is to instead consider the time average of the energy out of the grating, which we know is zero for the evanescent orders. It follows that to correctly describe the system in the presence of evanescent waves, we must balance the average energy into the grating with the average energy out (both transmitted and reflected). Now, we have already concluded that the radiance can only be uniformly rescaled. This means that the renormalised radiance of the transmitted field must be of the form

$$\hat{L}' = (1 + Q)\,\hat{L}, \tag{2.250}$$

where Q is some constant that is at most the sum of the evanescent energies in the transmitted and reflected fields. The amount of this evanescent energy is easy to bound. Unless the material after the grating has a lower refractive index than the incident material, the fundamental order of the transmitted field must always propagate. The fundamental order of the reflected field must also always propagate. It follows that Q is at most $O(\eta_{\pm 1})$, where $\eta_{\pm 1}$ is the first order diffraction efficiencies, and any redistribution of energy into the diffracted orders is at most $O\left(\eta_{\pm 1}^2\right)$. In the experimental systems that we seek to model later, the diffraction efficiencies of the first orders are very small, of the order $O(10^{-4})$. It follows that we should be able to safely neglect modelling these effects, and we do not worry about correctly accounting for this renormalisation.

Chapter 3

Modelling Diffraction in Thin PAAD Films

3.1 Introduction

We arrive now at the part of this thesis pertaining to the modelling of the photoaligned PAAD layers. To begin with, this work concerns the creation of optically controlled, dynamic half-wave plates. These are based on a twisted, nematic liquid crystal cell, and controlled using the photo-aligning properties of PAAD [43]. The modelling involved was to predict the expected transmission of various wavelengths through the cell in its twisted state, for comparison with the experimental results. The results raised some interesting questions about the nature of the photoalignment process, and so we discuss the results of the old diffraction efficiency experiments [35], in which diffraction efficiency measurements were made for the s- and p-polarisation at a single angle of incidence. The results of this were inconclusive. If anything, they raised more questions, being in direct contradiction with the standard model for realignment in thick layers [11].

The main work of this part of the thesis then begins by outlining the new experimental diffraction efficiency results obtained by our research group. My modelling, outlined here, is developed in tandem in the hope of fitting the model to the experimental results, and determining the photoalignment behaviour of the thin PAAD layers. First, a new method for the modelling of light propagation through multiple layers is presented, known as the Iterated Ray method. For a three layer system, say, a layer of glass in air, a standard textbook example shows that by considering the infinite number of reflected rays occuring within the glass, the total transmitted field, along with all its resonant effects, can be found by summing the transmission of all of the reflections. As far as I am aware, this methodology has not been extended beyond three layers systems. In such cases, the global standard is to adopt either a T- or

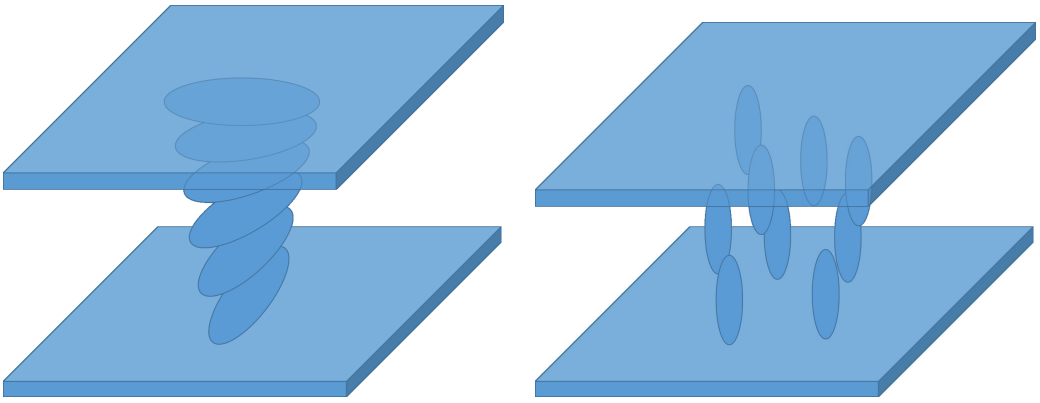
S-Matrix approach. The Iterated ray method works by replacing the $N-1$ layer of the N-layer system with its effective reflection and transmission coefficients from layer $N-2$ to N, thus reducing it to an effective boundary and leaving an $(N-1)$-layer system. The the $N-2$ layer is similarly reduced to an effective boundary, and so on and so forth until we have a single effective boundary between layer 1 and layer N. It is found to have the accuracy and stability of the current leading choice method, the S-matrix, but proves to be substantially more computationally efficient. The results of this are to be submitted to the Journal of the Optical Society of America A.

The Iterated Ray method is also intuitively extendable to multimode systems with mode mixing. This is used to describe the transmission across the anistropic aligned PAAD as a function of its optical axis and birefringence. The results of this are found to be in virtually exact agreement with numerical simulations of an anistropic layer. The results of the theory discussion on non-paraxial scalar diffraction are then coupled to the PAAD layer coefficients to describe the diffracted field caused by the periodically modulated PAAD. This leads to the generation of six modes, an s- and p-polarised wave for each of the 0 and ± 1 diffracted orders. Finally, the model is completed by utilising the multi-mode extension of the Iterated Ray method to describe the propagation across the entire multilayer system, including the repeat diffraction each time a reflected ray is incident upon the grating layer again. Final comparison of the experimental results, along with other results from one of our group's collaborators, is found to provide some consistent results and insight into the thin PAAD photoalignment behaviour.

3.2 Optically Controlled, Dynamic Wave Plates

3.2.1 The Nematic Twisted Cell and PAAD Alignment Layer

Having already discussed the advantages of an optically controlled system, we now present the work on the creation of dynamic liquid crystal waveplates. The waveplates are based on the twisted nematic cell, wherein the liquid crystal alignment at the opposite cell faces is coplanar (flat against the cell surface) but not parallel. The bulk of the liquid crystal then assumes a twisted, helical structure that rotates uniformly between the orientations prescribed at the two surfaces. This is depicted on the left of figure 3.1. In this state the polarisation of incident light rotates with the helical structure (provided the wavelength is "small enough" compared to the pitch, or twisting rate, of the liquid crystal). This is due to the incident light resolving into o- and e-waves which arrive at the far side of the cell out of phase with one another. The phase difference then means they recombine into a rotated polarisation upon exiting the cell. The right hand illustration shows the cell under the influence of an applied electric field (normal to the cell surfaces). The bulk of the molecules align along the



(A) (B)

FIGURE 3.1: Schematic of the bulk alignment of a nematic, twisted cell in its (a) default helical state induced by the surface alignments with no applied field and (b) the reoriented state under an applied voltage.

electric field and the helical structure is broken. As a result, the (normally) incident light sees only the ordinary refractive index, and its polarisation remains unchanged. If the cell is placed between crossed-polarisers that are parallel to the surface alignments, the light transmits through the cell in its twisted state, but does not transmit when the voltage is applied. As mentioned in the introduction, this system is limited by the surface alignments; it can only switch between the on and off (no rotation or $\pi/2$ rotation) states, and requires the continuous application of an external field to remain in the off state.

The optically controlled waveplates operate by replacing one of the rubbed alignment layers with PAAD. It is well known that, usually, the PAAD molecules reorient perpendicular to the polarisation of incident light [11, 4, 20]. This is a consequence of the fact that light is absorbed preferentially by molecules that are aligned parallel to the electric field. Upon absorbing a photon, the molecules transform from their trans state to higher energy cis state and, upon returning, assume a random orientation. The net effect over time is a collection of molecules that are all perpendicular to the probe polarisation, since these are non-absorbing. The application of a polarised external light source, then, allows for the surface alignment direction to be changed at will. The waveplates can thus operate at any twist angle between 0 and $\pi/2$ and, once the alignment direction is set, require no more power.

3.2.2 Model and Results

The experiments on these PAAD controlled waveplates involved monitoring the transmission of a probe beam through the cell for multiple, successive twist and untwist steps[43]. The experimental setup is shown in figure 3.2. For comparison, the expected transmission intensities were calculated using a formula derived from Jones

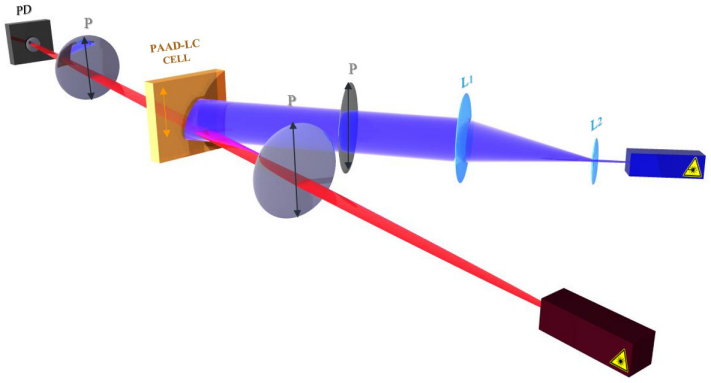


FIGURE 3.2: A schematic of the twist/untwist experimental setup. The blue beam and polariser are used to align the PAAD alignment layer on the near face of the cell. The cell is then probed with the red beam through parallel polarisers for successive twisted and untwisted alignments.

matrices [59]. A Jones matrix is a 2x2 matrix that describes the change in the spatial components of an electric field (perpendicular to the direction of propagation) due to an active optical element. By approximating the twisted cell as a series of uniform slices, each with an associated Jones matrix describing a small rotation of the polarisation, the final formula is given by the product of these matrices in the limit that the thickness of the slices goes to zero. The field at output, $\boldsymbol{E}_{out}$, is then given by

$$\boldsymbol{E}_o = \begin{pmatrix} \cos\Phi & -\sin\Phi \\ \sin\Phi & \cos\Phi \end{pmatrix} \begin{pmatrix} \cos\Theta - i\frac{\Gamma}{2\Theta}\sin\Theta & \frac{\sin\Theta}{\Theta}\Phi \\ -\frac{\sin\Theta}{\Theta}\Phi & \cos\Theta + i\frac{\Gamma}{2\Theta}\sin\Theta \end{pmatrix} \boldsymbol{E}_{in}, \quad (3.2.2.1)$$

where $\boldsymbol{E}_{in}$ is the incident electric field, Φ is the total twist angle of the liquid crystal, Θ is defined by

$$\Theta \equiv \left(\Phi^2 + \left(\frac{\Gamma}{2}\right)^2\right)^{\frac{1}{2}}, \quad (3.2.2.2)$$

and

$$\Gamma = \frac{2\pi}{\lambda}\left(n_e - n_o\right)L. \quad (3.2.2.3)$$

The quantities n_e and n_o are the extraordinary and ordinary refractive indices, respectively, of the liquid crystal, and L is the thickness of the cell.

The transmission through the cell was tested for a range of wavelengths spanning from green to near infrared; the expected values for the cell in the twist state between parallel polarisers is given in table 3.1. From this we see that, if the cell is perfectly twisted, a near 100% modulation of the incident light should be possible, for a large range of the spectrum. Some of the experimental results for the cell, with the PAAD alignment layer controlled with blue and green light are shown in figure 3.4. Clearly

λ/nm	I
405	0.7%
532	0.9%
632	2.1%
808	5.8%

TABLE 3.1: Expected transmitted intensity measured between parallel polarisers for a 6.5 µm thick, E7 LC cell in the twisted state

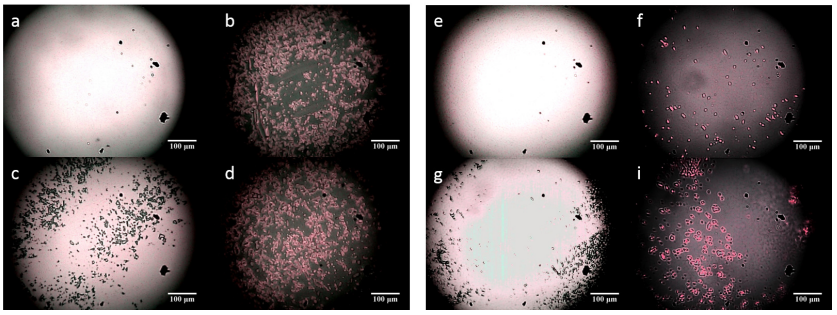


FIGURE 3.3: Microscope images of the PAAD aligned cell through multiple twist and untwist states, aligned with a green (a-d) and blue (e-i) beam.

the alignment produced by the PAAD layer was not perfect, producing lower than expected transmitted intensities. This was evident in microscope images of the cells showing rather patchy alignment shown in figure 3.3.

Although this may not be unexpected (given that the beam will in part be absorbed), the difference between the green and blue beams is. They are both absorbed by the PAAD (although the green not as well, shown in figure 3.5) and so you would expect that the green beam experiments would behave the same as the blue given enough time. Further experiments showed a similar difference in transmitted intensities between the blue beam and the same blue beam after being expanded. We would again expect the expanded blue beam (with a corresponding drop in intensity) to reach the same asymptotic intensities given enough time, but these were found to be significantly lower. This suggests that there are some more complicated dynamics in the realignment of the coupled liquid crystal/PAAD system. There is also the question of the origin of the transient behaviour in the first twist and untwist steps (which can be seen in figure 3.4) , which behave differently to the subsequent, and very repeatable steps. To begin gaining a better understanding of the system, the photoalignment of a standalone thin PAAD film, aligned using visible light, was considered.

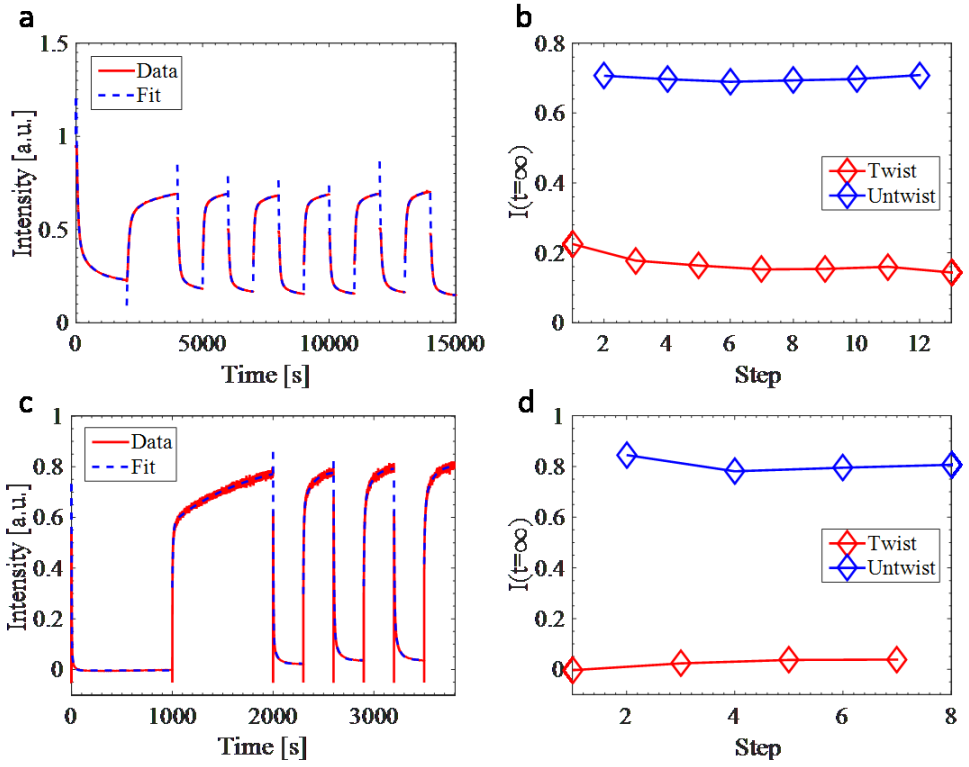


FIGURE 3.4: Transmitted intensity data for green (a) and blue (c) beams for repeated twist and untwist steps. (b) and (d) show the asymptotic intensities at each step for the green and blue beams, respectively.

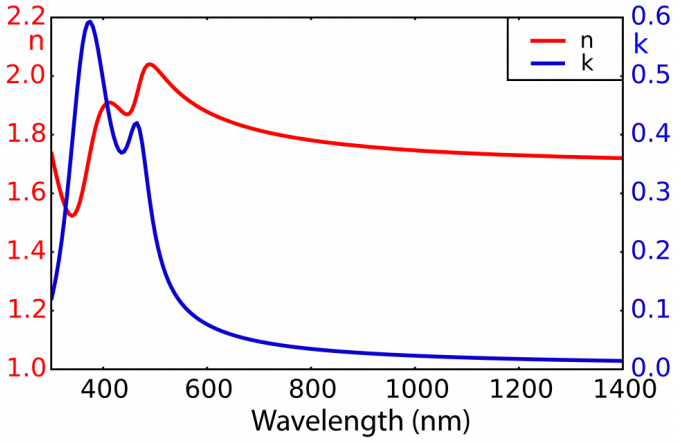


FIGURE 3.5: Complex refractive index of the PAAD as measured in [35].

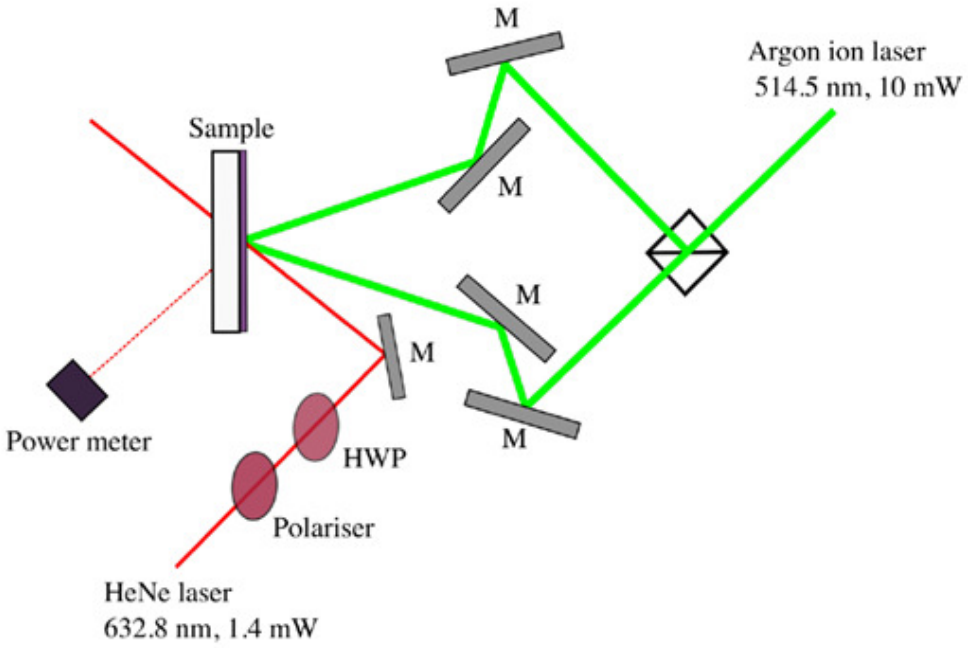


FIGURE 3.6: Schematic diagram of the diffraction efficiency set up. The green pump beam forms an interference pattern on the PAAD, inducing a photoaligned grating. The diffraction efficiency of the grating is measured using the red probe beam.

3.3 Experimental Photoalignment of Thin PAAD Films

3.3.1 The Experimental Setup and Results

Without the presence of the liquid crystal to modulate the probe beam, a different method was needed to test the photoalignment of the PAAD, and so the measurement of the diffraction efficiency of photo-recorded gratings was used [35]. The experiments used thin films of PAAD 22D, N, and E, deposited on a substrate by spin coating. The D, N, and E films had thicknesses of (20 ± 5) nm, (35 ± 5) nm, and (35 ± 15) nm, respectively, and were found to have no periodic surface relief. An interference pattern between the two recording beams was used to form a periodic illumination of the PAAD layers, with a grating period of 1µm. The grating was then probed with a laser at 632.8 nm incident at 18.5° for polarisations both parallel and perpendicular to the plane of incidence. The experimental setup is shown in figure 3.6 and the results are shown in table 3.2.

Given that the gratings were so thin, they were modelled in the Raman-Nath regime, meaning the expected diffraction efficiencies can be expressed as [47]

$$\eta = J_1^2\left(\frac{2\pi\Delta n d}{\lambda\cos\theta}\right), \tag{3.3.1.1}$$

Material	Thickness / nm	$\eta_\perp$	$\Delta n\ \perp$	$\eta_\parallel$	$\Delta n_\parallel$	R
22D	20 ± 5	3×10^{-6}	17×10^{-3}	6.1×10^{-8}	2.4×10^{-3}	7
22N	35 ± 5	3×10^{-5}	30×10^{-3}	2.1×10^{-7}	2.5×10^{-3}	12
22E	35 ± 15	1×10^{-6}	5.5×10^{-3}	5.5×10^{-8}	1.3×10^{-3}	4

TABLE 3.2: Experimental results for the measured diffraction efficiencies $\eta_\perp$ and $\eta_\parallel$, corresponding to probe polarisations perpendicular and parallel to the plane of incidence. $\Delta n_\perp$ and $\Delta n_\parallel$ are the calculated birefringence seen by the two polarisations [35].

where J_1 is the first order Bessel function of the first kind, Δn is the amplitude of the refractive index modulation, d the layer thickness, λ the probe wavelength, and θ the angle of incidence. The measured diffraction efficiencies were very small (of order 10^{-6}) where the Bessel functions behave as $J_1(x) \approx x/2$, and so the approximation is made that

$$\Delta n = \frac{\lambda \cos\theta}{\pi d}\sqrt{\eta}. \tag{3.3.1.2}$$

Using this approximation the respective $\Delta n_\perp$ and $\Delta n_\parallel$ for probe beams polarised perpendicular and parallel to the plane of incidence were calculated, which are shown in table 3.2. Calculating the ratio

$$R = \Delta n_\perp / \Delta n_\parallel \tag{3.3.1.3}$$

for PAAD 22E, 22D , and 22N gives 4, 7, and 12, respectively. This is inconsistent with the accepted model for the molecular reorientation of PAAD [11], namely, that the molecules align perpendicular to the polarisation of the recording beam. Under this model, the molecules are confined to the plane of incidence, but distributed uniformly. The refractive indices seen by the two polarisations of the probe beam are then

$$n_\parallel = \frac{n_e + n_o}{2} \tag{3.3.1.4}$$

$$n_\perp = n_o, \tag{3.3.1.5}$$

while the PAAD in the non-illuminated regions is isotropic and has refractive index

$$n_I = \frac{n_e + 2n_o}{3}. \tag{3.3.1.6}$$

This gives the concerned ratio as

$$R = \frac{|n_I - n_o|}{\left|n_I - \frac{n_e+n_o}{2}\right|} = 2. \tag{3.3.1.7}$$

This gives the right trend but it is not in agreement with the experimental results, and so we conclude that in these thin films the optical axis does not behave as expected.

3.4 New Experimental Results and Model

To complete a more in depth study of the PAAD molecular reorientation in thin layers, new experimental results were obtained and a new, more comprehensive model developed. PAAD-22E layers were once again deposited onto a substrate and the transmission was measured over a large range of incident angles in both the aligned and unaligned states. The substrate was a glass slide coated with an indium tin oxide (ITO) layer, an optically transparent and conductive material often used for electrode layers in liquid crystal cells. The coating was necessary to obtain a uniform, good quality layer since PAAD did not seem to adhere well to the glass alone. The normalised results for the fundamental s- and p-polarisations, s_0 and p_0, and the diffracted orders, $s_{\pm 1}$ and $p_{\pm 1}$, are shown later in section 3.7 figure 3.18

By developing a model of the system, the various system parameters can be fit to the experimental data with a minimisation routine to try and deduce the alignment of the PAAD optical axis. However, this is not a simple task owing to the large number of unknown parameters in the system. As well as the two angles describing the optical axis, the various refractive indices and thicknesses of the systems materials must also be obtained, which gives, in total, eight parameters to determine. Not only does the large solution space make for very long running times of the minimisation routine, but it also casts the accuracy of the results into question. It is highly likely that any errors in one parameter can easily be adjusted for by tweaking the other seven.

We deal with these problems in two ways. Firstly, we note that by considering subsystems of the full problem, we can first fit for the various thicknesses and isotropic refractive indices of the different materials with much greater ease and accuracy. As a result, further experiments were completed using only the substrate, and then the substrate coated with isotropic PAAD, meaning that we can reduce the number of fitting parameters for the diffraction data to three; the two angles parameterising the optical axis and the birefringence of the PAAD. This makes for a series of far simpler problems which, given some reasonable bounds on the expected values of the parameters, can be solved far quicker than the full problem. The second way that the process is sped up is with the implementation of a ray based method for calculating the transmission across layered systems. This method, discussed in the next section, proves to run much quicker than the T- and S-matrix approaches, has the accuracy of the S-matrix in the presence of evanescent fields, and also extends in an intuitive manner to more complicated systems containing anisotropic layers and diffraction gratings.

3.5 Ray Method for Multilayer Propagation

3.5.1 Transmission Across an Isotropic Layer

We begin the outline of the ray method by considering a system of three layers. The layers 1 and 3 are semi-infinite and sandwich layer 2, of finite thickness, and we are looking to describe the transmitted and reflected fields arising from a plane wave incident upon the layer. This is exactly the same three layer problem as in section 2.6, but this time we will use a different methodology to solve it.

Previously, we solved this problem from an entirely mathematical basis. We knew that each field must be a wavelike solution of Maxwell's equations, and we knew that all fields had to have the same tangential wavevector components as the incident wave (for phase continuity). Making the ansatz that each field could be written as the sum of a forwards and backwards propagating component, we applied the electromagnetic boundary conditions from section 2.5 to form a system of equations relating the amplitudes of the various waves. Solving this system of equations allowed us to write the transmitted and reflected fields as a function of the incident field, which could then be rewritten in terms of the Fresnel coefficients from section 2.5.4.

This time, however, we will use our prior knowledge of how a wave behaves at a single interface (the Fresnel coefficients), and take a more physical approach. By following the interaction of the incident wave and the layered structure forward in time, we can construct the fields in each layer as the sum of infinitely many waves. Each wave, or ray, can be easily described by a product of Fresnel coefficients, the formulae for which we already know, and we can avoid the process of again applying boundary conditions etc. This methodolgy, relying on the superposition of multiple transmitted and reflected rays, gives a much more intuitive understanding of the resonance phenomena of layered structures. It is also considerably more simple (and computationally eefficient) to apply to structures with more than three layers, where the field based methodologies (the T- and S-matrices) become very algebraically involved and computationally expensive.

We return now to the derivation of ray based solution to the three layer problem. The input field, F_1^1, is incident upon layer 1 in the forward direction, where one part of the field is reflected and the other part is transmitted into layer 1. Similarly, the transmitted field is split into reflected and transmitted fields at the interface between layer 2 and 3. In fact, every ray that is incident on an interface is split into two, with each of these split beams also being split each time they encounter an interface. This results in infinitely many rays arising in each layer, as depicted in figure 3.7a. By summing all of the rays in each layer that are in the same direction, we can describe each layer i as having a single forwards and backwards field F_i and B_i, respectively, as in figure 3.7b.

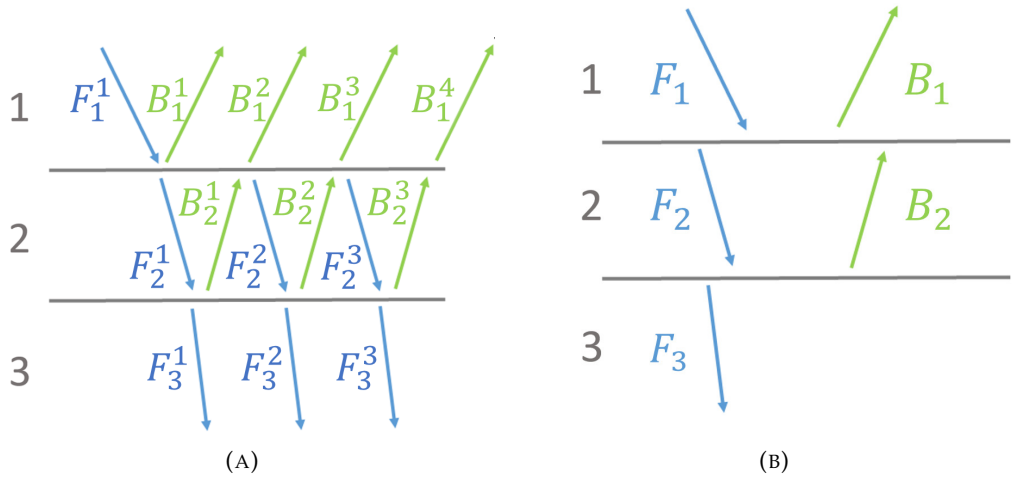


FIGURE 3.7: (A) Schematic of a three layer system, depicting the input field F_0, along with the rays resulting from subsequent reflections and transmissions at the two interfaces. (B) Schematic of a three layer system, depicting the input field, F_0, total reflected field, B_0, and total transmitted field, F_2. $F_1$and B_1 are the sum of all the forward and backward fields inside the centre layer.

The Fresnel coefficients can be used to describe the various rays in the system and calculate the total fields. From figure 3.7a, we can see that the n'th ray to propagate into layer 3 is given by

$$F_3^n = t_{12}t_{23}e^{i\frac{\delta}{2}}\left(r_{23}r_{21}e^{i\delta}\right)^{n-1}F_1, \tag{3.5.1.1}$$

while the n'th backward ray in layer 1 is given by

$$B_1^n \begin{cases} r_{12}F_1 & n = 0 \\ t_{12}r_{23}t_{21}e^{i\delta}\left(r_{21}r_{23}e^{i\delta}\right)^{n-2}F_1 & n > 0. \end{cases} \tag{3.5.1.2}$$

The quantity δ is the phase difference between two subsequent rays. To calculate δ, we consider the diagram in figure 3.8. The phase difference is calculated by the difference in optical path length between the adjacent beams and is given by

$$\begin{aligned} \delta &= k_0\left[\left(\left|\overrightarrow{AB}\right| + \left|\overrightarrow{BC}\right|\right)n_2 - \left|\overrightarrow{AD}\right|n_3\right] \\ &= k_0\left[2L\sec\theta_2 n_2 - 2L\tan\theta_2\sin\theta_3\right], \end{aligned} \tag{3.5.1.3}$$

where L is the thickness of layer 2. Applying Snell's Law then gives

$$\begin{aligned} \delta &= 2k_0 L n_2\cos\theta_2 \\ &= 2q_2 L, \end{aligned} \tag{3.5.1.4}$$

and the phase change is just given by the normal component of the wavevector multiplied by the layer thickness, agreeing with the field formulation of the problem in section 2.6.1.

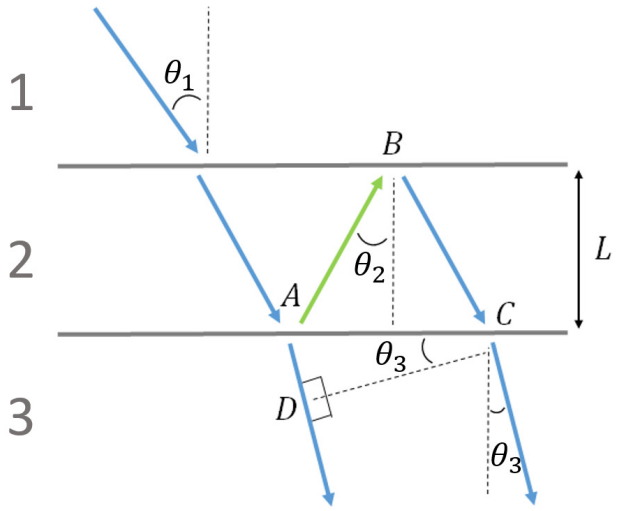


FIGURE 3.8: Diagram of the geometry needed to calculate the difference in optical path length between two subsequent transmitted or reflected rays.

We can now sum all of the transmitted rays to find

$$\begin{aligned} F_3 &= \left[t_{12}t_{23} \sum_{n=0}^{\infty} \left(r_{23}r_{21}e^{i\delta} \right)^n \right] F_1 \\ &= \frac{t_{12}t_{23}e^{i\frac{\delta}{2}}}{1 - r_{23}r_{21}e^{i\delta}} F_1. \end{aligned} \tag{3.5.1.5}$$

Summing all of the reflected rays we find

$$\begin{aligned} B_1 &= \left[r_{12} + t_{12}r_{23}t_{21}e^{i\delta} \sum_{n=0}^{\infty} \left(r_{21}r_{23}e^{i\delta} \right)^n \right] F_1 \\ &= \left[r_{12} + \frac{t_{12}r_{23}t_{21}e^{i\delta}}{1 - r_{23}r_{21}e^{i\delta}} \right] F_1 \\ &= \frac{r_{12} + r_{23}e^{i\delta}}{1 - r_{23}r_{21}e^{i\delta}} F_1 \end{aligned} \tag{3.5.1.6}$$

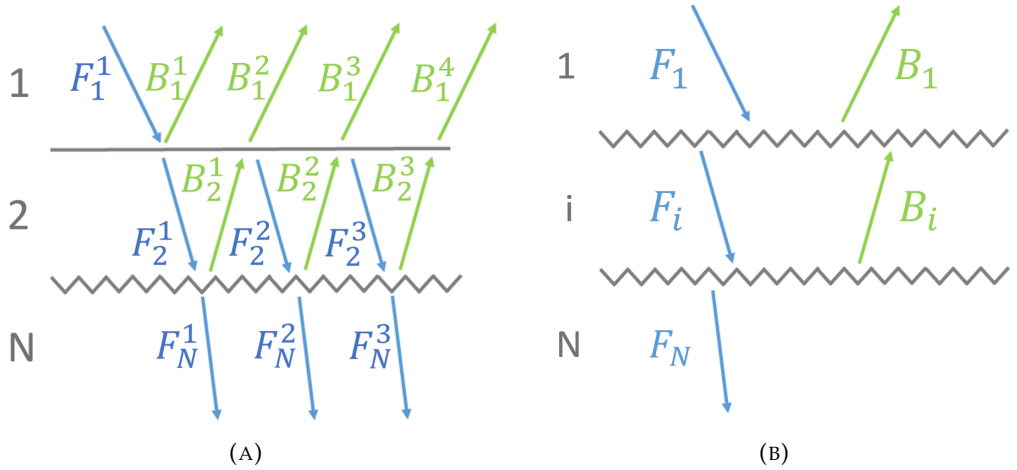
after applying the identity (2.100). In this way we can define the effective transmission coefficient of the system by

$$F_3 = t_{13}F_1 \, , \; t_{13} = \frac{t_{12}t_{23}e^{i\frac{\delta}{2}}}{1 - r_{23}r_{21}e^{i\delta}} \tag{3.5.1.7}$$

and the effective reflection coefficient by

$$B_1 = r_{13}F_1 \, , \; r_{13} = \frac{r_{12} + r_{23}e^{i\delta}}{1 - r_{23}r_{21}e^{i\delta}}. \tag{3.5.1.8}$$

Using the other Stokes' relation (2.101) we see that these are exactly the same as the formulae derived in section (2.6).



(A) (B)

FIGURE 3.9: (a) Schematic representation of the N layer system as a three layer system. The layers 3 to $N-1$ are replaced with an effective boundary, represented by a jagged interface. (b) Schematic representation of the N layer system as a general three layer system. The forwards and backwards fields, F_i and B_i, in any layer i can be evaluated by replacing layers 1 to $i-1$ and $i+1$ to $N-1$ with effective layers.

3.5.2 Transmission Across Multiple Layers - Iterated Ray Method

We now build on the ideas of a three layer system to find the effective coefficients of an N layer system. To do so, we consider the system of N layers as a 3 layer system, replacing the layers 3 to N with an effective boundary between layers 2 and N. This is depicted in figure (3.9a), where a jagged line indicates an effective boundary. Applying the same methodology as before, we derive the effective coefficients of the N layer system as

$$t_{1N} = \frac{t_{12}t_{2N}e^{i\frac{\delta_2}{2}}}{1 - r_{2N}r_{21}e^{i\delta_2}} \tag{3.5.2.1}$$

and

$$r_{1N} = \frac{r_{12} + r_{2N}e^{i\delta_2}}{1 - r_{2N}r_{21}e^{i\delta_2}}, \tag{3.5.2.2}$$

writing the phase difference as δ_2 to make clear that it depends on the refractive index and thickness of layer 2. To calculate the coefficients from layer 1 to N we must know the effective coefficients from layer 2 to N, which will in turn require the effective coefficients from layer 3 to N, and so on. This means that the coefficients can be built inductively, working backwards from the final layer.

To calculate these effective coefficients between two general layers j and k, we consider the same picture as in figure (3.9a), but replace the layers 0, 1, and N with j, $j+1$, and k, respectively. We can then write that, for any j and k in our N layer system (excluding $|j-k| < 2$, in which case we just need the regular Fresnel coefficients)

$$t_{jk} = \frac{t_{j,j+1}t_{j+1,k}e^{i\frac{\delta_{j+1}}{2}}}{1 - r_{j+1,k}r_{j+1,j}e^{i\delta_{j+1}}} \tag{3.5.2.3a}$$

and

$$\begin{aligned} r_{jk} &= r_{j,j+1} + \frac{t_{j,j+1} r_{j+1,k} t_{j+1,j} e^{i\delta_{i+1}}}{1 - r_{j+1,k} r_{j+1,j} e^{i\delta_{j+1}}} \\ &= \frac{r_{j,j+1} + r_{j+1,k} e^{i\delta_{i+1}}}{1 - r_{j+1,k} r_{j+1,j} e^{i\delta_{j+1}}} \end{aligned} \tag{3.5.2.3b}$$

if we are travelling forwards, i.e. $j < k$. If we are travelling backwards, that is, $j > k$,

$$t_{jk} = \frac{t_{j,j-1} t_{j-1,k} e^{i\frac{\delta_{j-1}}{2}}}{1 - r_{j-1,k} r_{j-1,j} e^{i\delta_{j-1}}} \tag{3.5.2.4a}$$

and

$$r_{jk} = \frac{r_{j,j-1} + r_{j-1,k} e^{i\delta_{i-1}}}{1 - r_{j-1,k} r_{j-1,j} e^{i\delta_{j-1}}}. \tag{3.5.2.4b}$$

In both cases, by working iteratively backwards from the final layer, the only terms that appear are the effective transmission and reflection calculated at the previous iteration, and the regular Fresnel coefficients for the new boundary that is added at each step. One simply needs to start with $r_{k-2,k}$ and $t_{k-2,k}$ and work back to r_{jk} and t_{jk} by using equations (3.5.2.3a)-(3.5.2.3b) for forward travel and (3.5.2.4a)-(3.5.2.4b) for backward travel.

Finally, we can also evaluate the forwards and backwards fields in each layer. To find the forwards and backwards fields F_i and B_i in layer i, we once again consider the N layer system as a 3 layer system, as depicted in figure 3.9b. By replacing layers 2 to $i-1$ with an effective interface between layers 1 and i, and replacing layers $i+1$ to $N-1$ with an effective interface between layers i and N, we can see that we must have that

$$F_N = t_{iN} F_i \tag{3.5.2.5}$$

and

$$B_i = r_{iN} F_i. \tag{3.5.2.6}$$

On the other hand, we also know that

$$F_N = t_{1N} F_1, \tag{3.5.2.7}$$

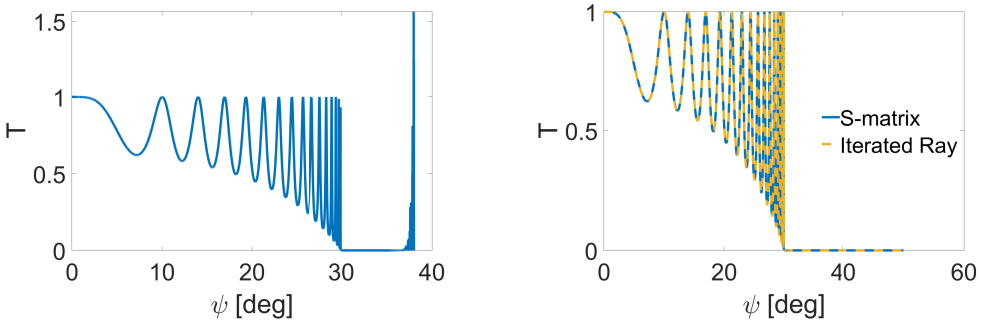
from which it follows that

$$F_i = \frac{t_{1N}}{t_{iN}} F_1 \tag{3.5.2.8a}$$

and

$$B_i = \frac{t_{1N} r_{iN}}{t_{iN}} F_1. \tag{3.5.2.8b}$$

An example of a three layer system with evanescent waves in the centre layer is shown in figures 3.10a and 3.10b. The system refractive indices are 2, 1, and 2, with a



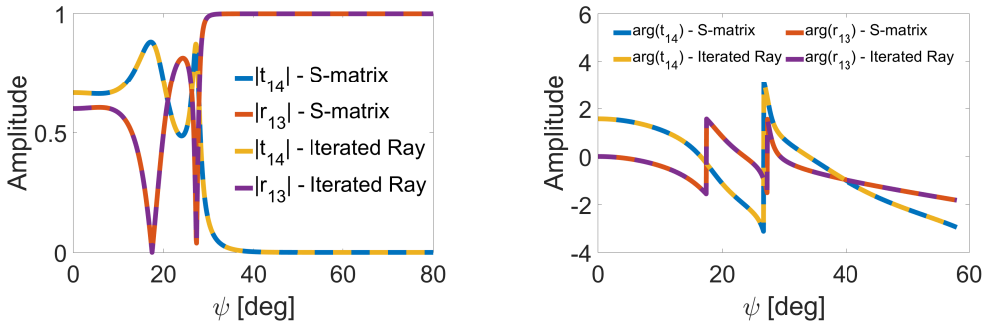
(A) (B)

FIGURE 3.10: A comparison of the transmission across a three layer system as a function of angle of incidence, ψ, for (a) the T-matrix method and (b) the S-matrix and Iterated Ray method. The material refractive indices are 2, 1, and 2 (in that order) and the layer thickness is eight wavelengths. The instability of the T-matrix method can be observed in (a), where the transmission can be observed rising sharply around 38°, despite the critical angle for total reflection being around 30°.

layer thickness of eight wavelengths. We see that all three outlined methods, the T- and S-matrices and the Iterated Ray method, give the same solution up until the critical angle that the middle field becomes evanescent, around 30°. We can see the instability of the T-matrix leading to a sudden surge in the transmission coefficient. By 70° the T-matrix solution has grown to $O\left(10^{38}\right)$, even with a layer thickness of only eight wavelengths. The S-matrix and ray method both remain at zero past the critical angle; as they should. The stability of the ray method is evident from equations (3.5.2.3a) to (3.5.2.4b). As there are only forward propagating exponential terms, there is only ever exponential decay, never growth as in the T-matrix, in the presence of evanescent fields.

An example four layer system is shown in figure (3.11). The system refractive indices are 2, 1, 2, and 2.5, and the thickesses of the two layers are 1.25 and 0.5 wavelengths, respectively. We once again observe the total internal reflection across the evanescent layer (layer 2 with refractive index 1) at around 30°, shown by r_{13} going to one. Correspondingly we see the transmission across the entire system, t_{14}, go to zero. The agreement between the Iterated Ray method and the S-matrix is exact.

Finally, we consider the computation speed of the Iterated Ray and S-matrix methods. Coding both methods in Matlab and optimising to the best of our abilities, the Matlab function `timeit` can be used to compare the execution speed of the methods. Comparing a sample of 2,750 random structures, with a number of layers between 3 and 20, thicknesses between 1 and 20 wavelengths, and refractive indices between 1 and 5, the speed ratio between the two algorithms is calculated as a function of the number of angles of incidence analysed, ranging from 100 to 1000, in the range 0° to 89°.



(A) (B)

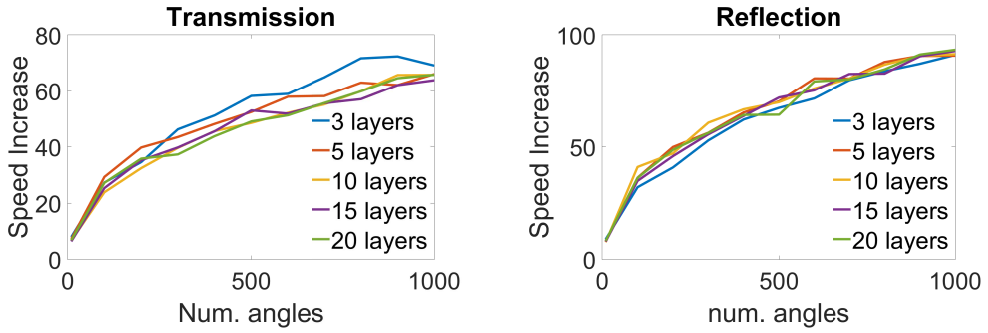
FIGURE 3.11: (a) A comparison of the amplitudes of the transmission coefficients across a four layer system and the reflection coefficients from the initial three layer subsystem, as a function of the angle of incidence, ψ, as given by the Iterated Ray and S-Matrix formulisms. (b) The angles of the coefficients. The system refractive indices are 2, 1, 2, and 2.5, and the layer thicknesses are 1.25 and 0.5 wavelengths.

The speed ratio between the two methods increases from approximately 25 to approximately 65 as the number of angles of incidence computed for each structure increases from 100 to 1,000, shown in figure 3.12a. This increase in speed is in part due to the fact that it is easy to vectorise the scalar equations (3.5.2.1) and (3.5.2.2) for an arbitrary number of incident angles. The S-matrix code, on the other hand, requires a matrix to be associated with each angle of incidence, with the steps performed at each iteration being more involved than the simple multiplication of vectors in the ray method.

Another reason for the increased efficiency is that, with the Iterated Ray Method, we need only calculate two effective coefficients at each step, rather than four as in the S-matrix formulation. Furthermore, in the case that that it is only the reflection from isotropic layers that is desired, it is clear from equation (3.5.2.2) that we only need to calculate the effective reflection at each iteration, since it does not depend on the effective transmission. This further increases the speed of the code, and the maximum speed ratio increase we have measured in this case is approximately 90, as shown in figure 3.12b.

3.5.3 Generalised Multi-Mode Ray Method

So far we have only considered systems of isotropic media. There is no polarisation rotation upon transmission and reflection and all rays have the same tangential wavevector component. As a result, all of these rays can be thought of as a single mode of the system, and the ray method is a scalar calculation. However, in the case that we have anisotropic media or diffraction, we have different polarisation states and fields with different tangential wavevector components to consider. These are all separate modes of the system that mix with one another each time a mode is incident



(A) (B)

FIGURE 3.12: Graphs showing the speed increase of the Iterated Ray method over the S-matrix approach when calculating (a) the transmission and (b) reflection. The values are calculated from the average run times provided by the Matlab function `timeit` for 2,750 samples of random structures of varying layer thicknesses and refractive indices for varying numbers of angles of incidence.

upon a boundary or grating layer. We therefore generalise the scalar/isotropic method of the previous sections to account for mode mixing.

We first define the mode vectors

$$\boldsymbol{F}^{(n)} = F_i^{(n)} \hat{\boldsymbol{e}}_i, \tag{3.5.3.1}$$

$$\boldsymbol{B}^{(n)} = B_i^{(n)} \hat{\boldsymbol{e}}_i, \tag{3.5.3.2}$$

where $F_i^{(n)}$ and $B_i^{(n)}$ are the amplitudes of the forwards and backwards i'th modes in the n'th layer. The transmission and reflection from layer j to k of the various modes are then derived in the same manner as the isotropic case, and are described by the matrices

$$\begin{aligned} \mathrm{T}_{jk} &= \mathrm{T}_{j+1,k} \left[\mathcal{I} - \Phi_{j+1}^{+} \mathrm{R}_{j+1,j} \Phi_{j+1}^{-} \mathrm{R}_{j+1,k} \right]^{-1} \Phi_{j+1}^{+} \mathrm{T}_{j,j+1} \\ &= \mathrm{T}_{j+1,k} \left[\mathcal{I} - \mathrm{R}_{j+1}^{tf} \right]^{-1} \Phi_{j+1}^{+} \mathrm{T}_{j,j+1}, \end{aligned} \tag{3.5.3.3a}$$

and

$$\begin{aligned} \mathrm{R}_{jk} &= \mathrm{R}_{j,j+1} + \mathrm{T}_{j+1,j} \left[\mathcal{I} - \Phi_{j+1}^{-} \mathrm{R}_{j+1,k} \Phi_{j+1}^{+} \mathrm{R}_{j+1,j} \right]^{-1} \Phi_{j+1}^{-} \mathrm{R}_{j+1,k} \Phi_{j+1}^{+} \mathrm{T}_{j,j+1} \\ &= \mathrm{R}_{j,j+1} + \mathrm{T}_{j+1,j} \left[\mathcal{I} - \mathrm{R}_{j+1}^{rf} \right]^{-1} \mathrm{R}_{j+1}^{r} \mathrm{R}_{j+1,j}^{-1} \mathrm{T}_{j,j+1} \end{aligned} \tag{3.5.3.3b}$$

for forward travel and

$$\begin{aligned} \mathrm{T}_{jk} &= \mathrm{T}_{j-1,k} \left[\mathcal{I} - \Phi_{j-1}^{-} \mathrm{R}_{j-1,j} \Phi_{j-1}^{+} \mathrm{R}_{j-1,k} \right]^{-1} \Phi_{j-1}^{+} \mathrm{T}_{j,j-1} \\ &= \mathrm{T}_{j-1,k} \left[\mathcal{I} - \mathrm{R}_{j+1}^{tb} \right]^{-1} \Phi_{j-1}^{+} \mathrm{T}_{j,j-1} \end{aligned} \tag{3.5.3.3c}$$

$$\begin{aligned}\mathrm{R}_{jk} &= \mathrm{R}_{j,j-1} + \mathrm{T}_{j-1,j}\left[\mathcal{I} - \Phi^{+}_{j-1}\mathrm{R}_{j-1,k}\Phi^{-}_{j-1}\mathrm{R}_{j-1,j}\right]^{-1}\Phi^{+}_{j-1}\mathrm{R}_{j-1,k}\Phi^{-}_{j-1}\mathrm{T}_{j,j-1} \\ &= \mathrm{R}_{j,j-1} + \mathrm{T}_{j-1,j}\left[\mathcal{I} - \mathrm{R}^{rb}_{j+1}\right]^{-1}\mathrm{R}^{rb}_{j+1}\mathrm{R}^{-1}_{j-1,j}\mathrm{T}_{j,j-1} \qquad (3.5.3.3d)\end{aligned}$$

for backward travel. The matrices T_{ij} and R_{ij} contain the transmission and reflection coefficients of the modes going from layer i to layer j. The kl element $\left(\mathrm{T}_{ij}\right)_{kl}$ is the transmission of the k'th mode to the l'th mode going across boundary/layer i to j (and similarly for the R matrices). The matrices Φ^{+}_{i} and Φ^{-}_{i} contain the phase changes of the modes going across layer i in the forwards and backwards direction, respectively. These phase changes are not, in general, the same, since the forwards and backwards e-waves in an anistropic layer have different refractive indices and normal wavevector components. While the Φ matrices are always diagonal the T and R matrices are only diagonal when they correspond to an isotropic/isotropic interface. Using this formulation we have, in analogy with equations 3.5.2.8a and 3.5.2.8b,

$$\boldsymbol{F}^{(i)} = \mathrm{T}^{-1}_{iN}\mathrm{T}_{0N}\boldsymbol{F}^{(0)} \qquad (3.5.3.4)$$

and

$$\boldsymbol{B}^{(i)} = \mathrm{R}_{iN}\mathrm{T}^{-1}_{iN}\mathrm{T}_{0N}\boldsymbol{F}^{(0)}. \qquad (3.5.3.5)$$

3.5.4 Transmission Across an Anisotropic Layer

We now consider the case of an anisotropic material bounded by isotropic media and employ the multi-mode ray method. We measure the input and output fields in terms of their s- and p-polarised components, and so we first look to write

$$\begin{pmatrix} F^{(3)}_{p} \\ F^{(3)}_{s} \end{pmatrix} = \mathrm{T}_{13}\begin{pmatrix} F^{(1)}_{p} \\ F^{(1)}_{s} \end{pmatrix}. \qquad (3.5.4.1)$$

The transmission is found to be

$$\mathrm{T}_{13} = \begin{pmatrix} t^{23}_{op} & t^{23}_{ep} \\ t^{23}_{os} & t^{23}_{es} \end{pmatrix}\left[\mathcal{I} - \mathrm{R}_{t}\right]^{-1}\begin{pmatrix} t^{12}_{po}e^{iq^{+}_{o}L} & t^{23}_{so}e^{iq^{+}_{o}L} \\ t^{12}_{pe}e^{iq^{+}_{e}L} & t^{12}_{se}e^{iq^{+}_{e}L} \end{pmatrix}, \qquad (3.5.4.2a)$$

where

$$\mathrm{R}_{t} = \begin{pmatrix} r^{23}_{oo}r^{21}_{oo}e^{i(q^{+}_{o}-q^{-}_{o})L} + r^{23}_{oe}r^{21}_{eo}e^{i(q^{+}_{o}-q^{-}_{e})L} & r^{23}_{eo}r^{21}_{oo}e^{i(q^{+}_{o}-q^{-}_{o})L} + r^{23}_{ee}r^{21}_{eo}e^{i(q^{+}_{o}-q^{-}_{e})L} \\ r^{23}_{oe}r^{21}_{ee}e^{i(q^{+}_{e}-q^{-}_{e})L} + r^{23}_{oo}r^{21}_{oe}e^{i(q^{+}_{e}+q^{+}_{o})L} & r^{23}_{ee}r^{21}_{ee}e^{i(q^{+}_{e}-q^{-}_{e})L} + r^{23}_{eo}r^{21}_{oe}e^{i(q^{+}_{e}+q^{+}_{o})L} \end{pmatrix} \qquad (3.5.4.2b)$$

and the coefficients t^{23}_{op} are the Fresnel coefficients for the transmission of the o- to p-wave across the layer 2 to 3 boundary, etc. The phase changes in R_t are found by considering figure 3.13, where the forward and backwards rays are described by their

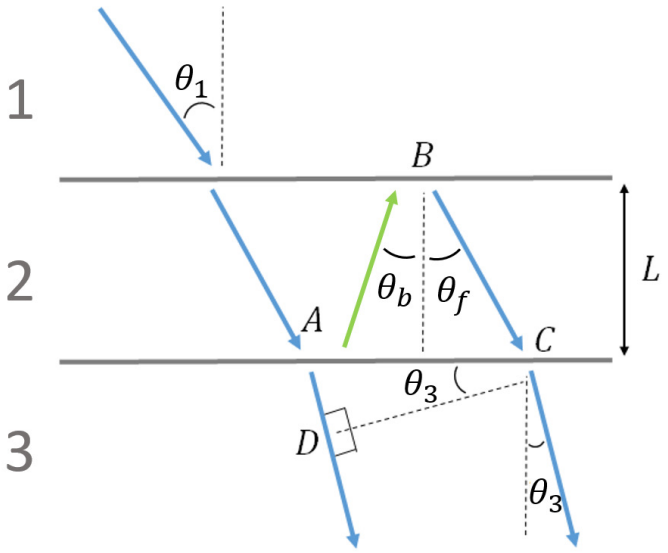


FIGURE 3.13: Diagram used to find the optical path difference between two emerging rays when there is mode mixing within a layer. The forwards and backwards angles of propagation, θ_f and θ_b, need not be the same, since the angle of the o-rays and forwards and backwards e-rays are all different.

angles θ_f and θ_b and their refractive indices n_f and n_b. The path difference is given by

$$\begin{aligned}\delta &= k_0 \left[\left|\overrightarrow{AB}\right| n_b + \left|\overrightarrow{BC}\right| n_f - \left|\overrightarrow{AD}\right| n_3 \right] \\ &= k_0 L \left[n_b \sec\theta_b + n_f \sec\theta_f - n_3 \left|\sin\theta_3\right| \left(\left|\tan\theta_b\right| + \left|\tan\theta_f\right| \right) \right], \end{aligned} \tag{3.5.4.3}$$

which, after applying Snell's law, becomes

$$\begin{aligned}\delta &= k_0 L \left[n_b \cos\theta_b + n_f \cos\theta_f \right] \\ &= \left(q_f - q_b \right) L, \end{aligned} \tag{3.5.4.4}$$

where the q's are the normal components of the wavevectors, as in section 2.5.7. The reflection from the layer, defined by

$$\begin{pmatrix} B_p^{(1)} \\ B_s^{(1)} \end{pmatrix} = \mathrm{R}_{13} \begin{pmatrix} F_p^{(1)} \\ F_s^{(1)} \end{pmatrix}, \tag{3.5.4.5}$$

is given by

$$\mathrm{R}_{13} = \begin{pmatrix} r_{pp}^{12} & r_{sp}^{12} \\ r_{ps}^{12} & r_{ss}^{12} \end{pmatrix} + \begin{pmatrix} t_{op}^{21} & t_{ep}^{21} \\ t_{os}^{21} & t_{es}^{21} \end{pmatrix} \left[\mathcal{I} - \mathrm{R}_r\right]^{-1} \mathrm{R}_r \begin{pmatrix} r_{oo}^{21} & r_{eo}^{21} \\ r_{oe}^{21} & r_{ee}^{21} \end{pmatrix}^{-1} \begin{pmatrix} t_{po}^{12} & t_{so}^{23} \\ t_{pe}^{12} & t_{se}^{12} \end{pmatrix}, \tag{3.5.4.6a}$$

where

$$\mathrm{R}_r = \begin{pmatrix} r_{oo}^{21} r_{oo}^{23} e^{i(q_o^+ - q_o^-)L} + r_{oe}^{21} r_{eo}^{23} e^{i(q_e^+ + q_o^+)L} & r_{eo}^{21} r_{oo}^{23} e^{i(q_o^+ - q_o^-)L} + r_{ee}^{21} r_{eo}^{23} e^{i(q_e^+ + q_o^+)L} \\ r_{oe}^{21} r_{ee}^{23} e^{i(q_e^+ - q_e^-)L} + r_{oo}^{21} r_{oe}^{23} e^{i(q_o^+ - q_e^-)L} & r_{ee}^{21} r_{ee}^{23} e^{i(q_e^+ - q_e^-)L} + r_{eo}^{21} r_{oe}^{23} e^{i(q_o^+ - q_e^-)L} \end{pmatrix}. \tag{3.5.4.6b}$$

The transmission and reflection for backward travel are given by interchanging all superscript 1's and 3's and all changing all $q_\alpha^\pm$ to $-q_\alpha^\mp$. An example comparison of this solution to the finite elements solution provided by Comsol is shown in figures 3.14 and (3.15). The layer is highly anisotropic and asymmetric (with respect to the angle of incidence) so as to induce very large, asymmetric polarisation rotation effects for the purposes of comparison; the agreement between the two is virtually exact.

The Comsol model requires the use of perfectly matched layers (PMLs) to simulate propagation into free space (the semi-infinite sandwiching layers). A PML is designed so that it is as perfectly absorbing as possible and reflects no waves, thus replicating propagation to infinity. However, for very large angles of incidence, the PMLs will begin reflecting some radiation back towards the centre layer, leading to non-physical resonance effects in the numerical simulation. It is for this reason that the comparison in figure 3.14 does not extend beyond 78°, which is where the finite elements solution starts to see such resonance effects. The Ray method, on the other hand, remains accurate at all angles of incidence, offering a distinct advantage over full numerical simulations, of which the accuracy of the results is often taken for granted.

3.6 Modelling Isotropic PAAD Layers

As discussed at the beginning of this section, studying the full diffraction problem and determining the alignment of the thin PAAD layers is made far more challenging by the amount of other unknowns in the system. Ideally, the fewer free parameters in the minimisation routine, the better. We begin, then, by considering simplified subsystems of the final diffractive problem, allowing the determination of a number of the parameters in a manner that isolates them from as many of the others as possible. The transmission through the substrate only is considered first, thus allowing the glass and ITO parameters to be found. This is followed by the substrate coated with isotropic PAAD, allowing the thickness and refractive index of the unaligned PAAD to be determined. Once these parameters are known, we can look for the optical axis and birefringence of the photoaligned PAAD, reducing the free parameters in the diffraction problem from eight to three.

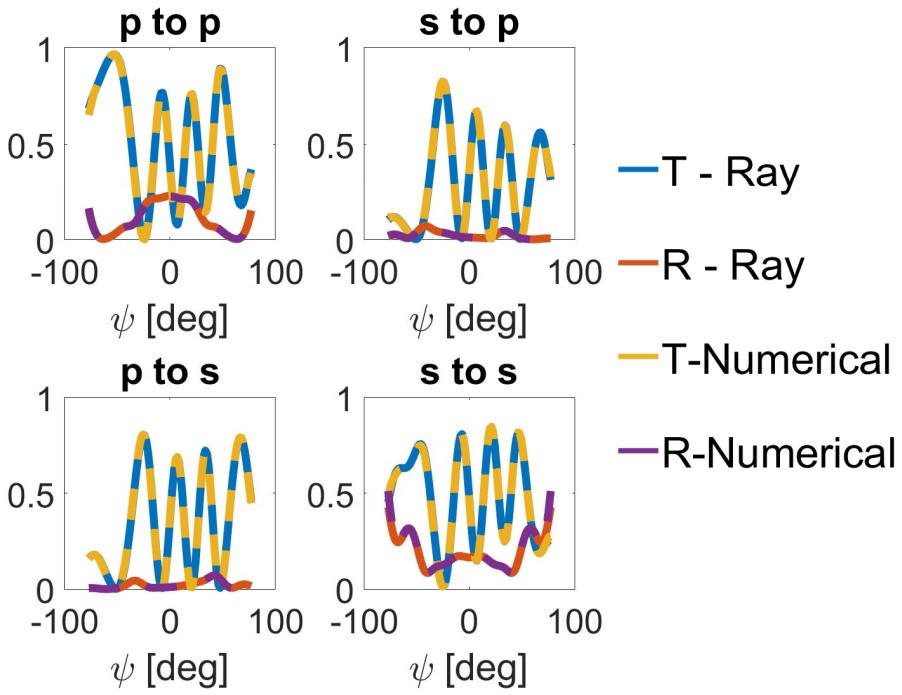


FIGURE 3.14: A comparison of the various transmittances, T, and reflectances, R, for s- and p-polarised incident fields to s- and p-polarised output fields, as calculated using the Iterated Ray Method and numerically with Comsol. The system refractive indices are $n_{in} = 1$, $n_e = 5$, $n_o = 1.5$, and $n_{out} = 2$. The layer thickness is two wavelengths and the optical axis is at an angle of $\pi/3$ with the y axis, with its projection in the xz plane (plane of incidence) equal to $\pi/5$.

3.6.1 Transmission Through a Glass Slide

As a test for the implementation of the fitting method and algorithmns, just a single glass slide was considered to begin with. Since all materials involved are isotropic (air and glass), we look to apply the multi-layer transmission formula 3.5.2.3a. However, this formula is derived for plane waves. When it is applied to such thick layers as that of the glass substrate (measured experimentally as $L_g = 1.06 \times 10^{-3}$), the phase change

$$\delta = 2k_0 n_g L_g \sqrt{1 - \frac{\sin^2 \psi}{n_g^2}} \tag{3.6.1.1}$$

varies rapidly with the angle of incidence ψ, owing to the factor of $k_0 L_g = O\left(10^4\right)$. Accordingly, the transmission through the glass layer also oscillates very rapidly with ψ. This does not, however, match the experimental results, shown in figure 3.16, which are mostly very smooth.

We see from the spikes at normal incidence that the interference effects cannot be ignored inside the glass, and so we must consider the Gaussian nature of the beam

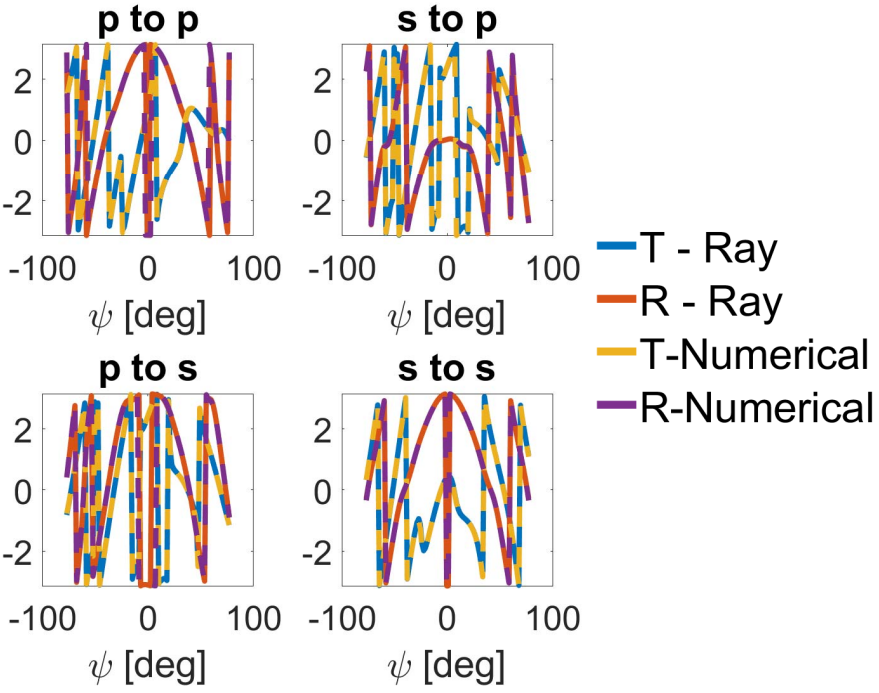


FIGURE 3.15: A comparison of the phase, in radians, of the Fresnel coefficients for the transmission and reflection of s- and p-polarised incident fields to s- and p-polarised output fields. The system refractive indices are $n_{in} = 1$, $n_e = 5$, $n_o = 1.5$, and $n_{out} = 2$. The layer thickness is two wavelengths and the optical axis is at an angle of $\pi/3$ with the y axis, with its projection in the xz plane (plane of incidence) equal to $\pi/5$.

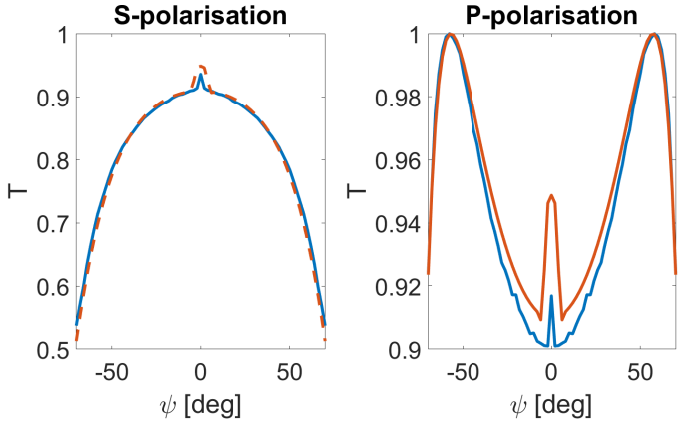


FIGURE 3.16: Graphs showing the experimental (blue) and fit (orange) transmission data for a HeNe 632.8nm laser through a glass slide.

used. Unlike the plane waves used to derive the transmission formula 3.5.2.3a, the Gaussian beams used experimentally do not extend infinitely and do not have parallel phase fronts. In fact, they have a decaying amplitude with transverse distance from their central axis, and also possess spherical wavefronts that change in radius and decay in amplitude with propagation distance. Their (normalised) electric field can be succinctly described by [51]

$$E = \sqrt{\frac{2}{\pi}} \frac{iz_R}{w_0} \frac{1}{q(z)} e^{-ik\left(\frac{x^2}{2q(z)} - z\right)}, \tag{3.6.1.2}$$

where the beam parameter

$$q(z) = z + iz_R \tag{3.6.1.3}$$

(capturing the change in amplitude and curvature) is given in terms of the Rayleigh range,

$$z_R = \frac{\pi W_0^2}{\lambda}, \tag{3.6.1.4}$$

and the beam waist W_0, which is a measure of the beam width at its narrowest point (and found from the laser specifications as 2×10^{-3}m). The Gaussian field can be written as a sum of plane waves with direction cosines α by taking the Fourier transform to find

$$\mathcal{F}\left\{\tilde{E}\right\} = \sqrt{\frac{W_0}{\sqrt{2\pi}}} e^{-\frac{\tilde{q}\alpha^2}{4\pi i}}. \tag{3.6.1.5}$$

In this way, the total transmitted field is a weighted average given by

$$E_T(\alpha_0) = \mathcal{F}^{-1}\left\{\mathcal{F}\left\{\tilde{E}\right\} * \delta(\alpha - \alpha_0)\, t_{13}(\alpha)\right\}, \tag{3.6.1.6}$$

where $*$ denotes convolution, $t_{13}(\alpha)$ is the effective transmission coefficient across the three layer system for a plane wave incident with cosine α, and α_0 is cosine of the beams propagation direction. However, since the experimental measurement is actually the intensity, we use Parseval's theorem to calculate

$$\begin{aligned} T(\alpha_0) &= \sqrt{\frac{W_0}{\sqrt{2\pi}}} \int_{-1}^{1} \left|\mathcal{F}\left\{\tilde{E}\right\} * \delta(\alpha - \alpha_0)\, t_{13}(\alpha)\right|^2 d\alpha \\ &= \sqrt{\frac{W_0}{\sqrt{2\pi}}} \int_{-1}^{1} \left|t_{13}(\alpha)\right|^2 e^{-\left(\frac{z_R}{4\pi}\right)(\alpha-\alpha_0)^4} d\alpha. \end{aligned} \tag{3.6.1.7}$$

It is found, however, that this averaging is not enough to obtain a smooth solution. This is likely due to factors such as small fluctuations in thickness (which could easily be of the order of wavelengths), the presence of other wavelengths in the beam, and any averaging effects that take place as the detector measures the field. As a result, a second Gaussian smoothing is applied to the calculated transmittance, T, with the width of the kernel left as fitting parameter.

These integrals are performed numerically in Matlab, and passed into the inbuilt lsqnonlin minimisation routine to fit the averaged model to the experimental data. The data for the s- and p-polarisations is fit simultaneously. A sampling of 15 of the 70 experimental data points are used, with the theoretical solution calculated and averaged in an interval around each of these points, with an angular resolution of $\pi/5000$ to capture the oscillations in enough detail for a convergent solution. It is found that the dependence on thickness is too oscillatory near normal incidence for a good solution to be found. As a result, the thickness is fixed at the experimental measurement, and the interval $[-8, 8]$ is excluded for an initial fit. Thus allowing the refractive index of the glass to be fit first. Afterwards, the thickness can be then be adjusted by fitting the centre interval. The refractive index of the glass is found to be 1.5549 and the second averaging window is found to have a total width of 0.0096°, corresponding to a Gaussian with $\sigma = 1.9 \times 10^{-3} \approx W_0$. The results of the fitting are shown in figure 3.16.

The next intended step was to proceed to consider the ITO coated glass substrate. However, the experimental data proved to be very noisy and it was not possible to obtain any good fitting results from the data. It is unknown if the noise was due to some experimental error, environmental factors, or even if we were seeing a sampling of the rapid oscillations due to interference in the ITO layer. Given that the refractive index of the ITO is expected to be higher than that of the glass, the angles of the rays inside the ITO layer will be smaller. This makes for slower oscillations as the angle of incidence is varied, and so we would expect the averaging process to be less effective. However, given that the data for the PAAD coated substrate is very smooth, it seems most likely that the noise is due to surface roughness of the ITO.

3.6.2 Transmission of Isotropic PAAD and Substrate

We proceed, then, to the next step; fitting the transmission data for isotropic PAAD deposited onto the substrate. While it would have been preferable to have the ITO parameters already, it is still possible to fit for the ITO and PAAD parameters simultaneously; albeit with far longer running times on the minimisation routine. The glass thickness and refractive index, however, are fixed at the results of the first fitting.

First, we note the slight asymmetry in the experimental data, which is shown in figure 3.17. Since the theoretical function is symmetric (with respect to angle of incidence), we know the least squares minimisation will end up fitting to the average of the data. We take advantage of this to speed up run times by feeding the routine the average of the experimental data beforehand, and only fitting to one side of normal incidence. Again, the centre region is excluded for an initial fit (width 11°), and a sampling of the outer regions are used to fit the thicknesses and refractive indices of the PAAD and ITO. The centre region is then used to fine tune the glass thickness.

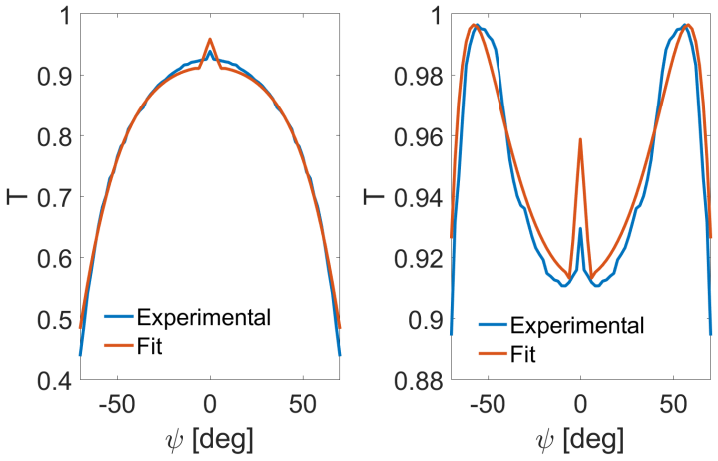


FIGURE 3.17: Graphs showing the experimental and fitted transmission curves for a (left) s-polarised and (right) p-polarised 632.8nm beam incident onto the PAAD/ITO/glass system.

It proves incredibly difficult (perhaps impossible) to obtain a good fitting at both high and low angles of incidence. In particular, obtaining a good fit for the p-polarised data above $\pm 56^\circ$ (the two peaks) cannot be achieved without incurring very large errors inside of the peaks, and vice-versa. However, since the p-polarised data above is 56° is practically vertical, the experimental and theoretical data can be very similar to one another but still have very large errors in the least squares sense. We therefore also exclude the data points outside of $\pm 56^\circ$ to try and obtain a better fitting.

This proves much more successful, with the results shown in figure 3.17. The ITO refractive index and thickness are found to be 1.85 and 142 nm. Unfortunately, the specifications of the substrates used are unknown, but the thickness is entirely consistent with the specifications of ITO/glass substrates offered by many manufacturers, usually in the 100-200nm range. The refractive index is also in good agreement with other experimental work, [37] and [29], for example, giving 1.87 and 1.78, respectively. The isotropic refractive index and thickness of the PAAD is found to be 1.73 and 27 nm. While the refractive index of the PAAD is unknown, the thickness is also consistent with the measured thicknesses in the original work from our group [35], ranging between 15 nm and 65 nm, depending on the speed at which the spin-coater was run.

3.7 Modelling PAAD diffraction

3.7.1 Surface Relief Gratings in Thick Layers

Before continuing with the modelling of the thin PAAD birefringence gratings, we first consider other experimental work from one of our research group's collaborators [24]. The work provides some nice results against which to compare our results and give some validation to our theoretical modelling techniques. Firstly, they measured the birefringence induced in the PAAD samples under illumination with polarised light. This was done using the same experimental setup as used in the twist-untwist experiments, shown in figure (3.2). No boundary alignments are fixed and so the entire PAAD layer is essentially one uniform slice. Applying the twist/untwist theory, the formula describing the polarisation rotation reduces to

$$\boldsymbol{E}_{out} = \begin{pmatrix} e^{-i\frac{\Gamma}{2}} & 0 \\ 0 & e^{i\frac{\Gamma}{2}} \end{pmatrix} \boldsymbol{E}_{in}. \tag{3.7.1.1}$$

The degree of rotation is given by $\frac{\Gamma}{2} = \frac{\pi \Delta n L}{\lambda}$, where L is the thickness of the layer and $\Delta n = n_e - n_o$ is the birefringence of the PAAD. The intensity of the rotated component is found from

$$I = I_0 \sin^2\left(\frac{\Gamma}{2}\right), \tag{3.7.1.2}$$

where I_0 is the baseline transmission intensity measurement when the PAAD is not present. The value of the birefringence that was found for PAAD-22 was 0.185 for a 75 nm thick layer, ranging to 0.168 for a 270 nm thick layer. As we will see in the next section, this is in good agreement with the magnitude of birefringence required to produce diffraction effects of the same magnitude found in our experimental work on thin PAAD birefringence gratings.

The second experiment considered in [24] was studying surface relief gratings in thick PAAD layers. These are systems where the realignment of the PAAD causes a periodic modulation of the layer thickness, leading to diffraction. The PAAD layers are realigned with the interference pattern in the same way as the diffraction efficiency experiment depicted in (3.6). This was carried out for PAAD-22 and PAAD-72, another photoaligning azobenzene polymer. As in the original diffraction efficiency experiments in our group [35], the PAAD-22 layers exhibited no surface relief. The PAAD-72, however, was able to form very high quality gratings with strong diffraction efficiency. In particular, a 270 nm layer with a 70 nm sinusoidal modulation was able to produce a diffraction efficiency of 5.7%.

The grating was recorded with circularly polarised beams. This is generally understood in the literature to give an isotropic alignment of the PAAD, e.g. [56] [62], however, during the reorientation process there is a net movement of the molecules

that leads to the formation of the grating [42]. It is relatively simple, then, to compare this result to the theoretical value, as calculated using the non-paraxial diffraction techniques outlined in section 2.7.4.

We take the peaks of the grating to lie at $z = 0$ and let the grating vary in x with period Λ. The thickness of the grating at any point can then be written as

$$\tilde{L} = \tilde{L}_0 + \Delta\tilde{L}\cos\left(\frac{2\pi\tilde{x}}{\tilde{\Lambda}}\right), \tag{3.7.1.3}$$

where $\tilde{L}_0$ is the average thickness, $\Delta\tilde{L}$ is the modulation, $\Lambda = 1\mu$m is the grating period, and we have renormalised with respect to λ. The optical path length of the field as it passes across the grating at normal incidence can be approximated by

$$\begin{aligned}\delta &= n_{air}\left(2\tilde{L_0} - \tilde{L}\right) + n_p\tilde{L} \\ &= \left(1 + n_p\right)\tilde{L}_0 + \left(n_p - 1\right)\Delta\tilde{L}\cos\left(\frac{2\pi\tilde{x}}{\tilde{\Lambda}}\right),\end{aligned} \tag{3.7.1.4}$$

where n_p is the isotropic PAAD refractive index and $n_{air} = 1$. This ignores any reflection and refraction effects at the surface of the grating. We therefore approximate the phase of the wave emerging from the grating as

$$E_{out} \propto e^{2\pi i\delta} \approx e^{2\pi i\left(1+n_p\right)\tilde{L}_0}\left[1 + 2\pi i\left(n_p - 1\right)\Delta\tilde{L}\cos\left(\frac{2\pi\tilde{x}}{\tilde{\Lambda}}\right) + O\left(\Delta\tilde{L}^2\right)\right] \tag{3.7.1.5}$$

where $\Delta\tilde{L} \approx 1/9$ for the probe beam with $\lambda = 632$ nm. This has Fourier transform

$$\mathcal{F}\left\{E_{out}\right\} = e^{2\pi i\left(1+n_p\right)\tilde{L}_0}\left\{1 + \pi i\left(n_p - 1\right)\Delta\tilde{L}\left[\delta\left(\alpha - \frac{1}{\tilde{\Lambda}}\right) + \delta\left(\alpha + \frac{1}{\tilde{\Lambda}}\right)\right] + O\left(\Delta\tilde{L}^2\right)\right\}, \tag{3.7.1.6}$$

and the diffraction efficiencies of the first diffracted orders are then given by their radiant intensities (as outlined in section 2.7.4) as

$$\begin{aligned}\eta_{\pm1} &\approx \frac{\pi^2\left(n_p - 1\right)^2}{81}\sqrt{1 - \tilde{\Lambda}^{-2}} \\ &\approx 0.1\left(n_p - 1\right)^2.\end{aligned} \tag{3.7.1.7}$$

Assuming the refractive index of the different PAAD variants are similar, a value of $n_p = 1.73$ (as found in the fitting in the previous section) gives a diffraction efficiency of 5.3%, and we we have very good agreement with the experimental results.

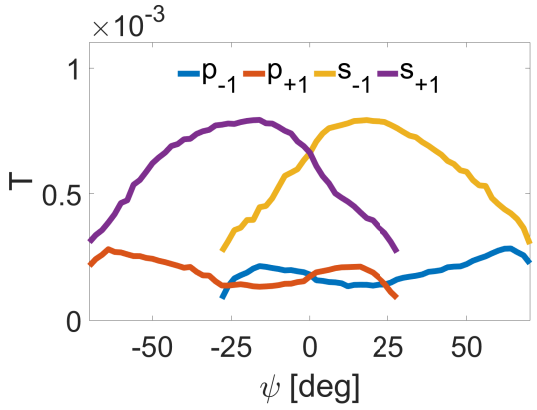


FIGURE 3.18: Normalised transmissions as a function of the angle of incidence, ψ, for the s- and p-polarisations of the first diffracted orders for an aligned PAAD grating.

3.7.2 Birefringence Gratings in Thin Layers

3.7.2.1 The Modelling Approximations

We now consider the thin, birefringence gratings studied in the experiments, the results of which are shown in figure 3.18. The transmission will be modelled using the the generalised, multi-mode ray method outlined earlier. In total we will model six modes, an s- and p-polarisation for the fundamental and the ± 1 diffracted orders. The various modes will be created in the grating layer, which will be modelled as an effective layer between the incident air layer and the ITO. We start now by deriving the effective coefficients of the grating layer.

To begin, we note that the grating layer is in fact very thin and wide. The thickness from the isotropic fits of 27 nm is around 0.04 wavelengths. From figure (3.18), we see that the orders look to be going to zero in the region of $30^\circ - 40^\circ$. This corresponds to a grating spacing of around 2 to 3 wavelengths, and each segment of the grating is around 50-80 times wider than it is thick. This leads us to our key modelling assumption. We assume that we have a square-wave grating, and that the field emerging from each isotropic/anisotropic section is simply given by the effective transmission coefficient t_{iso} or t_{ani} for an infinitely extending isotropic or anistropic layer, respectively. This is depicted in figure 3.19. While the field may not be approximated well near the boundaries of the isotropic and anistropic sections, the fact that each section is so wide and thin means that modelling them as infinite layers should be reasonable everywhere else.

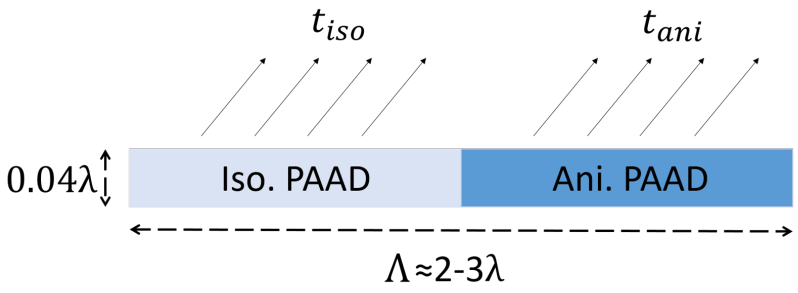


FIGURE 3.19: A schematic of the approximate grating geometry and the key modelling assumption, that the transmission through each thin section of isotropic or anisotropic PAAD can be approximated by the effective transmission coefficients t_{iso} or t_{ani} for an infinitely extending layer.

3.7.2.2 Transmission and Reflection of Anisotropic PAAD

We know how an incident field will transmit across and reflect from the isotropic sections of the PAAD. We directly apply the results for the effective coefficients of a three layer system using the value for the refractive index obtained from the fitting of the previous section. We also need to describe the transmission and reflection of the anistropic PAAD. This will be dependent on three unknown variables, the two angles that parameterise the optical axis, and the birefringence. The formulae describing the transmission and reflection of an anistropic layer, (3.5.4.2a)-(3.5.4.2b) and (3.5.4.6a)-(3.5.4.6b), depend on the Fresnel coefficients at each of the isotropic/anistropic interfaces, and also on the phase changes of the o- and e-waves inside the layer. This, in turn, requires knowledge of the wavevectors and fields of the o- and e-waves inside the layer. Our model, then, must first calculate these parameters as a function of the angle of incidence, the ordinary and extraordinary refractive indices, and the alignment of the optical axis.

We fix the boundary of the anistropic layer as the $z = 0$ plane and take the plane of incidence to be the xz plane. We then consider a field that is incident upon the layer with tangential wavevector component equal to K. Phase matching at the boundary means that the waves inside the anisotropic layer must have the same tangential components ($k_x = K$ and $k_y = 0$). Thus, we need to find the forward and backward o- and e-waves associated with K for a given n_e, n_o, and alignment of the optical axis. The optical/crystal axis of PAAD will be described by

$$\boldsymbol{c} = \begin{pmatrix} \cos\theta \sin\phi \\ \cos\phi \\ \sin\theta \sin\phi \end{pmatrix}, \tag{3.7.2.2.1}$$

where ϕ is the azimuthal angle made with the y axis and θ is the polar angle made with the positive x axis. This is shown in figure 3.20. Since we know the isotropic

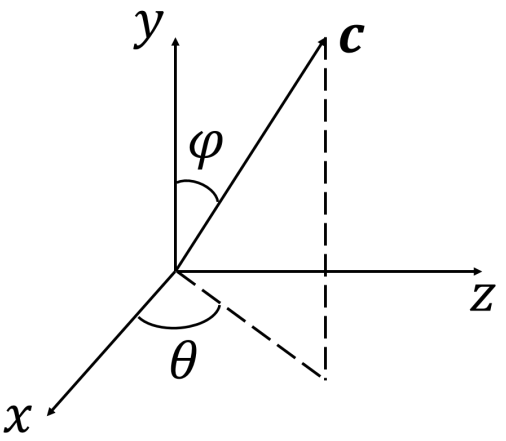


FIGURE 3.20: Diagram depicting the angles used to describe the optical axis of the PAAD.

refractive index from the fitting in the previous section, we determine n_e and n_o from our last unknown parameter, the birefringence, Δn, according to

$$n_o = n_{iso} - \Delta n \tag{3.7.2.2.2}$$

$$n_e = n_{iso} + 2\Delta n. \tag{3.7.2.2.3}$$

This follows from the fact that, for randomly oriented molecules, the resulting isotropic refractive index is given by $n_{iso} = \frac{2n_o+n_e}{3}$.

First we consider the o-waves. From the discussion in section 2.4.2, we know that the o-waves always see the refractive index n_o, regardless of their direction of propagation. It follows that the wavevectors of the forwards and backwards o-waves are given by

$$\boldsymbol{k}^{\pm o} = \begin{pmatrix} K \\ 0 \\ \pm q_o \end{pmatrix}, \tag{3.7.2.2.4}$$

where as usual $q_o = \sqrt{k_0^2 n_o^2 - K^2}$ is the normal component of the wavevector. We also know that the $\boldsymbol{D}$ field of an o-wave is perpendicular to both its wavevector and the optical axis of the medium, and furthermore, that the $\boldsymbol{E}$ and $\boldsymbol{D}$ fields are parallel. It follows that the normalised field orientations of the o-waves, as required by the Fresnel coefficients in section 2.5.7, are given by

$$\hat{\boldsymbol{E}}^{o\pm} = \frac{\boldsymbol{k}^{\pm o} \times \boldsymbol{c}}{|\boldsymbol{k}^{\pm o} \times \boldsymbol{c}|}. \tag{3.7.2.2.5}$$

The e-waves present a much greater challenge to determine. The refractive index changes with the direction of propagation, and so we cannot simply determine the normal component of the e-wave wavevectors as we did with the o-waves. Instead,

we must consider the wavenormal surface (2.47), which to remind the reader is

$$\frac{\tilde{k}_x^2 + \tilde{k}_y^2}{n_e^2} + \frac{\tilde{k}_z^2}{n_o^2} = k_0^2, \tag{3.7.2.2.6}$$

when written in the principal frame, $\tilde{\boldsymbol{x}}$, of the crystal. This surface defines the allowed wavevectors which satisfy Maxwell's equations in anisotropic media. Rewriting as the quadratic form

$$\tilde{\boldsymbol{k}}^T \mathrm{Q} \tilde{\boldsymbol{k}} = k_0^2, \tag{3.7.2.2.7}$$

where

$$\mathrm{Q} = \begin{pmatrix} n_e^{-2} & 0 & 0 \\ 0 & n_e^{-2} & 0 \\ 0 & 0 & n_o^{-2} \end{pmatrix}, \tag{3.7.2.2.8}$$

it can then be expressed in the lab frame as

$$(\mathrm{R}\boldsymbol{k})^T \tilde{Q} (\mathrm{R}\boldsymbol{k}) = k_0^2, \tag{3.7.2.2.9}$$

where

$$\mathrm{R} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \sin\phi & -\cos\phi \\ 0 & \cos\phi & \sin\phi \end{pmatrix} \begin{pmatrix} \sin\theta & 0 & -\cos\theta \\ 0 & 1 & 0 \\ \cos\theta & 0 & \sin\theta \end{pmatrix} \tag{3.7.2.2.10}$$

$$= \mathrm{R}_x \left(\frac{\pi}{2} - \phi\right) \mathrm{R}_y \left(\frac{\pi}{2} - \theta\right) \tag{3.7.2.2.11}$$

is the transformation that maps the optical axis the z axis. The e-wave wavevectors

$$\boldsymbol{k}^{e\pm} = \begin{pmatrix} K \\ 0 \\ q_{e\pm} \end{pmatrix} \tag{3.7.2.2.12}$$

can now be substituted into the wavenormal surface (3.7.2.2.9) to obtain a quadratic for $q_{e\pm}$. The solutions are

$$q_{e\pm} = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}, \tag{3.7.2.2.13a}$$

where

$$A = \frac{\mathrm{R}_{13}^2 - \mathrm{R}_{23}^2}{n_e^2} + \frac{\mathrm{R}_{33}^2}{n_o^2} \tag{3.7.2.2.13b}$$

$$B = 2K \left[\frac{\mathrm{R}_{11}\mathrm{R}_{13} + \mathrm{R}_{21}\mathrm{R}_{23}}{n_e^2} + \frac{\mathrm{R}_{31}\mathrm{R}_{33}}{n_o^2}\right] \tag{3.7.2.2.13c}$$

$$C = K^2 \left[\frac{\mathrm{R}_{11}^2 + \mathrm{R}_{21}^2}{n_e^2} + \frac{\mathrm{R}_{31}^2}{n_o^2}\right] - 1. \tag{3.7.2.2.13d}$$

Of the two solutions, the $+$ solution is positive and describes the forward propagating

mode, while the $-$ solution is negative and describes the backward propagating mode. The effective refractive index seen by the e-waves can be found from

$$n_{eff}^{\pm} = \frac{\sqrt{K^2 + q_{e\pm}^2}}{k_0}. \tag{3.7.2.2.14}$$

With the e-wave wavevectors now known, we can calculate the normalised fields needed for the Fresnel coefficients. For a given direction of propagation, the $\boldsymbol{D}$ fields of the associated o- and e-waves and the optical axis form an orthogonal triplet. Thus, the $\boldsymbol{D}$ fields of the e-waves are given by

$$\boldsymbol{D}^{e\pm} = \boldsymbol{k}^{e\pm} \times \boldsymbol{k}^{e\pm} \times \boldsymbol{c}, \tag{3.7.2.2.15}$$

and the normalised $\boldsymbol{E}$ fields are given by

$$\hat{\boldsymbol{E}}^{e\pm} = \frac{\varepsilon^{-1}\boldsymbol{k}^{e\pm} \times \boldsymbol{k}^{e\pm} \times \boldsymbol{c}}{|\varepsilon^{-1}\boldsymbol{k}^{e\pm} \times \boldsymbol{k}^{e\pm} \times \boldsymbol{c}|}, \tag{3.7.2.2.16}$$

where the permittivity tensor is given by

$$\varepsilon = \mathrm{R}^T \begin{pmatrix} n_o^2 & 0 & 0 \\ 0 & n_o^2 & 0 \\ 0 & 0 & n_e^2 \end{pmatrix} \mathrm{R}. \tag{3.7.2.2.17}$$

We now have everything we need to calculate the transmission and reflection of the PAAD layer, according to the formulae (3.5.4.2a)-(3.5.4.2b) and (3.5.4.6a)-(3.5.4.6b). It should be noted, however, that care must be taken when evaluating the Fresnel coefficients for waves heading backwards, i.e, t_{op}^{21} and r_{oe}^{21} etc. This is because the isotropic to anisotropic, and the anisotropic to isotropic coefficients were derived assuming that the waves were heading in the positive z direction. For these backwards coefficients, then, it is necessary to reverse the roles of the forward and backward waves, and to apply a rotation of π about the y axis to all quantities before entering them into the Fresnel formulae. The necessary changes are summarised by

$$\hat{\boldsymbol{E}}^{\alpha\pm} \to \begin{pmatrix} -\hat{E}_z^{\alpha\mp} \\ \hat{E}_y^{\alpha\mp} \\ -\hat{E}_x^{\alpha\mp} \end{pmatrix}, \tag{3.7.2.2.18a}$$

$$K \to -K, \tag{3.7.2.2.18b}$$

$$q_\alpha \to -q_\alpha \tag{3.7.2.2.18c}$$

$$n_{in} \leftrightarrow n_{out}, \tag{3.7.2.2.18d}$$

where α ranges over the $o\pm$-, $e\pm$-, $s\pm$-, and $p\pm$-waves.

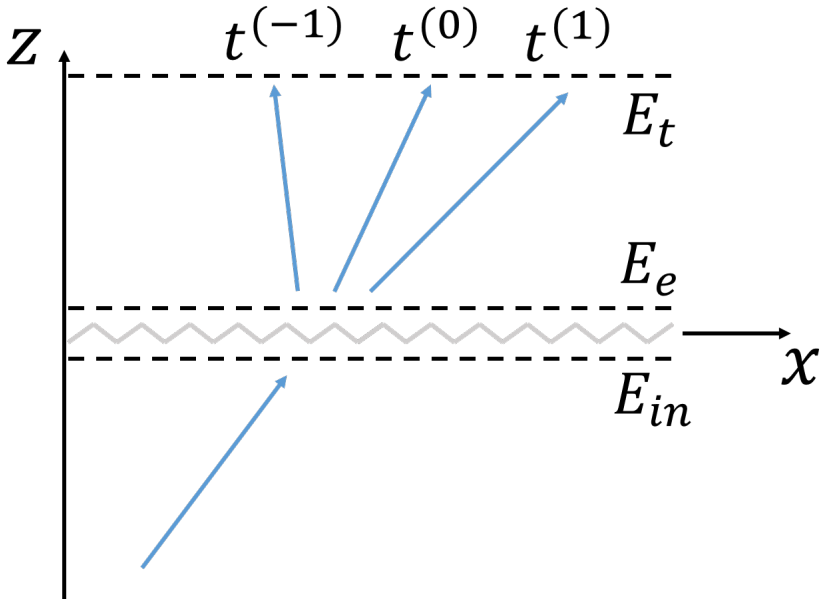


FIGURE 3.21: Schematic of the PAAD grating diffraction problem. The incident field, E_{in}, has its plane of incidence in the xz plane. The grating is treated as an effective layer (represented by a jagged line at $z = 0$) between the first and third layers. The emerging field, $E_e = E_t|_{z=0^+}$, is given by a square wave approximation applied to the isotropic and anistropic layer transmission coefficients, and gives rise to the transmitted field, E_t, comprised of the various diffracted orders described by their effective transmition coefficients $t^{(i)}$.

3.7.2.3 Diffraction from an Isotropic/Anisotropic Grating Layer

We now know the effective transmission coefficients t_{iso} and t_{ani} across an infinitely extending isotropic or anisotropic layer. Per our modelling assumption depicted earlier in figure 3.19, we know that the emerging field (the transmitted field, E_t, evaluated at the surface of the grating, $z = 0^+$) is given by

$$E_e = E_{in} \begin{cases} t_{ani}, & x \bmod \Lambda \leq f \\ t_{iso}, & \text{otherwise,} \end{cases} \tag{3.7.2.3.1}$$

where E_{in} is the field incident upon the grating at $z = 0^-$, and, for generality, we have included the factor f to describe the proportion of the grating that is anisotropic. We now look to find the effective transmission coefficients $t^{(i)}$ for the i'th diffracted order $E^{(i)} = t^{(i)} E_{in}$, where the grating is treated as an effective boundary between between the input and output layers, as shown in figure 3.21. The amplitude of the coefficients will be found using the non-paraxial scalar diffraction theory of section 2.7.4, while the angle of the coefficients will be found with the aid of the Fourier transform of E_e. This will allow us to compute E_t at an arbitrary distance z from the grating, and we can then verify the method numerically against the Rayleigh-Sommerfeld integrals in section 2.7.2.

The emerging field (3.7.2.3.1) can also be written as

$$E_e = E_{in}\left\{t_{iso} + (t_{ani} - t_{iso})\,\Pi\left(\frac{x}{f\Lambda}\right) * \frac{1}{\Lambda}\text{Ш}\left(\frac{x}{\Lambda}\right)\right\}, \tag{3.7.2.3.2}$$

where Π is the rectangle function

$$\Pi\left(\frac{x}{a}\right) = \begin{cases} 1, & \left|\frac{x}{a}\right| > \frac{1}{2}, \\ 0, & \text{otherwise}, \end{cases} \tag{3.7.2.3.3}$$

Ш is the comb function [15]

$$\text{Ш}\left(\frac{x}{a}\right) = |a| \sum_{n=-\infty}^{\infty} \delta\left(x - na\right), \tag{3.7.2.3.4}$$

and $*$ denotes convolution. Written in this way, we describe the emerging field as the sum of a constant field everywhere, given by the isotropic transmission, plus a square wave modulation due to the anisotropic transmission. Now, we must also consider that, due to the anisotropic layer, we have a mixing of the s- and p-polarisations emerging from the grating, and the diffracted field is the sum of two orthogonal fields. We therefore fully describe the emerging field by

$$\begin{pmatrix} E_e^{(p)} \\ E_e^{(s)} \end{pmatrix} = \left\{ \begin{pmatrix} t_{iso}^{pp} & 0 \\ 0 & t_{iso}^{ss} \end{pmatrix} + \left[\begin{pmatrix} t_{ani}^{pp} & t_{ani}^{sp} \\ t_{ani}^{ps} & t_{ani}^{ss} \end{pmatrix} - \begin{pmatrix} t_{iso}^{pp} & 0 \\ 0 & t_{iso}^{ss} \end{pmatrix} \right] \Pi\left(\frac{x}{f\Lambda}\right) * \frac{1}{\Lambda}\text{Ш}\left(\frac{x}{\Lambda}\right) \right\} \begin{pmatrix} E_{in}^{(p)} \\ E_{in}^{(s)} \end{pmatrix}, \tag{3.7.2.3.5}$$

where $E_e^{(p/s)}$ are the p- and s-polarised components of the emerging field (and analagously for $E_{in}^{(p/s)}$) , and $t_{iso/ani}^{\alpha\beta}$ denotes the isotropic/anistropic effective layer transmission coefficient for the α to β polarisation.

We apply the results of section 2.7.4, non-paraxial scalar diffraction using the angular spectrum, to each component of the emerging field. In general, each component is of the form

$$E_e = t_b + \Delta t \Pi\left(\frac{x}{f\Lambda}\right) * \frac{1}{\Lambda}\text{Ш}\left(\frac{x}{\Lambda}\right), \tag{3.7.2.3.6}$$

where $t_b = t_{iso}E_{in}$ is a constant background field and $\Delta t = t_{ani}E_{in} - t_b$ is the amplitude of the square wave modulation induced by the anisotropic segments of the grating. The field non-dimensionalises to

$$\tilde{E}_e = t_b + \Delta t \Pi\left(\frac{\tilde{x}}{f\tilde{\Lambda}}\right) * \frac{1}{\lambda\tilde{\Lambda}}\text{Ш}\left(\frac{\tilde{x}}{\tilde{\Lambda}}\right), \tag{3.7.2.3.7}$$

and we now seek to find the diffraction efficiencies of the various diffracted orders arising from this emerging field. The Fourier transforms of the rectangle and comb functions are [15]

$$\mathcal{F}_{k_x}\left\{\Pi\left(\frac{x}{a}\right)\right\} = a\frac{\sin\left(a\pi k_x\right)}{\pi k_x}$$

$$= a\text{sinc}\left(ak_x\right) \tag{3.7.2.3.8}$$

and

$$\mathcal{F}_{k_x}\left\{\frac{1}{a}\text{Ш}\left(\frac{x}{a}\right)\right\} = \text{Ш}\left(ak_x\right), \tag{3.7.2.3.9}$$

giving

$$\begin{aligned}\mathcal{F}_{k_x}\left\{\Pi\left(\frac{\tilde{x}}{f\tilde{\Lambda}}\right) * \frac{1}{\lambda\tilde{\Lambda}}\text{Ш}\left(\frac{\tilde{x}}{\tilde{\Lambda}}\right)\right\} &= f\tilde{\Lambda}\text{sinc}\left(f\tilde{\Lambda}\alpha\right)\text{Ш}\left(\tilde{\Lambda}\alpha\right) \\ &= f\text{sinc}\left(f\tilde{\Lambda}\alpha\right)\sum_{m=-\infty}^{\infty}\delta\left(\alpha - \frac{m}{\tilde{\Lambda}}\right) \\ &= \sum_{m=-\infty}^{\infty} f\text{sinc}\left(fm\right)\delta\left(\alpha - \frac{m}{\tilde{\Lambda}}\right),\end{aligned} \tag{3.7.2.3.10}$$

by definition of the comb function (3.7.2.3.4). We can therefore write

$$\tilde{E}_e = t_b + \Delta t\sum_{m=-\infty}^{\infty} f\text{sinc}\left(fm\right)e^{2\pi i\frac{m}{\tilde{\Lambda}}\tilde{x}}, \tag{3.7.2.3.11}$$

and it follows immediately that the autocorrelation is simply

$$\rho_X\left(\tilde{E}_t\right) = \left|t_b + f\Delta t\right|^2 + \left|\Delta t\right|^2\sum_{-\infty<m<0}^{0<m<\infty} f^2\text{sinc}^2\left(fm\right)e^{2\pi i\frac{m}{\tilde{\Lambda}}X}. \tag{3.7.2.3.12}$$

Taking the Fourier transform to find the electromagnetic spectral power density, S_α, and multiplying by γ, we find that the diffraction efficiencies, given by the radiant intensity distribution as in section 2.7.4, are given by

$$I\left(\alpha\right) = \frac{1}{2Z}\left\{\left|t_b + f\Delta t\right|^2\delta\left(\alpha\right) + \left|\Delta t\right|^2\sum_{-\infty<m<0}^{0<m<\infty} f^2\text{sinc}^2\left(fm\right)\gamma_m\delta\left(\alpha - \frac{m}{\tilde{\Lambda}}\right)\right\}. \tag{3.7.2.3.13}$$

As a sanity check, we calculate the total electromagnetic energy as

$$\begin{aligned}\frac{1}{2Z}\int_{-\infty}^{\infty} S_\alpha\left(\tilde{E}\right)d\alpha &= \frac{1}{2Z}\left\{\left|t_b + f\Delta t\right|^2 + \left|\Delta t\right|^2 f^2\left(\sum_{m=-\infty}^{\infty}\text{sinc}^2\left(fm\right) - 1\right)\right\} \\ &= \frac{1}{2Z}\left\{\left|t_b + f\Delta t\right|^2 + \left|\Delta t\right|^2 f^2\left(\frac{1}{f} - 1\right)\right\} \\ &= \frac{1}{2Z}\left(\left|t_b\right|^2 + f\left|t_{ani}\right|^2\right).\end{aligned} \tag{3.7.2.3.14}$$

As you would expect, the total energy is a weighted average of the energies of the emerging fields. The background field is present throughout the whole grating and has a weighting of 1, while the modulating field from the anisotropic layer emerges only from a fraction f of the grating.

We have the power that each diffracted order will propagate away from the grating,

giving us the phasor amplitude of each diffracted wave. To correctly describe the field, and not just the intensity distribution, we also need the phase. It seems a reasonable assumption that the transmitted propagating field E_t must be in phase with the emerging field E_e. We therefore need the m'th order of the transmitted field to have the amplitude

$$\left|\tilde{E}_t^{(m)}\right| = \sqrt{\frac{1}{2Z}} \begin{cases} |t_b + f\Delta t|\,, & m = 0, \\ |\Delta t|\, f\mathrm{sinc}\,(fm)\,\sqrt{\gamma_m}, & m \neq 0, \end{cases} \tag{3.7.2.3.15}$$

and the phase

$$\arg\left(\tilde{E}_t^{(m)}\right) = \begin{cases} \frac{t_b+\Delta t f}{|t_b+\Delta t f|}, & m = 0, \\ \frac{\Delta t}{|\Delta t|}, & m \neq 0, \end{cases} \tag{3.7.2.3.16}$$

as given by the Fourier transform (3.7.2.3.11). Multiplying the two together we get

$$\tilde{E}_t^{(m)} = \sqrt{\frac{1}{2Z}} E_{in} \begin{cases} [(1-f)t_{iso} + f t_{ani}]\, e^{2\pi i \tilde{z}}, & m = 0, \\ (t_{ani} - t_{iso}) f\mathrm{sinc}\,(fm)\,\sqrt{\gamma_m} e^{2\pi i(\alpha_m \tilde{x} + \gamma_m \tilde{z})}, & m \neq 0, \end{cases} \tag{3.7.2.3.17}$$

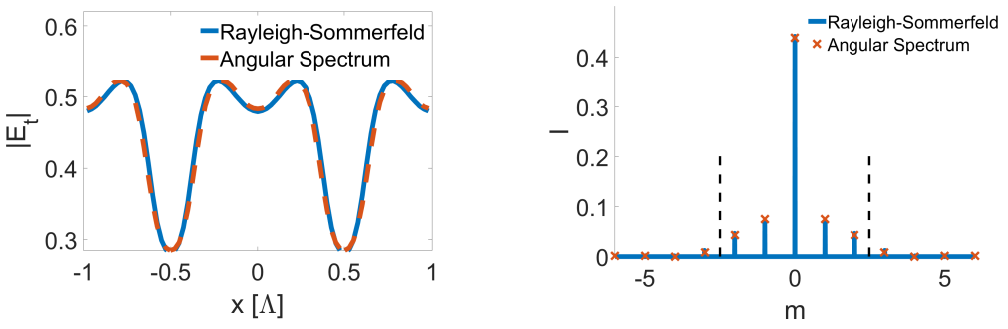
where

$$\alpha_m = \frac{m}{\tilde{\Lambda}} \tag{3.7.2.3.18}$$

$$\gamma_m = \sqrt{1 - \alpha_m^2}. \tag{3.7.2.3.19}$$

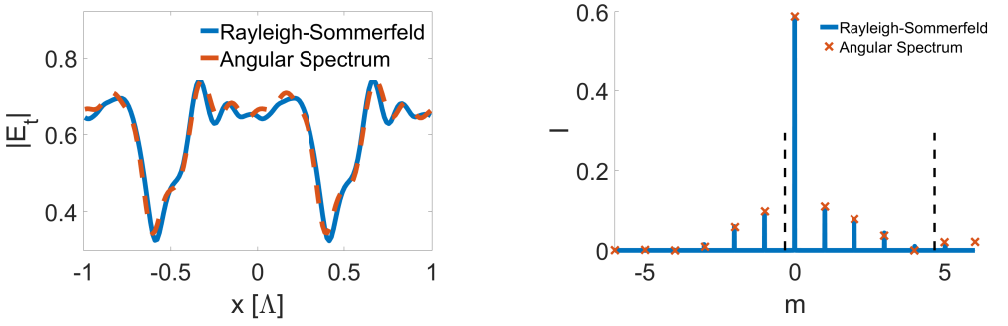
We compare these results to the numerical solution given by the first Rayleigh-Sommerfeld integral (2.194a). We choose a grating with $f = 3/4$ and a grating spacing $\Lambda = 2.5\lambda$, in keeping with the approximate dimensions of the PAAD grating. The propagation distance is set to 0.224λ, which is the thickness of the ITO from the fitting, and the refractive indices of the incident and exit media is set to $n_{in} = n_{out} = 1$. We choose the transmission coefficients arbitrarily, setting $t_b = 0.5$ and $\Delta t = 0.7e^{i\frac{\pi}{4}}$. Figure 3.22 shows a comparison of the total diffracted field $E_t = \sum_m E_t^{(m)}$ calculated with the two methods, where we see excellent agreement. The magnitude of the diffracted orders according to the Rayleigh-Sommerfeld method are found by applying the FFT to the calculated field. We now apply the results for diffraction at non-normal incidence, and also consider that the input and exit media have different refractive indices. Assuming that the incident field has direction cosine α_0' in the input medium, the phase variation across the emerging field is $e^{2\pi i \frac{n_{in}}{n_{out}} \alpha_0' \tilde{x}}$ when non-dimensionalised according to $\lambda_2 = \lambda_0 n_2^{-1}$. Using the same method as for normal incidence, the field is found to be

$$\tilde{E}_t = \sqrt{\frac{1}{2Z}} E_{in} \sqrt{\gamma_m} e^{2\pi i(\alpha_m \tilde{x} + \gamma_m \tilde{z})} \begin{cases} (1-f)t_{iso} + f t_{ani}, & m = 0, \\ (t_{ani} - t_{iso}) f\mathrm{sinc}\,(fm)\,, & m \neq 0, \end{cases} \tag{3.7.2.3.20}$$



(A) (B)

FIGURE 3.22: A comparison of the diffracted field as calculated by the Angular Spectrum and Rayleigh-Sommerfeld methods at normal incidence. (a) Shows the norm of the electric field while (b) shows the magnitude of the individual orders, with the black dashed lines marking the evanescent cut-off. The system parameters are $f = 3/4$, $\Lambda = 2.5\lambda$, and $z = 0.224\lambda$.



(A) (B)

FIGURE 3.23: A comparison of the diffracted field as calculated by the Angular Spectrum and Rayleigh-Sommerfeld methods at a 60° incidence angle. (a) Shows the norm of the electric field while (b) shows the magnitude of the individual orders, with the black dashed lines marking the evanescent cut-off. The system parameters are $f = 3/4$, $\Lambda = 2.5\lambda$, and $z = 0.224\lambda$.

where

$$\alpha_m = \left(n_{in}\alpha_0' + \frac{m}{\tilde{\Lambda}} \right) n_{out}^{-1}, \tag{3.7.2.3.21}$$

$$\gamma_m = \sqrt{1 - \alpha_m^2}. \tag{3.7.2.3.22}$$

Figure 3.23 shows a comparison of the diffracted fields for an incident angle of $\alpha_0' = 60°$, $n_{in} = 1$, and $n_{out} = 2$. The transmission coefficients t_{ani} and t_b are as in the normal incidence case. The quality of agreement between the two methods is much the same.

3.7.2.4 Diffraction from an Aligned PAAD Grating

We now have an accurate description of how each of the s- and p-components that emerge from an isotropic/anisotropic grating are further split up into their diffracted orders. In the formalism of the generalised multi-mode ray method, we now consider the full experimental system as a three layer system comprised of air, ITO, and then air again. The PAAD layer acts as a effective boundary between the first air and ITO layers, and the glass layer acts as an effective boundary between the ITO and second air layers. The results of the previous section are used for the effective transmission and reflection coefficients to describe the mode mixing at the PAAD effective boundary, and the isotropic formulae of section 2.6 are used at the glass effective boundary.

We will assume that the factor f of the experimental PAAD gratings is $\frac{1}{2}$, which seems reasonable given the symmetries of the interference pattern formed by the two recording beams. By equation (3.7.2.3.20), this means that all even orders have zero diffraction efficiency when the incident beam first passes through the grating.

Furthermore, since we are only interested in the behaviour of the first orders, it is unnecessary to model any orders higher than the first. Whenever a ray reflects from the effective glass layer, it heads back towards the grating where the modes will again be mixed upon reflection. From equation (3.7.2.3.20), we see that any time a ray diffracts (i.e., it reflects back to a different order and $m \neq 0$) the amplitude of the diffracted ray is proportional to Δt times the amplitude of the incident ray. We know that Δt must be small since the experimental results were $O(10^{-4})$, and so the effects of higher diffracted orders on the first orders would be at least $O(\Delta t^2)$. We conclude it is safe to neglect these diffraction effects from higher orders. Moreover, we also do not model diffraction effects in the first orders for the same reason, and any diffracted orders can only reflect back to the same order, according to the $m = 0$ case of equation (3.7.2.3.20). This is all visualised and described in figure 3.24.

The fundamental and first order modes that enter into the ITO when the incident beam first crosses through the grating can be described in the formalism of the generalised multi-mode ray method by the mode matrix

$$\mathrm{F}^{(3)} = \begin{pmatrix} F_{p0}^{(3)(p)} & F_{p0}^{(3)(s)} \\ F_{s0}^{(3)(p)} & F_{s0}^{(3)(s)} \\ F_{p-1}^{(3)(p)} & F_{p-1}^{(3)(s)} \\ F_{s-1}^{(3)(p)} & F_{s-1}^{(3)(s)} \\ F_{p+1}^{(3)(p)} & F_{p+1}^{(3)(s)} \\ F_{s+1}^{(3)(p)} & F_{s+1}^{(3)(s)} \end{pmatrix}. \tag{3.7.2.4.1}$$

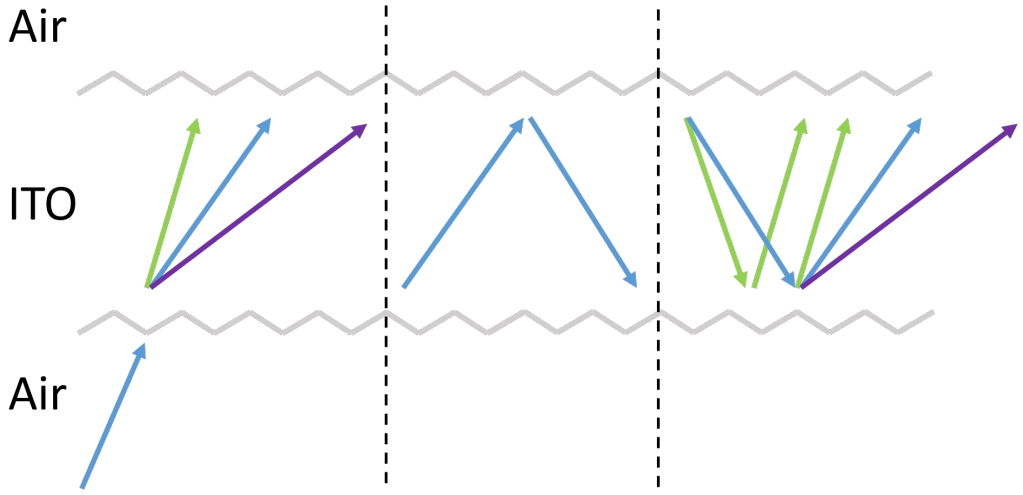


FIGURE 3.24: A schematic of the model of the full experimental system. The system is represented as three layers, air, ITO, and air, separated by effective boundaries (jagged lines) that account for the effects of the PAAD and glass layers. We neglect diffracted orders higher than the first, and the left section shows the fundamental order (blue) crossing the effective PAAD boundary and entering the ITO layer after diffracting into the negative (green), fundamental, and positive (purple) orders (with both s- and p-polarisations of each order making up the 6 modes of the model). The second effective boundary accounting for the glass layer is a purely isotropic subsystem, so any mode reflects back as the same mode, shown in the centre section. We ignore higher orders and repeat diffraction effects, so a backwards fundamental mode reflects and diffracts to fundamental, positive and negative modes, while a backwards diffracted mode can only reflect back to the same mode. This is shown in the right section.

The component $F_{p0}^{(3)(p)}$ is the amplitude of the forward propagating 0 order p-wave in the third (ITO) layer assuming the incident field is p-polarised. The component $F_{s-1}^{(3)(s)}$ is the amplitude of the -1 order s-wave in the third layer assuming the incident field is s-polarised, etc. We can apply (3.7.2.3.20) with the relevant effective coefficients to determine each of the modes as

$$\mathrm{F}^{(3)} = \mathrm{I}\Delta\mathrm{t}, \tag{3.7.2.4.2}$$

where the matrix I is diagonal and contains the relevant radiant intensity values for the given order, that is,

$$\mathrm{I} = \sqrt{\frac{n_3}{8}}\mathrm{diag}\begin{pmatrix} \sqrt{\gamma_0} \\ \sqrt{\gamma_0} \\ \mathrm{sinc}\left(\frac{1}{2}\right)\sqrt{\gamma_{-1}} \\ \mathrm{sinc}\left(\frac{1}{2}\right)\sqrt{\gamma_{-1}} \\ \mathrm{sinc}\left(\frac{1}{2}\right)\sqrt{\gamma_{+1}} \\ \mathrm{sinc}\left(\frac{1}{2}\right)\sqrt{\gamma_{+1}} \end{pmatrix}, \tag{3.7.2.4.3}$$

The second option ($\theta = 0$ and $\phi = \pi/2$) can be ruled out. In this case the model gives larger diffraction efficiencies for the input p-polarisations. This leaves the range of possible alignments where $\theta = \pi/2$. For small ϕ, where the optical axis is upwards in

the direction of the y axis, we also obtain that the p-polarised diffraction efficiencies are larger. For large ϕ, where the optical axis is close to normal to the cell faces, the diffraction efficiencies are approximately equal between the s- and p-polarisations. As ϕ is varied from 0 to $\pi/2$ the maxima of the p-polarised and s-polarised curves come together, with the s-polarised curves become higher at around 40°, before coming back together at 90°and the matrix Δt contains the relevant combinations of effective transmission coefficients,

$$\Delta \mathrm{t} = \begin{pmatrix} t_{13}^{pp} + t_{13}^{p} & t_{13}^{sp} \\ t_{13}^{ps} & t_{13}^{ss} + t_{13}^{s} \\ t_{13}^{pp} - t_{13}^{p} & t_{13}^{sp} \\ t_{13}^{ps} & t_{13}^{ss} - t_{13}^{s} \\ t_{13}^{pp} - t_{13}^{p} & t_{13}^{sp} \\ t_{13}^{ps} & t_{13}^{ss} - t_{13}^{s} \end{pmatrix}. \tag{3.7.2.4.4}$$

The left hand column of Δt assumes that the input polarisation is p-polarised, while the right column assumes it is s-polarised. The components t_{13}^{pp} are the effective transmission coefficients of the p- to p- polarisation from layers 1 to 3 across the anistropic part of the layer. The components t_{13}^{s} are the effective transmission coefficients of the s-polarisation across the isotropic part of the layer, etc.

3.7.2.5 Modelling the Experimental System

We now complete the full model of the experimental system by adding the glass layer after the ITO. We model the ITO/glass/air subsystem as an effective layer using the multi-mode formalism. Since all the materials involved are isotropic, we construct a diagonal 6x6 matrix whose entries are the isotropic three layer effective coefficients for each of the modes.

The transmission matrix from layer 3 (ITO) to layer 5 (air) is

$$\mathrm{T}_{35} = \mathrm{diag} \begin{pmatrix} t_{35}^{p}(\alpha_0) \\ t_{35}^{s}(\alpha_0) \\ t_{35}^{p}(\alpha_{-1}) \\ t_{35}^{s}(\alpha_{-1}) \\ t_{35}^{p}(\alpha_{+1}) \\ t_{35}^{s}(\alpha_{+1}) \end{pmatrix}, \tag{3.7.2.5.1}$$

where α_0 is the tangential wavevector component of the incident field, and $t_{35}^{p}(\alpha_{\pm 1})$ is the isotropic effective transmission coefficient from layer 3 to layer 5 for a p-wave with

direction cosine $\alpha_{\pm 1} = n_3^{-1}\left(\alpha_0 \pm \tilde{\Lambda}^{-1}\right)$, etc. The corresponding reflection matrix is

$$\mathrm{R}_{35} = \mathrm{diag}\begin{pmatrix} r_{35}^{p}(\alpha_0) \\ r_{35}^{s}(\alpha_0) \\ r_{35}^{p}(\alpha_{-1}) \\ r_{35}^{s}(\alpha_{-1}) \\ r_{35}^{p}(\alpha_{+1}) \\ r_{35}^{s}(\alpha_{+1}) \end{pmatrix}, \tag{3.7.2.5.2}$$

and the phase changes when traveling from one side of the ITO layer to the other are

$$\Phi_3 = \mathrm{diag}\begin{pmatrix} e^{ik_0 L_3\sqrt{n_3^2-\alpha_0^2}} \\ e^{ik_0 L_3\sqrt{n_3^2-\alpha_0^2}} \\ e^{ik_0 L_3\sqrt{n_3^2-\alpha_{-1}^2}} \\ e^{ik_0 L_3\sqrt{n_3^2-\alpha_{-1}^2}} \\ e^{ik_0 L_3\sqrt{n_3^2-\alpha_{+1}^2}} \\ e^{ik_0 L_3\sqrt{n_3^2-\alpha_{+1}^2}} \end{pmatrix}, \tag{3.7.2.5.3}$$

where L_3 and n_3 are the thickness and refractive index of the ITO layer, respectively.

We now need to consider the effective reflection matrix of the grating for rays that are heading in the backwards direction. As in calculating the initial transmission across the grating, we neglect second order diffraction effects. While fundamental orders can reflect back to any of the modes, the diffracted orders can only reflect back to the same diffracted order, though they can change polarisation. The multi-mode effective layer reflection matrix is then given by

$$\mathrm{R}_{31} = \mathrm{I}\Delta\mathrm{r}, \tag{3.7.2.5.4}$$

where I is the radiant intensity matrix defined in (3.7.2.4.3), and

$$\Delta\mathrm{r} = \begin{pmatrix} r_{31}^{pp}(-\alpha_0)+r_{31}^{p}(-\alpha_0) & r_{31}^{sp}(-\alpha_0) & 0 \\ r_{31}^{ps}(-\alpha_0) & r_{31}^{ss}(-\alpha_0)+r_{31}^{s}(-\alpha_0) & 0 \\ r_{31}^{pp}(-\alpha_0)-r_{31}^{p}(-\alpha_0) & r_{31}^{sp}(-\alpha_0) & r_{31}^{pp}(-\alpha_{-1})+r_{31}^{p}(-\alpha_{-1}) \\ r_{31}^{ps}(-\alpha_0) & r_{31}^{ss}(-\alpha_0)-r_{31}^{s}(-\alpha_0) & r_{31}^{ps}(-\alpha_{-1}) \\ r_{31}^{pp}(-\alpha_0)-r_{31}^{p}(-\alpha_0) & r_{31}^{sp}(-\alpha_0) & 0 \\ r_{31}^{ps}(-\alpha_0) & r_{31}^{ss}(-\alpha_0)-r_{31}^{s}(-\alpha_0) & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ r_{31}^{sp}(-\alpha_{-1}) & 0 & 0 \\ r_{31}^{ss}(-\alpha_{-1})+r_{31}^{s}(-\alpha_{-1}) & 0 & 0 \\ 0 & r_{31}^{pp}(-\alpha_{+1})+r_{31}^{p}(-\alpha_{+1}) & r_{31}^{sp}(-\alpha_{+1}) \\ 0 & r_{31}^{ps}(-\alpha_{+1}) & r_{31}^{ss}(-\alpha_{+1})+r_{31}^{s}(-\alpha_{+1}) \end{pmatrix}. \tag{3.7.2.5.5}$$

The ij element of R_{31} describes the effective reflection of mode i when mode j is incident upon the grating. The modes are ordered as in $F^{(3)}$, the input mode matrix (3.7.2.4.1) into the ITO layer.

Once again it is important to remember that the reflection coefficients are formulated assuming that the direction of travel is in the positive z direction (using the same coordinates as section as 3.7.2.2 where we derived the o- and e-wave fields in terms of the optical axis). To evaluate the reflection coefficients, where the waves are incident from the far side of the layer, we must view the system as rotated around the y axis by π. Firstly, this means evaluating both the isotropic and anistropic effective coefficients with negative values of the tangential wavevector components (since the x axis has been reversed). The s-polarisation is unchanged and our earlier stipulation that the positive orientation of the p-polarisation is always to the right of the wave vector ensures a consistent positive orientation between the forwards and backwards waves. For the anisotropic coefficients, the fields, and hence the effective coefficients, need to be evaluated according to (3.5.4.6a)-(3.5.4.6b) after switching $\phi \to \pi - \phi$.

These matrices can now be assembled to describe the transmission of the various modes across the experimental system. The output mode matrix in the exit air layer is

$$\mathrm{F}^{(5)} = \begin{pmatrix} F_{p0}^{(5)(p)} & F_{p0}^{(5)(s)} \\ F_{s0}^{(5)(p)} & F_{s0}^{(5)(s)} \\ F_{p-1}^{(5)(p)} & F_{p-1}^{(5)(s)} \\ F_{s-1}^{(5)(p)} & F_{s-1}^{(5)(s)} \\ F_{p+1}^{(5)(p)} & F_{p+1}^{(5)(s)} \\ F_{s+1}^{(5)(p)} & F_{s+1}^{(5)(s)} \end{pmatrix}, \tag{3.7.2.5.6}$$

where, as in (3.7.2.4.1), the component $F_{p-1}^{(5)(p)}$ is forward (output) p-polarised -1 diffracted order in layer 5 assuming that the input was p-polarised, etc. The output mode matrix is found according to

$$\mathrm{F}^{(5)} = \mathrm{T}_{35} \left[\mathcal{I} - \Phi_3 \mathrm{R}_{31} \Phi_3 \mathrm{R}_{35}\right]^{-1} \Phi_3 \mathrm{F}^{(3)}. \tag{3.7.2.5.7}$$

The output mode matrix is dependent on the direction of the incident field, that is, we can write $\mathrm{F}^{(5)} = \mathrm{F}^{(5)}(\alpha_0)$. Accounting for the Gaussian nature of the beam, we once again average the output field over the spectrum of the Gaussian beam to find the radiance of each mode as

$$\mathrm{L}(\alpha_0) = \int \left|\mathrm{F}^{(3)}(\alpha)\right|^2 e^{-\left(\frac{z_R}{4\pi}\right)(\alpha-\alpha_0)^4} d\alpha, \tag{3.7.2.5.8}$$

where the modulus squared is applied element-wise to $\mathrm{F}^{(3)}$. The experimental measurements were of the intensity of the diffracted orders, independent of

polarisation. Therefore, the quantities that are compared to the experimental results are the radiant intensities of the total field of each order:

$$\eta_{-1}^{(p)}(\alpha_0) = \frac{1}{2}\gamma_{-1}\left(\mathrm{L}_{31}^{(5)} + \mathrm{L}_{41}^{(5)}\right) \tag{3.7.2.5.9a}$$

$$\eta_{+1}^{(p)}(\alpha_0) = \frac{1}{2}\gamma_{+1}\left(\mathrm{L}_{51}^{(5)} + \mathrm{L}_{61}^{(5)}\right) \tag{3.7.2.5.9b}$$

$$\eta_{-1}^{(s)}(\alpha_0) = \frac{1}{2}\gamma_{-1}\left(\mathrm{L}_{32}^{(5)} + \mathrm{L}_{42}^{(5)}\right) \tag{3.7.2.5.9c}$$

$$\eta_{+1}^{(s)}(\alpha_0) = \frac{1}{2}\gamma_{+1}\left(\mathrm{L}_{52}^{(5)} + \mathrm{L}_{62}^{(5)}\right), \tag{3.7.2.5.9d}$$

where $\eta_{-1}^{(p)}(\alpha_0)$ is the diffraction efficiency of the -1 order for a p-polarised field incident at α_0, etc, and

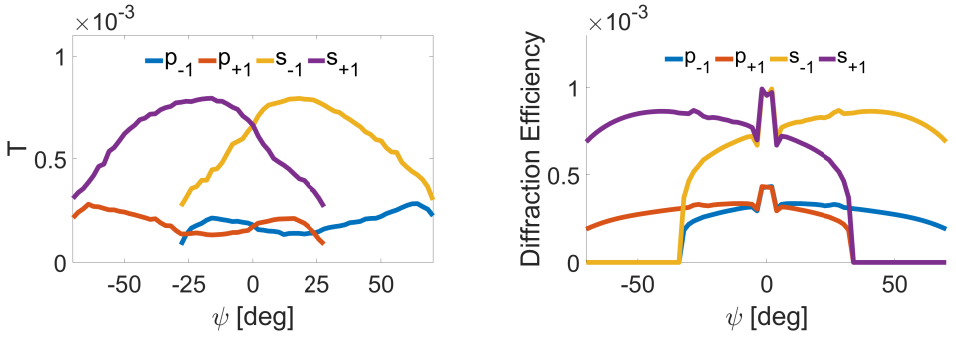
$$\gamma_{\pm 1} = \sqrt{1 - n_3^2\alpha_{\pm 1}^2}. \tag{3.7.2.5.10}$$

As before a second Gaussian smoothing was then applied to the data.

3.7.2.6 Comparison with Experimental Results

With a complete model describing the multi-layer diffraction problem of thin photo-aligned PAAD films, we can now compare with the experimental results, depicted in figure (3.18), to see what we can deduce about the molecular alignment of the anisotropic sections of the grating. The first feature to note is the symmetry between the positive and negative orders. This tells us that the optical axis must be confined to (or close to) one of two orientations. The first option is that it is in the plane perpendicular to both the plane of the cell and the plane of incidence ($\theta = \pi/2$). The second option is that the optical axis is parallel to the plane of the cell and in the plane of incidence ($\theta = 0$ and $\phi = \pi/2$). This is a direct consequence of the theory on propagation in anistropic media. The various o-waves that propagate inside the aligned PAAD are orthogonal to the wavevectors and the optical axis. The e-waves are in the plane of the optical axis and wavevector. These fields can only be symmetric with respect to the angle of incidence if the optical axis is in one the two listed configurations. It is necessary for the fields to be symmetric since, if they are not, it follows that the effective Fresnel coefficients and the diffraction efficiencies will also be asymmetric.

It is found that for a fixed alignment of the optical axis, the diffraction efficiency curves seem to maintain their shape, and only scale in magnitude as the birefringence is varied. Therefore, we next look for alignments of the optical axis that give the right relative amplitudes between the s- and p-polarised diffraction efficiencies, and also the right shape trends. This means looking for alignments which give s-polarised curves that are single large humps, and p-polarised curves that form an 'N' shape. We also require that the maximum of the s-polarised curves is around 3 to 4 times larger



(A) (B)

FIGURE 3.25: (a) The experimental diffraction efficiencies and (b) the best fit results from the theoretical model. The optical axis is at $\theta = 90°$, $\phi = 22°$, all other system parameters are according to the isotropic fit results.

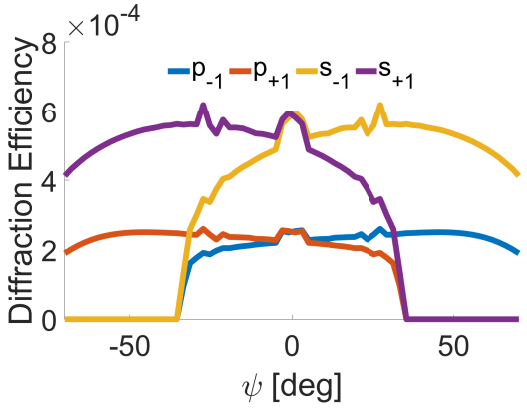


FIGURE 3.26: The theoretical diffraction efficiencies for the optical axis at $\theta = 90°$, $\phi = 22°$, $L_{ITO} = 100$ nm, and all other system parameters according to the isotropic fit results.

than the p-polarised. The closest match in terms of relative magnitudes (using the parameters from the isotropic fitting) is obtained around $\phi = 22°$. Furthermore, if a birefringence value of 0.185 is used, as found in the other experimental work of [24], the maximum values of the diffraction efficiency are also in very good agreement, this is shown in figure (3.25), along with the experimental data once again.

The overall shapes of the curves, however, are not in the best agreement. The shapes also do not seem to vary much as the system parameters are changed, with the exception of the optical axis and the ITO thickness. Even after performing both of the averaging processes, the theoretical transmissions are still quite susceptible to resonance effects. Figure (3.26) shows the theoretical transmissions when the ITO thickness is set to 100 nm (instead of 142 nm), which is within the common range for ITO/glass substrates available from manufacturers. With this one small change we see far better agreement between the curves. The large resonance in the centre seems to disappear and the peaks describing the s-polarised data become somewhat narrower

and in better shape agreement. The 'N' shape of the experimental p-polarised data can also be seen forming. Given that we have already seen that the ITO can be troublesome (it proved impossible to fit the transmission data for a single ITO/glass slide) it is likely that the ITO layer is still causing problems now. Regardless, even after the change of thickness the dependence on the PAAD variables remains mostly unchanged, and an angle around $\phi = 22^\circ$ still seems to give the best fit, suggesting that this is most likely the best estimate this model can provide.

We also have one last validation we can perform by considering the lack of surface relief seen in the PAAD layers. The z component of the optical axis is given by (3.7.2.2.1) as

$$c_z = \sin\phi \sin\theta. \tag{3.7.2.6.1}$$

In the isotropic PAAD the molecules are uniformly distributed in ϕ and θ, and it follows that the average "thickness" of each molecule is proportional to

$$\frac{1}{\pi^2}\int_0^\pi \int_0^\pi \sin\phi \sin\theta d\phi d\theta = \frac{4}{\pi^2}. \tag{3.7.2.6.2}$$

That is, for a molecular length l, the average contribution to the thickness of the PAAD layer is $\frac{4l}{\pi^2}$ for each molecule. We now consider the "thickness" of the aligned molecules. Maintaining our assumption that the molecules must be aligned in the normal plane to the cell ($\theta = \pi/2$), we find that the ϕ angle required to give the same thickness (and there be no surface relief) is

$$\phi = \sin^{-1}\left(\frac{4}{\pi^2}\right) = 24^\circ, \tag{3.7.2.6.3}$$

in very good agreement with the diffraction efficiency results.

While we have not been able to exactly replicate the experimental measurements, this is perhaps not surprising given the number of unknowns in the system. Possibly the main issue to consider to obtain more accurate results, though, is the use of a square wave approximation for the grating. Some very rough simulations simply replacing the sinc functions with Bessel functions in the diffraction efficiency calculations shows greater modulation in both polarisations with varying angle of incidence. This leads to the ability to produce more 'N' shaped curves for the p-polarisation and higher arched s-polarised data. As in the case of Raman-Nath diffraction, considered in section 2.7.3, it is well known that sinusoidal modulations of the emerging field give rise to diffracted orders governed by the Bessel functions. It seems that, given the intensity pattern of the grating recording beams is sinusoidal, the approximation of a square wave governing the grating transmission is not good enough for accurate simulation of the diffraction process. How to go about calculating the transmission across a sinusoidally varying anistropic layer is not obvious, and a possible area for future work. However, from the theory and understanding gained of anistropic media

throughout this work, it is hard to see that the general requirement of symmetry between the optical axis and angle of incidence can be relaxed if symmetric diffraction data is to be obtained. It would seem, then, that the results of square wave diffraction give us at least a reasonable estimate of the PAAD photoalignment behaviour, especially given the agreement between various system parameters across different experimental systems. It is also worth noting that, if further work is to be completed on this topic, the completion of more experiments measuring the separate polarisations of the diffracted orders would also prove highly beneficial.

Chapter 4

Multiscale Modelling of Doped Liquid Crystal Systems

4.1 Introduction

This final part of this thesis concerns the multiscale modelling of dielectric particles in liquid crystal. Although perhaps a focused topic, the results are, in principle, very general to many anistropic systems. Modelling the average behaviour of composite systems has been an area of interest for an incredibly long time, with a multitude of existing effective medium theories used in science and engineering. The majority of these are based on physical principles and scientific theory. In this thesis my work on using the method of homogenisation to consider this problem is presented. Homogenisation is an entirely mathematical principle used in the study of PDEs with fast oscillating coefficients. However, when solving Maxwell's equation $\boldsymbol{\nabla} \cdot (\varepsilon \boldsymbol{\nabla} \phi) = 0$ in the presence of a micro-structure, we have, assuming periodicity, exactly the homogenised problem, with the permittivity tensor oscillating rapidly between the permittivity of the host and the inclusions. The result of the homogenised problem is an effective permittivity tensor, ε_{eff}, which describes a smooth, slow varying electric field in a single, homogeneous, effective medium. The effective medium behaves as the average of composite system, and solving for the field inside the effective medium is virtually instantaneous, as opposed to incredibly time consuming when the full microscopic behaviour is considered.

We present a study, then, of homogenisation as an alternative to traditional theories. In this part of the thesis, the homogenised equations for dielectric particles in a host medium are developed. This leads to the so-called cell problem, a sub-problem that separates from the main equations, and is responsible for describing the effects of the micoscopic geometry. A numerical solution of the cell problem is first considered to validate the accuracy of the homogenised solution against full microscopic simulation.

Analytic approximations to the cell problem are then considered, based on the assumption of low particle volume fraction. This is done for 2D elliptical particles in anisotropic media, and then for 3D spheroidal particles in isotropic media. We then proceed to relax the assumption of periodicity to see if homogenisation is still applicable when there is disorder in the particles arrangements, all in comparison with traditional theories and numerical simulations.

4.2 Homogenisation and Effective Medium Theories

A prevalent method employed to simplify the solution of such host/inclusion problems is the use of an effective medium theory. This is a method in which the composite system of materials with differing properties is reduced to a system comprised of one single "effective" medium. This effective medium has no micro-structure but, in principle, has the same macroscopic properties as the composite system. In the case of modelling the electric field inside composite systems, this amounts to reducing the permittivity, which varies microscopically between the different domains, into an effective permittivity for the system as a whole. This is, in fact, a problem that has been the subject of considerable research over the years, with many different theories relying entirely on, or being variations of, the theories of Maxwell-Garnett or Bruggeman [10]. In each of these theories, the model is formed by considering a single inclusion lying within a homogeneous host material. The interaction of the inclusion with the electric field in the host then gives rise to the final form of the theory. In this way, the theories are essentially mean field theories, assuming that each inclusion interacts with the average field of the remainder of the system. The way in which they accomplish this differs, however, and gives rise to theories that, although in some ways are similar, produce some very different results.

Maxwell-Garnett, for instance, follows as a direct consequence of the Lorentz local field concept and the Clausius-Mossotti relation, which assumes that all inclusions experience the same local field. This essentially amounts to volume averaging of the dielectric permittivities with the assumption that the polarisability of the inclusions is known, i.e., that all of the molecules are of the same shape, usually a sphere or ellipsoid. The method proves to be accurate to first order in the volume fraction of the inclusions and, owing to the simplicity of its calculation, can prove a very efficient theory. It also proves to be extendable to more complicated problems such as randomly oriented and distributed anisotropic particles [32], and can be shown to be valid for arbitrary particle shapes (provided the medium is, on average, uniform and isotropic) [34]. It does, however, possess some quirks. While simply relabeling the host as the inclusion and the inclusions as the host (while maintaining the volume fractions) should give the same results, the Maxwell-Garnett formula gives vastly

differing results. This is due to the way the formula/derivation treat the different components asymmetrically.

The theory by Bruggeman [6] on the other hand, does treat the components of the system symmetrically. It considers the boundary flux of the fields inside each medium, and posits that the average deviation of these fluxes away from that of the homogenous field must be zero [10]. In this way, the model can be easily extended to mixtures of multiple kinds of inclusion with no restrictions on their respective volume fractions or which is considered the host or inclusion. Moreover, this also allows for the prediction of a critical volume fraction threshold in which an insulating system becomes conducting with enough conducting inclusions being added etc (which is also not present in Maxwell-Garnett theory). Although the answer agrees qualitatively with experimental results and with theoretical results from percolation theory, it often predicts critical thresholds that are too high, and gives other strange results such as a near zero threshold for needle shaped particles etc [10]. Like Maxwell-Garnett, it is also accurate only to first order in the volume fraction and can be viewed as volume averaging.

The method used in this work takes a somewhat different approach to the above methods and their variants. The method in question, first presented in [25], is known as homogenisation and, given that the solution it provides is a macroscopic approximation to a system possessing a microstructure, can be viewed as an effective medium theory when used in contexts such as these. The way in which it does so, however, is far less physical and completely mathematical. As defined in [41], homogenisation is a second-order perturbation theory for linear differential operators, while averaging, on the other hand, is a first-order theory. This agrees with the above observations that Maxwell-Garnett and Bruggeman are volume averaging techniques. Formally, homogenisation is a method for approximating the solution to problems of the form

$$\mathcal{L}^\eta u^\eta = f \quad \text{or} \quad \frac{\partial u^\eta}{\partial t} = \mathcal{L}^\eta u^\eta,$$

where

$$\mathcal{L}^\eta = \frac{1}{\eta^2}\mathcal{L}_0 + \frac{1}{\eta}\mathcal{L}_1 + \mathcal{L}_2$$

is a linear differential operator and $\eta \ll 1$. The homogenised solution u converges exactly to the true solution u^η in the limit $\eta \to 0$.

In the case of particles in medium, the problem arrives in this form under the assumption (or approximation) that the microstructure, and hence the permittivity, is periodic, with the period of the permittivity naturally defining the length scale η. A power series solution in η is then sought that depends on the macroscopic position $\boldsymbol{x}$ and the microscopic coordinate $\boldsymbol{y} = \boldsymbol{x}/\eta$, which are assumed to be independent. Thus, the problem is explicitly defined on the slow varying macroscopic ($\boldsymbol{x}$) and fast varying microscopic ($\boldsymbol{y}$) scales. The homogenised solution first yields a leading order

background solution as a necessary consequence of the separation of scales and periodicity. The microscopic interactions then appear at second order as a correction to this macroscopic field. The nature of the corrections are determined at first order through the solution of the so called cell problem, a derived system of equations on a representative volume of medium containing a single inclusion. This volume matches exactly with the assumed periodic geometry, and herein lies the fundamental difference between the schemes. Under the periodic boundary conditions of the cell problem we are assuming that all of the particles interact directly with one another to produce a correction to the background field. This is in contrast to the traditional effective medium theory techniques; they assume the existence of a single homogeneous field from the rest of the effective medium in its entirety, and that its interaction with an individual particle can then be used to find the field and describe the effective medium.

It should now be noted that the requirement for periodicity can actually be relaxed somewhat. It proves sufficient that "periodic" really means that the microstructure varies only on the macroscopic scale. Still, though, this clearly puts some serious restrictions on the systems that can be modelled. Under the framework as it stands, randomly oriented and distributed particles, as in [32], are not possible to model with homogenisation (although as will be considered later, drawing on work from [7], we can model disordered particles in the low volume fraction limit). On the other hand, provided that the periodicity requirements are met, the modelling of an arbitrarily complex microstructure (that cannot be dealt with using the traditional effective medium techniques) should be possible. This will, however, likely require the numerical solution of the cell problem. Moreover, homogenisation is by nature algebraically intensive, requiring multiple systems of differential equations to be solved and, in this respect, Maxwell-Garnett and Bruggeman are advantageous for more simple systems in that they offer far easier calculation.

There is a benefit to this rigorous framework, though, in that it provides a systematic method to solve far more complex problems. In the setting of doped liquid crystal systems, we must consider inclusions in an anisotropic medium. The original effective medium theories have been extended to deal with such problems, with anisotropic, spherical particles in an anisotropic medium considered in [52]. Reference [23] considers the solution for the electric field inside an ellipsoidal particle in an anisotropic medium, providing the basis for relevant effective medium theories. From the point of view of homogenisation, nothing changes, the homogenised equations remain the same, though they do become harder to solve since the scalar permittivities become tensors. To model physically realistic systems, the variable liquid crystal and dopant alignments must also be considered. This is, again, a problem that has previously been considered in the context of effective medium theories. In [50], for instance, the authors consider ferroelectric particles in liquid crystal, with the dopant

particles assumed to be aligned parallel to the local director field. It is found that simply using the derived effective permittivity in the usual alignment equations leads to troubles, and further work is required to use the effective permittivity correctly. More general particle/liquid crystal interactions have been considered, as in [33, 8] for instance, which both formulate an effective free energy to derive the governing equations of the effective medium. In [33] this interaction occurs between the volume average of the orientations, reducing to considering order parameters for the liquid crystal and for the particles. The model presented in [8] is more complex, directly considering surface interactions and changes in concentration, for example. These theories, however, define parameters with unknown values, and these must somehow be found, e.g., experimentally. Here lies another advantage of homogenisation for such systems of complex interactions. Treating the entire problem in the homogenised regime [3], the solution yields an effective medium with both an effective permittivity and alignment, naturally avoiding the complications in [50]. It also allows for a clear way in which to couple the liquid crystal and platelet alignment through a weak anchoring condition, as in [8], but without the need for defining and finding new parameters. All macroscopic properties and behaviour arise directly from the original equations and properties of the microstructure. It does, however, necessitate the solution of a second cell problem.

Given the power of the method, and how it can be used to account for complex geometries and interactions, we continue in this work using the method of homogenisation. Starting with the simplified problem of a dielectric particle in an isotropic medium, the homogenised potential equations are derived and the properties of the effective permittivity tensor explored. This work is then built upon to derive an analytical expression for the effective permittivity of ellipsoidal, dielectric particles in an anisotropic medium, with a view to coupling this to the homogenised liquid crystal alignment equations for future work. The model is derived for a disordered configuration of particles, that are allowed to vary in shape and orientation on the macroscopic scale, which is not accounted for in the traditional effective medium theories.

4.3 Dielectric Particles in Isotropic Medium

4.3.1 Setting Up the Problem

We consider a system comprising of a voltage applied across two parallel plates with dielectric particles suspended in an isotropic medium between them. The fields are static and have no time dependence, so, from equation (2.1c), we have that $\boldsymbol{\nabla} \times \boldsymbol{E} = 0$ and there exists a scalar potential ϕ such that $\boldsymbol{E} = -\boldsymbol{\nabla}\phi$. This greatly simplifies the

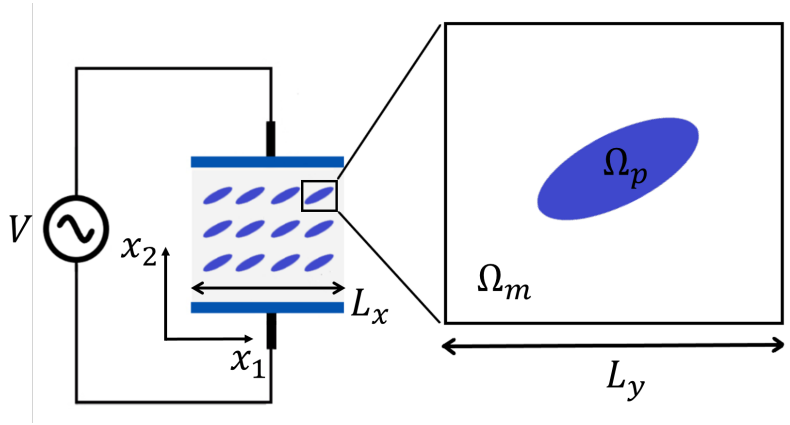


FIGURE 4.1: Periodic approximation of the microscopic geometry of the doped system.

treatment of the problem, which is now to find this electric potential inside the system. The first Maxwell equation (2.1a) then becomes

$$\nabla \cdot (\varepsilon_m \nabla \phi) = 0, \qquad \boldsymbol{x} \in \Omega_m, \tag{4.3.1.0.1}$$

$$\nabla \cdot (\varepsilon_p \nabla \phi) = 0, \qquad \boldsymbol{x} \in \Omega_p, \tag{4.3.1.0.2}$$

where ε_m and ε_p are the dielectric permittivities of the medium and particles respectively, and Ω_m and Ω_p are the domains of the medium and particles. The problem is closed by adding the boundary conditions that the normal component of $\boldsymbol{D}$ must be continuous at the boundary between medium and particle and fixing the voltage drop from one plate to another:

$$\hat{\boldsymbol{n}} \cdot (\varepsilon_m \nabla \phi) = \hat{\boldsymbol{n}} \cdot (\varepsilon_p \nabla \phi), \qquad \boldsymbol{x} \in \partial\Omega_p, \tag{4.3.1.0.3}$$

$$\phi(0) = 0\,,\ \phi(d) = V, \tag{4.3.1.0.4}$$

where $\hat{\boldsymbol{n}}$ is the normal to $\partial\Omega_p$, the boundary of the particle.

4.3.2 The Homogenised Equations

As discussed in the previous section, the method of homogenisation requires only the assumption that our system is locally periodic. This means that, when viewed at the microscopic level, any part of the medium will look like a tiled array of one microscopic cell, with the properties of the averaged, macroscopic solution depending on the geometry and properties of this cell. Assuming that the particles of our composite system lie on a square lattice, the obvious choice for our unit cell is just one particle in the centre of a square tile, as shown in figure 4.1.

With both the macroscopic and microscopic geometries now defined, the new parameter

$$\eta = \frac{L_y}{L_x} \ll 1 \tag{4.3.2.0.1}$$

can also defined, measuring the difference in the scale of the two geometries. We can use this parameter to define the new spatial variable

$$\boldsymbol{y} = \frac{1}{\eta}\boldsymbol{x}, \tag{4.3.2.0.2}$$

and we assume that the potential depends explicitly on both $\boldsymbol{x}$ and $\boldsymbol{y}$. The dependence on the slow varying $\boldsymbol{x}$ describes the macroscopic dynamics of the problem, while the dependence on the fast varying $\boldsymbol{y}$ describes the microscopic dynamics. Introducing this $\boldsymbol{y}$ dependence means that the partial derivatives in our governing equations (4.3.1.0.1) and (4.3.1.0.2) (which are derivatives with respect to $\boldsymbol{x}$) change according to

$$\boldsymbol{\nabla} \to \boldsymbol{\nabla}_{\boldsymbol{x}} + \frac{1}{\eta}\boldsymbol{\nabla}_{\boldsymbol{y}}. \tag{4.3.2.0.3}$$

The end goal in homogenisation is to find a solution that is purely macroscopic, with no dependence on the $\boldsymbol{y}$ scale. To that end, we search for a solution of the form

$$\phi = \phi_0\left(\boldsymbol{x},\boldsymbol{y}\right) + \eta\phi_1\left(\boldsymbol{x},\boldsymbol{y}\right) + \eta^2\phi_2\left(\boldsymbol{x},\boldsymbol{y}\right) + O\left(\eta^3\right), \tag{4.3.2.0.4}$$

where ϕ_0 is the macroscopic potential and the remaining terms are small corrections due to the microscopic geometry. Substituting this into the governing equations (4.3.1.0.1)-(4.3.1.0.4) then gives

$$\left(\boldsymbol{\nabla}_{\boldsymbol{x}} + \frac{1}{\eta}\boldsymbol{\nabla}_{\boldsymbol{y}}\right)\cdot\left[\varepsilon_\alpha\left(\boldsymbol{\nabla}_{\boldsymbol{x}} + \frac{1}{\eta}\boldsymbol{\nabla}_{\boldsymbol{y}}\right)\left(\phi_0 + \eta\phi_1 + \eta^2\phi_2 + O\left(\eta^3\right)\right)\right] = 0\,, \quad \boldsymbol{y} \in \Omega_\alpha, \tag{4.3.2.0.5}$$

where $\alpha = m,\ p$, and

$$\left[\left(\varepsilon_m - \varepsilon_p\right)\left(\boldsymbol{\nabla}_{\boldsymbol{x}} + \frac{1}{\eta}\boldsymbol{\nabla}_{\boldsymbol{y}}\right)\left(\phi_0 + \eta\phi_1 + \eta^2\phi_2 + O\left(\eta^3\right)\right)\right]\cdot\hat{\boldsymbol{n}} = 0\,, \qquad \boldsymbol{y} \in \partial\Omega_p. \tag{4.3.2.0.6}$$

Consider now a single unit cell, as in figure 4.1, and the series of problems obtained by equating powers of η in equations (4.3.2.0.5) and (4.3.2.0.6). The leading order problem, given by the $O\left(\eta^{-2}\right)$ terms, is

$$\boldsymbol{\nabla}_{\boldsymbol{y}}\cdot\left(\varepsilon_\alpha\boldsymbol{\nabla}_{\boldsymbol{y}}\phi_0\right) = 0, \qquad \boldsymbol{y} \in \Omega_\alpha, \tag{4.3.2.0.7}$$

$$\hat{\boldsymbol{n}}\cdot\left(\varepsilon_m\boldsymbol{\nabla}_{\boldsymbol{y}}\phi_0\right) = \hat{\boldsymbol{n}}\cdot\left(\varepsilon_p\boldsymbol{\nabla}_{\boldsymbol{y}}\phi_0\right), \qquad \boldsymbol{y} \in \partial\Omega_p. \tag{4.3.2.0.8}$$

To check the solvability of the problem, we must check that these equations are consistent with one another. To do so, we integrate equation (4.3.2.0.7) over Ω_α and

check that the boundary conditions are compatible as follows:

$$\begin{aligned} 0 = \int_{\Omega_\alpha} \boldsymbol{\nabla}_{\boldsymbol{y}} \cdot \left(\varepsilon_\alpha \boldsymbol{\nabla}_{\boldsymbol{y}} \phi_0\right) dV &= \int_{\partial\Omega_\alpha} \left(\varepsilon_\alpha \boldsymbol{\nabla}_{\boldsymbol{y}} \phi_0\right) \cdot \hat{\boldsymbol{n}} dS \\ &= -\int_{\partial\Omega_{\bar{\alpha}}} \left(\varepsilon_{\bar{\alpha}} \boldsymbol{\nabla}_{\boldsymbol{y}} \phi_0\right) \cdot \hat{\boldsymbol{n}} dS \\ &= -\int_{\Omega_{\bar{\alpha}}} \boldsymbol{\nabla}_{\boldsymbol{y}} \cdot \left(\varepsilon_{\bar{\alpha}} \boldsymbol{\nabla}_{\boldsymbol{y}} \phi_0\right) dV \\ &= 0, \end{aligned} \tag{4.3.2.0.9}$$

as required. $\Omega_{\bar{\alpha}}$ denotes the complement of Ω_α and the result follows from the divergence theorem and the leading order boundary conditions (4.3.2.0.8) to switch between domains. This then tells us that the problem is solvable with no added restrictions on ϕ_0 other than the given equations. The solution at leading order is then

$$\phi_0 = \phi_0 \left(\boldsymbol{x}\right), \tag{4.3.2.0.10}$$

which is a function of the macroscopic position only.

Proceeding to the next order, $O\left(\eta^{-1}\right)$, in equations (4.3.2.0.5) and (4.3.2.0.6) gives

$$\boldsymbol{\nabla}_{\boldsymbol{y}} \cdot \left(\varepsilon_\alpha \boldsymbol{\nabla}_{\boldsymbol{x}} \phi_0 + \varepsilon_\alpha \boldsymbol{\nabla}_{\boldsymbol{y}} \phi_1\right) = 0 \qquad \boldsymbol{y} \in \Omega_\alpha \tag{4.3.2.0.11}$$

$$\hat{\boldsymbol{n}} \cdot \left[\varepsilon_m \left(\boldsymbol{\nabla}_{\boldsymbol{x}} \phi_0 + \boldsymbol{\nabla}_{\boldsymbol{y}} \phi_1\right)\right] = \varepsilon_p \left(\boldsymbol{\nabla}_{\boldsymbol{x}} \phi_0 + \boldsymbol{\nabla}_{\boldsymbol{y}} \phi_1\right) \qquad \boldsymbol{y} \in \partial\Omega_p. \tag{4.3.2.0.12}$$

Checking for solvability in the same way as before yields no extra constraints on ϕ_1. Rearranging equation (4.3.2.0.11) gives

$$\boldsymbol{\nabla}_{\boldsymbol{y}} \cdot \left(\varepsilon_\alpha \boldsymbol{\nabla}_{\boldsymbol{y}} \phi_1\right) = -\left(\boldsymbol{\nabla}_{\boldsymbol{y}} \varepsilon_\alpha\right) \cdot \left(\boldsymbol{\nabla}_{\boldsymbol{x}} \phi_0\right)$$

and so we make the ansatz that ϕ_1 has the form

$$\phi_1 = \boldsymbol{\Gamma}_\alpha \left(\boldsymbol{y}\right) \cdot \boldsymbol{\nabla}_{\boldsymbol{x}} \phi_0 \tag{4.3.2.0.13}$$

for some vector field $\boldsymbol{\Gamma}_\alpha$ in each of the domains Ω_α. If this expression for ϕ_1 is now substituted into equations (4.3.2.0.11) and (4.3.2.0.12) then the following cell problem is obtained:

$$\partial y_i \left(\varepsilon_\alpha\right)_{ij} \left[\partial y_j (\Gamma_\alpha)_k + \delta_{jk}\right] = 0, \qquad \boldsymbol{y} \in \Omega_\alpha, \tag{4.3.2.0.14a}$$

$$\hat{n}_i (\varepsilon_m)_{ij} \left[\partial y_j (\Gamma_m)_k + \delta_{jk}\right] = \varepsilon_p \hat{n}_j \left[\partial y_j (\Gamma_p)_k + \delta_{jk}\right], \qquad \boldsymbol{y} \in \partial\Omega_p, \tag{4.3.2.0.14b}$$

$$\boldsymbol{\Gamma}_m = \boldsymbol{\Gamma}_p, \qquad \boldsymbol{y} \in \partial\Omega_p, \tag{4.3.2.0.14c}$$

$$\boldsymbol{\Gamma}_m \text{ periodic on opposite sides of cell,} \tag{4.3.2.0.14d}$$

This problem is dependent only on the microscopic variable $\boldsymbol{y}$ and the unit cell can be solved to find the vector fields $\boldsymbol{\Gamma}_\alpha$. This reduces all of the $\phi_1 \left(\boldsymbol{x}, \boldsymbol{y}\right)$ dependence of the

problem to a dependence on $\phi_0(\boldsymbol{x})$. As a result, we are part way to obtaining a purely macroscopic problem, but must still proceed to the next order.

At $O(\eta^0)$ we have that

$$\boldsymbol{\nabla}_{\boldsymbol{x}} \cdot \left[\varepsilon_\alpha \left(\boldsymbol{\nabla}_{\boldsymbol{x}}\phi_0 + \boldsymbol{\nabla}_{\boldsymbol{y}}\phi_1\right)\right] + \boldsymbol{\nabla}_{\boldsymbol{y}} \cdot \left[\varepsilon_\alpha \left(\boldsymbol{\nabla}_{\boldsymbol{x}}\phi_1 + \boldsymbol{\nabla}_{\boldsymbol{y}}\phi_2\right)\right] = 0, \qquad \boldsymbol{y} \in \Omega_\alpha \quad (4.3.2.0.15)$$

$$\hat{\boldsymbol{n}} \cdot \left[\varepsilon_m \left(\boldsymbol{\nabla}_{\boldsymbol{x}}\phi_1 + \boldsymbol{\nabla}_{\boldsymbol{y}}\phi_2\right)\right] = \varepsilon_p \hat{\boldsymbol{n}} \cdot \left(\boldsymbol{\nabla}_{\boldsymbol{x}}\phi_1 + \boldsymbol{\nabla}_{\boldsymbol{y}}\phi_2\right), \qquad \boldsymbol{y} \in \partial\Omega_p. \quad (4.3.2.0.16)$$

We can substitute in 4.3.2.0.13 to obtain

$$\partial x_i (\varepsilon_\alpha)_{ij} \left[\delta_{jk} + \partial y_j (\Gamma_\alpha)_k\right] \partial x_k \phi_0 + \partial y_i (\varepsilon_\alpha)_{ij} \left[\partial x_j \phi_1 + \partial y_j \phi_2\right] = 0, \qquad \boldsymbol{y} \in \Omega_\alpha \quad (4.3.2.0.17)$$

$$\hat{n}_i (\varepsilon_m)_{ij} \left[\partial x_j \phi_1 + \partial y_j \phi_2\right] = \varepsilon_p \hat{n}_i \left[\partial x_i \phi_1 + \partial y_i \phi_2\right], \qquad \boldsymbol{y} \in \partial\Omega_p. \quad (4.3.2.0.18)$$

and we can then integrate the first equation over the microscopic domain, apply the divergence theorem and the the boundary condition, and then undo the divergence theorem to give

$$\partial x_i \int_{\Omega_\alpha} (\varepsilon_\alpha)_{ij} \left[\delta_{jk} + \partial y_j (\Gamma_\alpha)_k\right] dV \partial x_k \phi_0 - \int_{\Omega_{\overline{\alpha}}} \partial y_i (\varepsilon_{\overline{\alpha}})_{ij} \left[\partial x_j \phi_1 + \partial y_j \phi_2\right] dV = 0. \quad (4.3.2.0.19)$$

If we add to this the integral of the $\overline{\alpha}$ equations, the second terms cancels and we find

$$\partial x_i \sum_\alpha \int_{\Omega_\alpha} (\varepsilon_\alpha)_{ij} \left[\delta_{jk} + \partial y_j (\Gamma_\alpha)_k\right] dV \partial x_k \phi_0 = 0. \quad (4.3.2.0.20)$$

Defining the effective permittivity tensor

$$(\varepsilon_{\text{eff}})_{ij} = \sum_\alpha \int_{\Omega_\alpha} (\varepsilon_\alpha)_{ik} \left[\partial_{kj} + \partial_k (\Gamma_\alpha)_j\right] dV \quad (4.3.2.0.21)$$

the complete homogenised problem is then

$$\boldsymbol{\nabla}_{\boldsymbol{x}} \cdot (\varepsilon_{\text{eff}} \cdot \boldsymbol{\nabla}_{\boldsymbol{x}} \phi_0) = 0, \quad (4.3.2.0.22a)$$

$$\phi_0(d) = V, \quad (4.3.2.0.22b)$$

$$\phi_0(0) = 0. \quad (4.3.2.0.22c)$$

The final homogenised equations depend only on $\boldsymbol{x}$ and all of the effects of the microscopic geometry have become a material property contained in ε_{eff}, calculated independently from solution of the cell problem. Unlike the scalar ε_m and ε_p, ε_{eff} may be a tensor, reflecting the symmetries of the particle geometries. This results in the "averaged out" composite material (made from two isotropic materials) behaving as a single, anisotropic material. The emergence of this anisotropy is a clear consequence of the broken symmetries introduced by the addition of the particles. Considering the

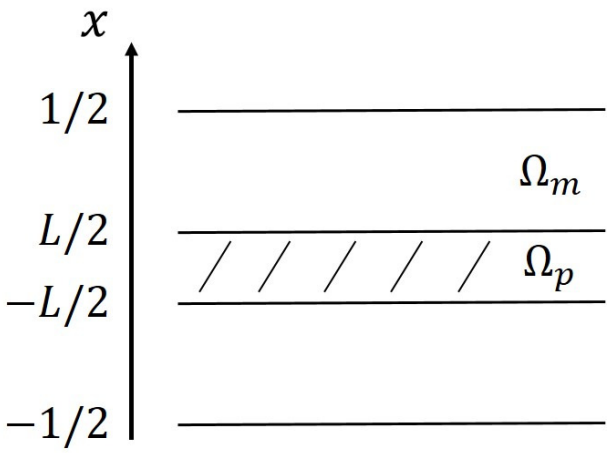


FIGURE 4.2: The unit cell of the 1D problem.

average permittivity when travelling a straight path through the system, this will be different depending on which direction the path takes due to the fraction of the path that passes through each domain. This idea is similar to that of a series of capacitors, where the capacitance (and hence permittivity) scales with the thickness of the dielectric.

4.3.3 1D Simplification

To validate this capacitance idea, we consider now the case where the particles are in fact flat layers that stretch across the entire cross section of the cell. The problem is then only one dimensional and the system is simply a stack of unit cells as shown in figure 4.2. The cell problem (4.3.2.0.14a-4.3.2.0.14d) reduces to

$$\frac{d^2\Gamma_\alpha}{dx^2} = 0, \qquad x \in \Omega_\alpha, \tag{4.3.3.0.1a}$$

$$\varepsilon_m\left(\frac{d\Gamma_m}{dx}+1\right) = \varepsilon_p\left(\frac{d\Gamma_p}{dx}+1\right), \qquad x = \pm\frac{1}{2}, \tag{4.3.3.0.1b}$$

$$\Gamma_m\left(\pm\frac{L}{2}\right) = \Gamma_p\left(\pm\frac{L}{2}\right), \tag{4.3.3.0.1c}$$

$$\Gamma_m\left(\frac{1}{2}\right) = \Gamma_m\left(-\frac{1}{2}\right), \tag{4.3.3.0.1d}$$

where $0 < L < 1$ is the thickness of the particles and the unit cell thickness is 1. Equation (4.3.3.0.1a) tells us that $\Gamma_\alpha = A_\alpha x + B_\alpha$ and, since the periodicity allows us to choose an integration constant, we set $B_p = 0$. Together with the continuity and periodicity conditions this gives

$$\Gamma_p = A_p x_1, \tag{4.3.3.0.2}$$

$$\Gamma_m^+ = \frac{\left(\frac{1}{2} - x_1\right) A_p L}{1-L}, \tag{4.3.3.0.3}$$

and

$$\Gamma_m^- \frac{\left(x_1 - \frac{1}{2}\right) A_p L}{1 - L}, \tag{4.3.3.0.4}$$

where Γ_m^+ and Γ_m^- are the parts of Γ_m above and below the "particle". Substituting these into the boundary conditions on the derivative gives

$$A_p = \frac{\left(\varepsilon_p - \varepsilon_m\right)(L-1)}{\varepsilon_m L - \varepsilon_p (L-1)}. \tag{4.3.3.0.5}$$

With this expression for A_p, ε_{eff} can then be calculated according to equation (4.3.2.0.21), giving the final homogenised equation

$$\underbrace{\frac{\varepsilon_m \varepsilon_p}{\varepsilon_m L + \varepsilon_p (1-L)}}_{=\varepsilon_{\text{eff}}} \frac{d^2\phi_0}{dx^2} = 0. \tag{4.3.3.0.6}$$

This seems a valid and reasonable solution. For a capacitor made up of a single material with a constant dielectric constant, the potential will be linear in x, just as we have here. It can also be seen that in the limits $L \to 1$ and $L \to 0$, the effective permittivity goes to ε_p and ε_m, in agreement with the geometry. Furthermore, calculating the total capacitance of the system as capacitors in series, it can be seen that

$$\begin{aligned} C_{\text{tot}} &= \frac{1}{\frac{1}{c_1} + \frac{1}{c_2} + \cdots} \\ &= \frac{1}{N\left(\frac{L}{\varepsilon_p A} + \frac{1-L}{\varepsilon_m A}\right)} \end{aligned} \tag{4.3.3.0.7}$$

for a system of N particles and plates of area A. Each unit cell has unit length, and so it follows that $N = d$, the thickness of the whole system, and this can be written as

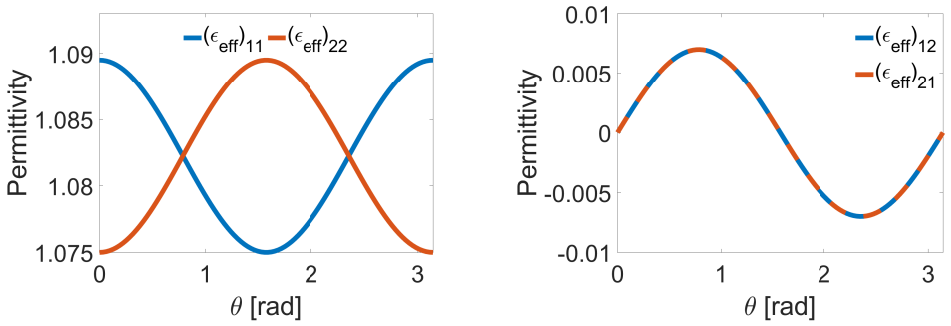
$$\begin{aligned} C_{\text{tot}} &= \frac{\varepsilon_m \varepsilon_p}{\varepsilon_m L + \varepsilon_p (1-L)} \frac{A}{d} \\ &= \varepsilon_{\text{eff}} \frac{A}{d}, \end{aligned}$$

which is the equation for capacitance in the presence of a uniform dielectric with permittivity ε_{eff}. The homogenised solution, which approximates the suspension as a single material, is therefore in exact agreement with standard results.

4.3.4 2D Particles in Medium - Homogenisation vs Finite Elements

4.3.4.1 The Effective Permittivity

Given that homogenisation has proven itself a good choice of method in one dimension, we consider now the 2D case, where the cell problem

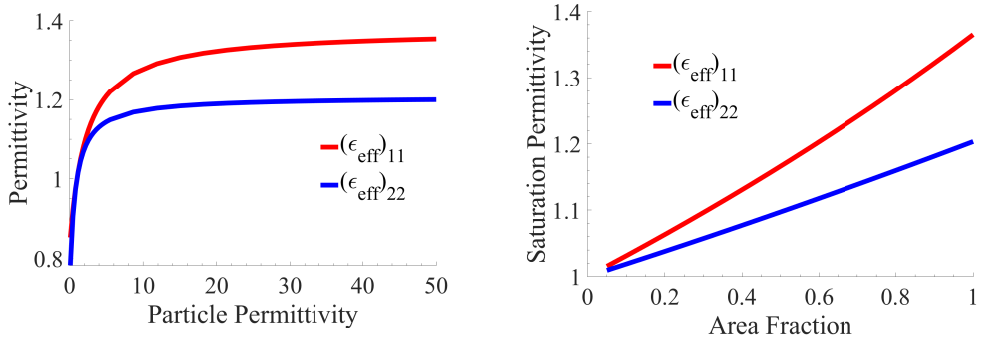


(A) (B)

FIGURE 4.3: The components of ε_{eff} as a function of particle angle for ellipsoidal particles with major and minor semiaxes of 0.25 and 0.15 and effective permittivity 2 suspended in an isotropic medium of permittivity 1.

(4.3.2.0.14a-4.3.2.0.14d) is solved using the finite elements package Comsol. The particles are modelled to be ellipses, with major and minor semiaxes of 0.25 and 0.15 (relative to a square cell of side length 1), and with a permittivity of 2. The permittivity of the isotropic medium is 1. Parameterising the particle orientation by θ, the angle between the particle's major axis and the x axis, the resulting effective permittivity is shown in figure 4.3. The effect of the particles, whose permittivity is higher than that of the isotropic medium, is to raise the overall permittivity of the system. It can also be observed that the orientation of the particles plays a clear roll in the effective permittivity, with the modulation being π periodic, as you would expect given the symmetries of the particles. It can be observed that the direction parallel to the orientation of the particles gains a greater increase in permittivity than the orthogonal direction. The converse effect can be observed when the permittivity of the particles is lower than that of the medium. Both of these results agree in analogy with the 1D example. Considering a straight path through the system and the relative fraction of the path that is particle, ε_{eff} along this path would have a form exactly the same as that in equation (4.3.3.0.6). By making the fraction of the path that is in the particles larger (making L larger), ε_{eff} goes to ε_p and the permittivity increases (or decreases when $\varepsilon_p < \varepsilon_m$).

Intuitively, you would expect that the size of the modulation in the effective permittivity will be dependent on the relative difference in permittivities of the medium and particles. This can clearly be observed in figure 4.4a, where the particle's size is as before, the orientation is fixed at $\theta = 0$, and $\varepsilon_m = 1$. The intersection of the xx and yy permittivities occurs at 1 and coincides with the system being isotropic. Interestingly, the effective permittivity saturates and approaches an upper limit asymptotically. This limiting is due to the size of the particles. By considering the form of the effective permittivity in equation (4.3.3.0.6), it can be seen that when $\varepsilon_p \gg \varepsilon_m$ then ε_{eff} goes to $\frac{\varepsilon_m}{1-L}$. Taking L to be the size of the particle in the x and y directions, the permittivity tends to 1.33 and 1.18 (for $\theta = 0$) as can be seen in figure (4.4a).



(A) (B)

FIGURE 4.4: (a) Effective permittivity as a function of particle permittivity. $\varepsilon_m = 1$, particle major and minor semiaxes 0.25 and 0.15, and $\theta = 0$. (b) Saturation permittivity ($\varepsilon_p = 100$) as a function of area fraction, with $\varepsilon_m = 1$ and particle aspect ratio set at 0.25:0.15.

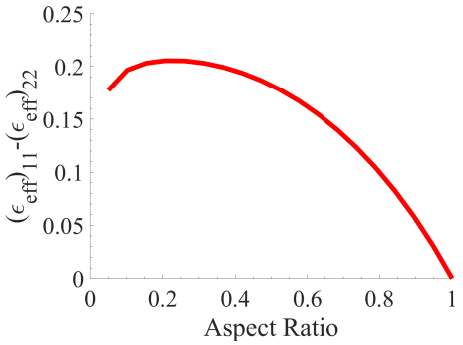
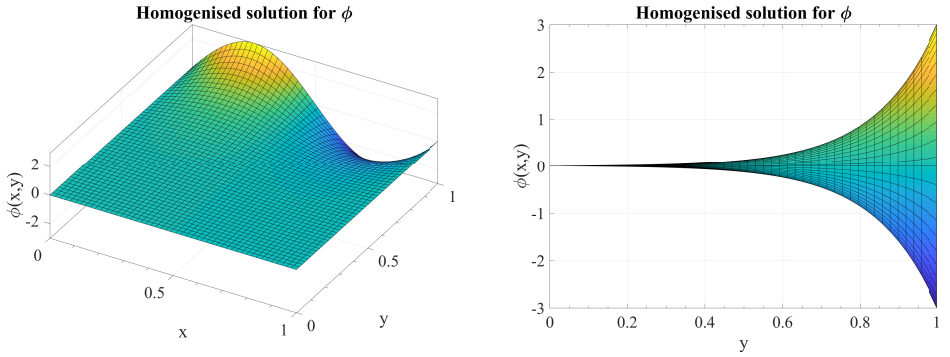


FIGURE 4.5: Anisotropy in ε_{eff} as a function of particle aspect ratio with the particle major semiaxis at 0.25. $\varepsilon_m = 1$, $\varepsilon_p = 100$, $\theta = 0$

Moreover, maintaining the aspect ratio of the semiaxes, 0.25:0.15, and scaling the area of the particle, it can be observed that this limit scales linearly with particle size, as shown in figure 4.4b. It can also be seen that the relative anisotropy in the x and y directions depends on the xy aspect ratio of the particles. This can be seen in figure 4.5 showing the saturation permittivity when the major semiaxis is 0.25. Note that the dip as the aspect ratio goes to zero is due to the particle area going to zero and not resembling a physically realistic particle.

4.3.4.2 Accuracy of the Homogenised Solution

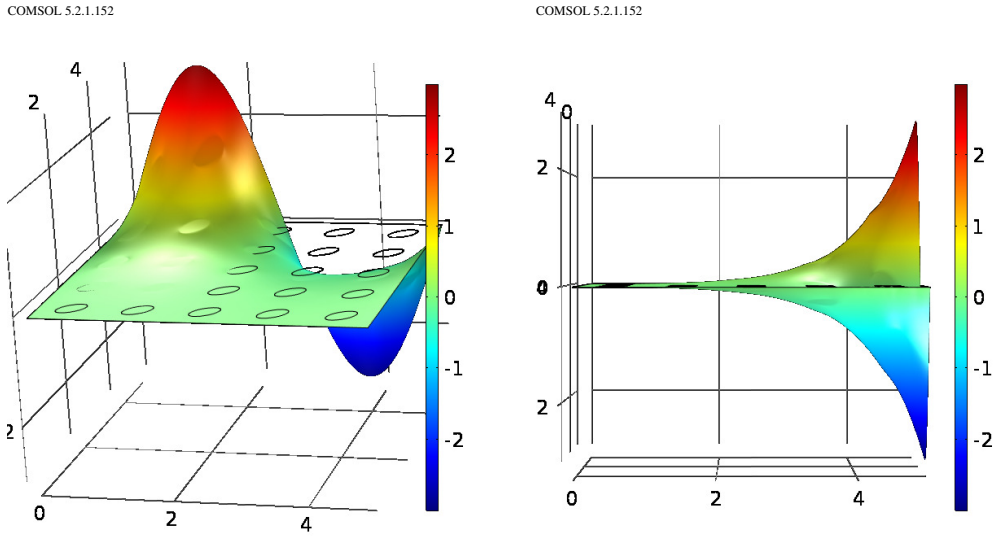
With ε_{eff} in hand, we can now solve the macroscopic problem given in equations (4.3.2.0.22b)-(4.3.2.0.22c). Using spectral differentiation [55], this problem can be solved very efficiently in MATLAB. Applying a sinusoidal modulation of the voltage at the top plate, the homogenised solution can be seen in figure 4.6. To see if the accuracy is as good as in the 1D case, the homogenised solution is compared to that from direct finite elements calculation including the full microscopic geometry.



(A) (B)

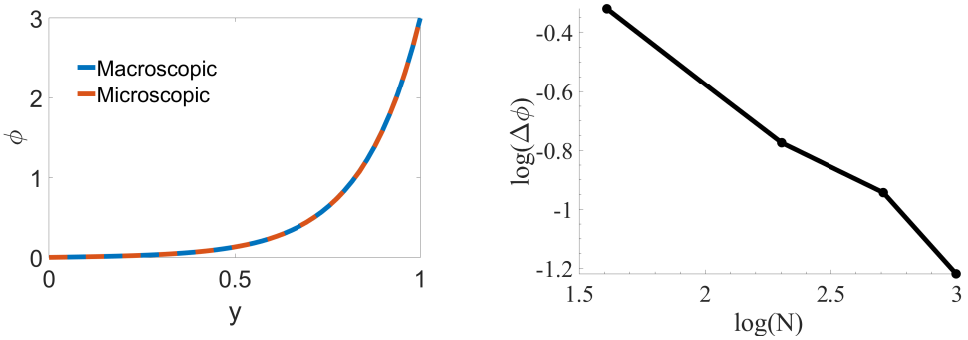
FIGURE 4.6: The homogenised potential subject to a sinusoidal voltage modulation $\phi(x,1) = 2\sin(2\pi x)$ at the top plate. Calculated for particles at 60° with major and minor semiaxes equal to 0.25 and 0.15, $\varepsilon_p = 2$, and $\varepsilon_m = 1$. (a) Front view (b) Side (yz) view.

We once again use Comsol to solve the problem. The full microscopic solution for a 5 by 5 array of particles can be seen in figure 4.7. Comparing the error between the two solutions, $\Delta\phi$, as a function of grid size, as shown in figure (4.8b), it can be seen that modelling the full microscopic geometry does not differ from the homogenised solution all that greatly. This is even where the number of (rather large) particles is small, which does not meet the requirements for the homogenisation approximation particularly well. It does, however, take far longer to compute. On a 2.4GHz Intel i5 processor, fully computing both the effective permittivity and the homogenised solution takes no more than 10 seconds of real time. Computing the microscopic solution for a 20 by 20 array of particles takes a few minutes. Considering that a real system could have hundreds or even thousands of particles, the homogenised solution will not only be incredibly accurate, but also take only a tiny fraction of the time to compute, thus proving itself an obvious method of choice in such problems. The difference in solutions can be best seen from the side view. There are slight bumps in the potential caused by the discontinuity of the potential derivative, and these reveal the presence of the particles. In the 1D case, the homogenised solution gives a linear potential that averages the piecewise linear solution of the microscopic solution. In the same way, you would expect that as the 2D array increases in size and the particles become smaller, the solution will smooth out and tend towards the homogenised solution. It can be seen, in fact, that the two solutions have excellent agreement with as little as a 20 by 20 array of particles, as shown in figure 4.8a.



(A) (B)

FIGURE 4.7: The microscopic potential, as calculated by Comsol, subject to a sinusoidal voltage modulation $\phi(x,1) = 3\sin(2\pi x)$ at the top plate. Calculated for a 5x5 array of particles at 60° with major and minor semi-axes equal to 0.25 and 0.15, $\varepsilon_p = 2$, and$\varepsilon_m = 1$. (a) Front view (b) Side (yz) view.



(A) (B)

FIGURE 4.8: (a) Comparison of the macroscopic and microscopic potential subject to a sinusoidal voltage modulation $\phi(x,1) = 3\sin(2\pi x)$ at the top plate. Calculated for a microscopic geometry of a 20 by 20 array of particles at 60° with major and minor semi-axes equal to 0.25 and 0.15, $\varepsilon_p = 2$, and $\varepsilon_m = 1$. (b) Difference in the macroscopic and microscopic solutions as a function of the microscopic array size, calculated with the infinity norm applied to the entire system.

4.4 Homogenised Electrostatic Problem for Disordered, Spatially Varying Ellipsoidal Particles

It is clear that homogenisation provides an accurate method with which to efficiently describe these composite systems of dielectric particles in liquid crystal. We have, however, put some serious restrictions on these systems with the key modelling assumption in homogenisation, that the particles can be viewed as a period array of identical particles. Fortunately, the method can be adjusted to take this into account. In [7], for instance, the authors demonstrate that in the low particle volume fraction limit, diffusion in a disordered porous medium is equivalent to the standard homogenised lattice setup. Given the similarity in the governing equations of the diffusion and electrostatic problems, and also what we know from the previous section, it seems reasonable to expect that the same methodology can be applied to our dielectric composite problem.

At small volume fractions we have small particles that are, on average, well spaced out (in the real, disordered problem). Each particles will give rise to a small scattered field and, except for the few cases of particularly close particles, each field will decay exponentially. This means that the scattered field from the majority of particles will not interact with any other particles, and there will mostly be no inter-particle coupling. In the homogenised regime, the particles will each occupy a small volume fraction of their unit cell, and the solution to the cell problem will also be small and exponentially decaying. Since the particles are located at the centre of each unit cell, the solution to the cell problem will decay completely by the time it reaches the cell boundary, and there will also be no inter-particle coupling (which arises through the periodic boundary conditions of the cell problem). It is the lack of coupling in both cases that means the average, macroscopic behaviour of both the real, disordered system and the ordered, homogenised system should be approximately the same. As mentioned, however, there will be particles in the real, disordered system that do lie close to one another and experience coupling effects in their scattered fields. The following work will therefore derive the low volume fraction homogenised solution, and compare the results with numerical simulations of disordered systems to see to what extend homogenisation can correctly describe the average macroscopic behaviour of disordered systems.

This work will also consider another methodology presented in [7], one in which the geometry of the particles is allowed to vary on the macroscopic scale. This is achieved by defining the particle normals to be a function of their macroscopic position, leading to a multiple scales expression for the normal (which was previously only microscopic). This presents no problems for the implementation of the homogenisation method, which is valid provided that the properties of the microscale vary only on a much larger scale than that of the unit cells.

4.4.1 Varying Particle Geometries

To begin, we must first define the geometries of the problem. We already know that the outer boundary of the system is described by the parallel plate boundaries at $x_2 = 0,\ d$, and we assume the system to be infinitely extending in the x_1 direction. For the microscopic geometry, we will first allow the particles to have varying shape and size by letting their semi-axes, a and b, vary with macroscopic position, assuming that $a \geq b$. The particle boundaries then satisfy the ellipse equation

$$\left(\frac{y_1}{a\left(\boldsymbol{x}\right)}\right)^2 + \left(\frac{y_2}{b\left(\boldsymbol{x}\right)}\right)^2 = 1, \tag{4.4.1.0.1}$$

which can alternatively be written as

$$\left| \begin{pmatrix} a\left(\boldsymbol{x}\right)^{-1} & 0 \\ 0 & b\left(\boldsymbol{x}\right)^{-1} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right|^2 = 1. \tag{4.4.1.0.2}$$

We also want the orientation of the particles to be able to vary. Describing their orientation by the angle θ that their major axes make with the positive x_1 axis, the particle boundaries are then the points that satisfy the ellipse equation after the plane has been rotated by an angle $-\theta$, i.e.,

$$\left| \begin{pmatrix} a\left(\boldsymbol{x}\right)^{-1} & 0 \\ 0 & b\left(\boldsymbol{x}\right)^{-1} \end{pmatrix} \begin{pmatrix} \cos\theta\left(\boldsymbol{x}\right) & \sin\theta\left(\boldsymbol{x}\right) \\ -\sin\theta\left(\boldsymbol{x}\right) & \cos\left(\boldsymbol{x}\right) \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right|^2 = 1. \tag{4.4.1.0.3}$$

This can equivalently be written as

$$\left|\mathrm{E}\boldsymbol{y}\right|^2 = \left(\mathrm{E}\boldsymbol{y}\right)^T \mathrm{E}\boldsymbol{y} = \boldsymbol{y}^T \mathrm{Q}\boldsymbol{y} = 1, \tag{4.4.1.0.4}$$

where

$$\mathrm{E}\left(\boldsymbol{x}\right) = \begin{pmatrix} a\left(\boldsymbol{x}\right)^{-1}\cos\theta\left(\boldsymbol{x}\right) & a\left(\boldsymbol{x}\right)^{-1}\sin\theta\left(\boldsymbol{x}\right) \\ -b\left(\boldsymbol{x}\right)^{-1}\sin\theta\left(\boldsymbol{x}\right) & b\left(\boldsymbol{x}\right)^{-1}\cos\theta\left(\boldsymbol{x}\right) \end{pmatrix} \tag{4.4.1.0.5}$$

and

$$\mathrm{Q} = \mathrm{E}^T\mathrm{E}. \tag{4.4.1.0.6}$$

This allows us to simply define the particle boundaries with the level set

$$\chi\left(\boldsymbol{x},\boldsymbol{y}\right) = \boldsymbol{y}^T\mathrm{Q}\boldsymbol{y} - 1, \tag{4.4.1.0.7}$$

which gives the particles boundaries at $\chi = 0$. By definition of the gradient, the surface normals are parallel to the gradient of this level set, which in the homogenised regime possesses two components:

$$\boldsymbol{n}\left(\boldsymbol{x},\boldsymbol{y}\right) = \left(\frac{1}{\eta}\boldsymbol{\nabla}_y + \boldsymbol{\nabla}_x\right)\chi\left(\boldsymbol{x},\boldsymbol{y}\right) \tag{4.4.1.0.8}$$

$$= \frac{1}{\eta}\boldsymbol{n}_0\left(\boldsymbol{x},\boldsymbol{y}\right) + \boldsymbol{n}_1\left(\boldsymbol{x},\boldsymbol{y}\right). \tag{4.4.1.0.9}$$

Note that we do not need these to be unit normals, this is a result of the boundary conditions having an $\boldsymbol{n}$ on each side and any scaling factors cancelling out. The first of these components, $\boldsymbol{n}_0$, is just the usual vector normal to the surface of an ellipse, and describes the geometry of the particle boundary inside the unit cell at the micro scale. The second of the components, $\boldsymbol{n}_1$, accounts for the variation in the particle geometries at the macroscopic scale. For our ellipsoidal particles the first of these terms is

$$\boldsymbol{n}_0 = \boldsymbol{\nabla}_{\boldsymbol{y}}\chi = \mathrm{Q}\boldsymbol{y} + \boldsymbol{y}^T\mathrm{Q} = 2\mathrm{Q}\boldsymbol{y}, \tag{4.4.1.0.10}$$

owing to Q being symmetric. The second component is given by

$$\begin{aligned}
\boldsymbol{n}_1 &= \boldsymbol{\nabla}_{\boldsymbol{x}}\chi \\
&= \boldsymbol{\nabla}_{\boldsymbol{x}}\left[\left(\mathrm{E}\boldsymbol{y}\right)\cdot\left(\mathrm{E}\boldsymbol{y}\right)\right] \\
&= 2\left[\boldsymbol{\nabla}_{\boldsymbol{x}}\mathrm{E}\boldsymbol{y}\right]\mathrm{E}\boldsymbol{y} \\
&= 2\left(\mathrm{E}\boldsymbol{y}\right)^T\cdot\left(\mathrm{E}\mathrm{E}^{-1}\right)\left[\boldsymbol{\nabla}_{\boldsymbol{x}}\mathrm{E}\boldsymbol{y}\right]^T \\
&= \boldsymbol{n}_0\cdot\mathrm{E}^{-1}\left[\boldsymbol{\nabla}_{\boldsymbol{x}}\mathrm{E}\boldsymbol{y}\right]^T.
\end{aligned} \tag{4.4.1.0.11}$$

For brevity in our calculations, we define the term

$$\mathrm{N}\left(\boldsymbol{x},\boldsymbol{y}\right) = \mathrm{E}^{-1}\left[\boldsymbol{\nabla}_{\boldsymbol{x}}\mathrm{E}\boldsymbol{y}\right]^T, \tag{4.4.1.0.12}$$

and then our multiple scales expression for the varying particle normals is

$$\boldsymbol{n} = \boldsymbol{n}_0\cdot\left(\frac{\mathcal{I}}{\eta} + \mathrm{N}\right). \tag{4.4.1.0.13}$$

In the case that the particles are all the same, $\mathrm{N} = 0$ and the problem is exactly the same as in the previous section.

4.4.2 Homogenised Equations

With the geometry of our problem now defined we are in a position to derive the equations governing the potential inside our system. As before, the starting point is the electrostatic problem

$$\boldsymbol{\nabla}\cdot\left(\varepsilon_\alpha\boldsymbol{\nabla}\phi\right) = 0\,, \qquad \boldsymbol{x}\ \in\ \Omega_\alpha, \tag{4.4.2.0.1a}$$

$$\boldsymbol{n}\left(\boldsymbol{x}\right)\cdot\left(\varepsilon_p\boldsymbol{\nabla}\phi_p\right) = \boldsymbol{n}\left(\boldsymbol{x}\right)\cdot\left(\varepsilon_m\boldsymbol{\nabla}\phi_m\right)\,, \quad \boldsymbol{x}\ \in\ \partial\Omega_p, \tag{4.4.2.0.1b}$$

$$\phi(0) = 0, \tag{4.4.2.0.1c}$$

$$\phi(d) = V. \tag{4.4.2.0.1d}$$

The method proceeds very much as in the previous section, and so this derivation will be brief for the first two steps.

Expanding in powers of η as before, this problem becomes

$$\left\{\boldsymbol{\nabla}_{\boldsymbol{x}}\cdot\left(\varepsilon_{\alpha}\boldsymbol{\nabla}_{\boldsymbol{x}}\right)+\frac{1}{\eta}\left[\boldsymbol{\nabla}_{\boldsymbol{y}}\cdot\left(\varepsilon_{\alpha}\boldsymbol{\nabla}_{\boldsymbol{x}}\right)+\boldsymbol{\nabla}_{\boldsymbol{x}}\cdot\left(\varepsilon_{\alpha}\boldsymbol{\nabla}_{\boldsymbol{y}}\right)\right]+\frac{1}{\eta}\boldsymbol{\nabla}_{\boldsymbol{y}}\cdot\left(\varepsilon_{\alpha}\boldsymbol{\nabla}_{\boldsymbol{y}}\right)\right\}\left(\phi_{0}+\eta\phi_{1}+...\right)=0 \tag{4.4.2.0.2a}$$

for $\boldsymbol{y}\in\Omega_{\alpha}\left(\boldsymbol{x}\right)$, with the boundary condition

$$\begin{aligned}&\left[\boldsymbol{n}_{0}\cdot\left(\frac{\mathcal{I}}{\eta}+\mathrm{N}\right)\right]\cdot\left[\varepsilon_{p}\left(\boldsymbol{\nabla}_{\boldsymbol{x}}+\frac{1}{\eta}\boldsymbol{\nabla}_{\boldsymbol{y}}\right)\left(\phi_{0}+\eta\phi_{1}+...\right)\right]\\&\quad=\left[\boldsymbol{n}_{0}\cdot\left(\frac{\mathcal{I}}{\eta}+\mathrm{N}\right)\right]\cdot\left[\varepsilon_{m}\left(\boldsymbol{\nabla}_{\boldsymbol{x}}+\frac{1}{\eta}\boldsymbol{\nabla}_{\boldsymbol{y}}\right)\left(\phi_{0}+\eta\phi_{1}+...\right)\right]\end{aligned} \tag{4.4.2.0.2b}$$

for $\boldsymbol{y}\in\partial\Omega_{p}\left(\boldsymbol{x}\right)$. At leading order, $O\left(\eta^{-2}\right)$, we once again find that

$$\phi_{0}=\phi_{0}\left(\boldsymbol{x}\right). \tag{4.4.2.0.3}$$

Proceeding to the next order, and again making the ansatz that

$$\phi_{1}\left(\boldsymbol{x},\boldsymbol{y}\right)=\boldsymbol{\Gamma}\left(\boldsymbol{x},\boldsymbol{y}\right)\cdot\boldsymbol{\nabla}_{x}\phi_{0}, \tag{4.4.2.0.4}$$

we arrive at the cell problem

$$\partial y_{i}\left(\varepsilon_{\alpha}\right)_{ij}\left[\partial y_{j}(\Gamma_{\alpha})_{k}+\delta_{jk}\right]=0, \qquad \boldsymbol{y}\in\Omega_{\alpha}, \tag{4.4.2.0.5a}$$

$$(\hat{n}_{0})_{i}(\varepsilon_{m})_{ij}\left[\partial y_{j}(\Gamma_{m})_{k}+\delta_{jk}\right]=\varepsilon_{p}(\hat{n}_{0})_{j}\left[\partial y_{j}(\Gamma_{p})_{k}+\delta_{jk}\right], \qquad \boldsymbol{y}\in\partial\Omega_{p}, \tag{4.4.2.0.5b}$$

$$\boldsymbol{\Gamma}_{m}=\boldsymbol{\Gamma}_{p}, \qquad \boldsymbol{y}\in\partial\Omega_{p}, \tag{4.4.2.0.5c}$$

$$\boldsymbol{\Gamma}_{m}\text{ periodic on opposite sides of cell.} \tag{4.4.2.0.5d}$$

The only difference from the previous cell problem is that the $\boldsymbol{n}_{0}$ in the boundary condition was previously an $\boldsymbol{n}$. Proceeding to the final order we obtain

$$\boldsymbol{\nabla}_{\boldsymbol{x}}\cdot\left[\varepsilon_{\alpha}\left(\boldsymbol{\nabla}_{\boldsymbol{x}}\phi_{0}+\boldsymbol{\nabla}_{\boldsymbol{y}}\phi_{1}\right)\right]+\boldsymbol{\nabla}_{\boldsymbol{y}}\cdot\left[\varepsilon_{\alpha}\left(\boldsymbol{\nabla}_{\boldsymbol{x}}\phi_{1}+\boldsymbol{\nabla}_{\boldsymbol{y}}\phi_{2}\right)\right]=0\,, \qquad \boldsymbol{y}\in\Omega_{\alpha}\left(\boldsymbol{x}\right) \tag{4.4.2.0.6a}$$

and

$$\begin{aligned}&\boldsymbol{n}_{0}\cdot\left[\mathrm{N}\varepsilon_{p}\left(\boldsymbol{\nabla}_{\boldsymbol{x}}\phi_{0}+\boldsymbol{\nabla}_{\boldsymbol{y}}\phi_{1}\right)+\varepsilon_{p}\left(\boldsymbol{\nabla}_{\boldsymbol{x}}\phi_{1}+\boldsymbol{\nabla}_{\boldsymbol{y}}\phi_{2}\right)\right]\\&\quad=\boldsymbol{n}_{0}\cdot\left[\mathrm{N}\varepsilon_{m}\left(\boldsymbol{\nabla}_{\boldsymbol{x}}\phi_{0}+\boldsymbol{\nabla}_{\boldsymbol{y}}\phi_{1}\right)+\varepsilon_{m}\left(\boldsymbol{\nabla}_{\boldsymbol{x}}\phi_{1}+\boldsymbol{\nabla}_{\boldsymbol{y}}\phi_{2}\right)\right]\,, \qquad \boldsymbol{y}\in\partial\Omega_{p}\left(\boldsymbol{x}\right).\end{aligned} \tag{4.4.2.0.6b}$$

As in the previous section where the microscopic geometry was constant, we integrate equation (4.4.2.0.6a), apply the divergence theorem and equation (4.4.2.0.6b), giving us

that

$$0 = \int_{\Omega_\alpha(\boldsymbol{x})} \boldsymbol{\nabla}_{\boldsymbol{x}} \cdot \left[\varepsilon_\alpha \left(\boldsymbol{\nabla}_{\boldsymbol{x}}\phi_0 + \boldsymbol{\nabla}_{\boldsymbol{y}}\phi_1\right)\right] dV - \int_{\Omega_{\overline{\alpha}}(\boldsymbol{x})} \boldsymbol{\nabla}_{\boldsymbol{y}} \cdot \left[\varepsilon_{\overline{\alpha}} \left(\boldsymbol{\nabla}_{\boldsymbol{x}}\phi_1 + \boldsymbol{\nabla}_{\boldsymbol{y}}\phi_2\right)\right] dV$$
$$- \sum_{\alpha=m,p} \int_{\partial\Omega_p(\boldsymbol{x})} \boldsymbol{n}_0 \cdot \left[\mathrm{N}\varepsilon_\alpha \left(\boldsymbol{\nabla}_{\boldsymbol{x}}\phi_0 + \boldsymbol{\nabla}_{\boldsymbol{y}}\phi_1\right)\right] dA, \tag{4.4.2.0.7}$$

where, again, $\Omega_{\overline{\alpha}}$ is the complement of Ω_α (and noting that $\boldsymbol{n}_0$ of one domain is $-\boldsymbol{n}_0$ of the other). Next we add the α and $\overline{\alpha}$ equations together to obtain

$$\sum_{\alpha=m,p} \left\{ \int_{\Omega_\alpha(\boldsymbol{x})} \boldsymbol{\nabla}_{\boldsymbol{x}} \cdot \left[\varepsilon_\alpha \left(\boldsymbol{\nabla}_{\boldsymbol{x}}\phi_0 + \boldsymbol{\nabla}_{\boldsymbol{y}}\phi_1\right)\right] dV - \int_{\partial\Omega_p(\boldsymbol{x})} \boldsymbol{n}_0 \cdot \left[\mathrm{N}\varepsilon_\alpha \left(\boldsymbol{\nabla}_{\boldsymbol{x}}\phi_0 + \boldsymbol{\nabla}_{\boldsymbol{y}}\phi_1\right)\right] dA \right\} = 0. \tag{4.4.2.0.8}$$

At this point, the derivation deviates from the previous section owing to the fact that the domain we are integrating over is itself a function of $\boldsymbol{x}$. We cannot, then, simply take the derivative with respect to $\boldsymbol{x}$ outside of the integrals like before. Instead, we must use the transport theorem to change the order of integration and differentiation. The transport theorem usually concerns the rate of change of a quantity inside a material element that is changing with time. Simply put, the total rate of change is equal to the rate of change of the quantity inside the element plus the flux arising from the change of the element's boundary. Replacing time with the spatial position $\boldsymbol{x}$, the theorem applied to a vector quantity $\boldsymbol{f}$ is then

$$\boldsymbol{\nabla}_{\boldsymbol{x}} \cdot \int_{\Omega(\boldsymbol{x})} \boldsymbol{f} dV = \int_{\Omega(\boldsymbol{x})} \boldsymbol{\nabla} \cdot \boldsymbol{f} dV + \int_{\partial\Omega(\boldsymbol{x})} \left(\boldsymbol{n} \cdot \boldsymbol{\nabla}_{\boldsymbol{x}}\boldsymbol{b}\right) \boldsymbol{f} dS, \tag{4.4.2.0.9}$$

where $\boldsymbol{b}\left(\boldsymbol{x}\right) = \boldsymbol{y} \in \partial\Omega_p\left(\boldsymbol{x}\right)$ is the boundary of the particles at $\boldsymbol{x}$. The first integral in equation (4.4.2.0.8) becomes

$$\boldsymbol{\nabla}_{\boldsymbol{x}} \cdot \int_{\Omega_\alpha(\boldsymbol{x})} \varepsilon_\alpha \left(\boldsymbol{\nabla}_{\boldsymbol{x}}\phi_0 + \boldsymbol{\nabla}_{\boldsymbol{y}}\phi_1\right) dV - \int_{\partial\Omega_p(\boldsymbol{x})} \left(\boldsymbol{n}_0 \cdot \boldsymbol{\nabla}_{\boldsymbol{x}}\boldsymbol{b}\right) \cdot \varepsilon_\alpha \left(\boldsymbol{\nabla}_{\boldsymbol{x}}\phi_0 + \boldsymbol{\nabla}_{\boldsymbol{y}}\phi_1\right) dA \tag{4.4.2.0.10}$$

after applying the theorem. We now need to evaluate the quantity $\boldsymbol{n}_0 \cdot \boldsymbol{\nabla}_{\boldsymbol{x}}\boldsymbol{b}$. By differentiating χ with respect to $\boldsymbol{x}$ and applying the chain rule we can say that, at the boundaries,

$$\boldsymbol{\nabla}_{\boldsymbol{x}}\chi + \boldsymbol{\nabla}_{\boldsymbol{y}}\chi \cdot \boldsymbol{\nabla}_{\boldsymbol{x}}\boldsymbol{b}\left(\boldsymbol{x}\right) = 0. \tag{4.4.2.0.11}$$

By definition $\boldsymbol{\nabla}_{\boldsymbol{y}}\chi = \boldsymbol{n}_0$ and $\boldsymbol{\nabla}_{\boldsymbol{x}}\chi = \boldsymbol{n}_0 \cdot \mathrm{N}$, and so it follows that

$$\boldsymbol{n}_0 \cdot \boldsymbol{\nabla}_{\boldsymbol{x}}\boldsymbol{b}\left(\boldsymbol{x}\right) = -\boldsymbol{n}_0 \cdot \mathrm{N}. \tag{4.4.2.0.12}$$

Substituting this back into equation (4.4.2.0.10), we see that the surface integral arising from the application of the transport theorem cancels out with the surface integral in equation (4.4.2.0.8). We are then left with the final homogenised equation

$$\boldsymbol{\nabla}_{\boldsymbol{x}} \cdot \left(\varepsilon_{\mathrm{eff}} \boldsymbol{\nabla}_{\boldsymbol{x}}\phi_0\right) = 0, \tag{4.4.2.0.13}$$

where, as before,

$$\varepsilon_{\text{eff}} = \sum_{\alpha=m,p} \int_{\Omega_\alpha(\boldsymbol{x})} \varepsilon_\alpha \left[\mathcal{I} + \boldsymbol{\nabla}_{\boldsymbol{y}} \boldsymbol{\Gamma}_\alpha\right] dV. \tag{4.4.2.0.14}$$

The problem is then almost exactly the same as before, we just have that ε_{eff} is now dependent on $\boldsymbol{x}$ (as it should be), and this $\boldsymbol{x}$ dependence arises from the variable geometry of the unit cell. We must therefore solve the cell problem given by equations 4.4.2.0.5a-4.4.2.0.5d for an arbitrary particle with semi-axes $a(\boldsymbol{x})$ and $b(\boldsymbol{x})$, with an arbitrary orientation $\theta(\boldsymbol{x})$, and in a background medium with an arbitrary permittivity tensor.

4.5 Calculating the Effective Permittivity

4.5.1 Non-Rotating Particles in Isotropic Medium

4.5.1.1 The General Cell Problem

To model a system of disordered particles, we make the assumption that the volume fraction of the particles is low, as in [7]. This is equivalent to the dimension of the particles being small with respect to the unit cell. Mathematically, the particles then have major and minor semi-axes βa and βb, where $\beta \ll 1$. In this limit, the solution in the unit cell is determined only by the small inner region containing the particle and is constant at the periodic cell boundaries. This means the effect of the particle is not felt at the boundaries and so presents no information about the particle's position. As a result, a disordered configuration is equivalent to the usual homogenised setup with a small volume fraction. This does not present a problem since this reflects real doped liquid crystal systems.

We will solve the cell problem given in equations (4.4.2.0.5a)-(4.4.2.0.5d) for low volume fraction particles. This can be achieved analytically via a matched asymptotic expansion. We do so by dividing the cell into an outer region, where$|\boldsymbol{y}| \gg O(\beta)$, and an inner region, with coordinates $\boldsymbol{Y} = \boldsymbol{y}/\beta$. This allows us to look for solutions γ and $\boldsymbol{\Gamma}$ in the inner and outer regions, respectively, where

$$\boldsymbol{\Gamma}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{\Gamma}^{(0)}(\boldsymbol{x}, \boldsymbol{y}) + \beta \boldsymbol{\Gamma}^{(1)}(\boldsymbol{x}, \boldsymbol{y}) + ... \tag{4.5.1.1.1}$$

and

$$\gamma(\boldsymbol{x}, \boldsymbol{Y}) = \gamma^{(0)}(\boldsymbol{x}, \boldsymbol{Y}) + \beta \gamma^{(1)}(\boldsymbol{x}, \boldsymbol{Y}) + ... \tag{4.5.1.1.2}$$

The cell problem in the inner region then becomes

$$\boldsymbol{\nabla}_{\boldsymbol{Y}} \cdot \left[\varepsilon_\alpha \boldsymbol{\nabla}_{\boldsymbol{Y}} \left(\gamma^{(0)} + \beta \gamma^{(1)} + ...\right)\right] = 0\,, \qquad \boldsymbol{Y} \in \Omega_\alpha(\boldsymbol{x})\,, \tag{4.5.1.1.3a}$$

$$\varepsilon_p \boldsymbol{n}_0 \cdot \left[\beta\mathcal{I} + \boldsymbol{\nabla}_{\boldsymbol{Y}}\gamma_p\right] = \boldsymbol{n}_0 \cdot \left\{\varepsilon_m \left[\beta\mathcal{I} + \boldsymbol{\nabla}_{\boldsymbol{Y}}\gamma_m\right]\right\}\,, \quad \boldsymbol{Y} \in \partial\Omega_p\left(\boldsymbol{x}\right), \tag{4.5.1.1.3b}$$

$$\gamma_m = \gamma_p\,, \qquad \boldsymbol{Y} \in \partial\Omega_p\left(\boldsymbol{x}\right), \tag{4.5.1.1.3c}$$

while the cell problem in the outer region becomes

$$\boldsymbol{\nabla}_{\boldsymbol{y}} \cdot \left[\varepsilon_\alpha \boldsymbol{\nabla}_{\boldsymbol{y}} \left(\boldsymbol{\Gamma}^{(0)} + \beta\boldsymbol{\Gamma}^{(1)} + ...\right)\right] = 0\,, \quad \boldsymbol{y} \in \Omega_\alpha\left(\boldsymbol{x}\right), \tag{4.5.1.1.3d}$$

$$\boldsymbol{\Gamma} \text{ periodic in } \boldsymbol{y}. \tag{4.5.1.1.3e}$$

The problem is closed with the condition that γ matches $\boldsymbol{\Gamma}$ as $|\boldsymbol{Y}| \to \infty$.

We now look to solve the problem at each order in β. At all orders the outer solution is governed by

$$\boldsymbol{\nabla}_{\boldsymbol{y}} \cdot \left[\varepsilon_\alpha \boldsymbol{\nabla}_{\boldsymbol{y}} \boldsymbol{\Gamma}^{(i)}\right] = 0\,, \quad \boldsymbol{y} \in \Omega_\alpha\left(\boldsymbol{x}\right), \tag{4.5.1.1.4}$$

$$\boldsymbol{\Gamma}^{(i)} \text{ periodic in } \boldsymbol{y}, \tag{4.5.1.1.5}$$

which has the solution that

$$\boldsymbol{\Gamma}^{(i)} = \boldsymbol{\Gamma}^{(i)}(\boldsymbol{x}), \tag{4.5.1.1.6}$$

and hence

$$\boldsymbol{\Gamma} = \sum_{i=0}^{\infty} \beta^i \boldsymbol{\Gamma}^{(i)} = \boldsymbol{\Gamma}(\boldsymbol{x}). \tag{4.5.1.1.7}$$

Turning to the inner solution, at leading order we have

$$\boldsymbol{\nabla}_{\boldsymbol{Y}} \cdot \left[\varepsilon_\alpha \boldsymbol{\nabla}_{\boldsymbol{Y}} \gamma^{(0)}\right] = 0\,, \qquad \boldsymbol{Y} \in \Omega_\alpha\left(\boldsymbol{x}\right), \tag{4.5.1.1.8a}$$

$$\boldsymbol{n}_0 \cdot \left[\varepsilon_m \boldsymbol{\nabla}_{\boldsymbol{Y}} \gamma_m^{(0)}\right] = \varepsilon_p \boldsymbol{n}_0 \cdot \boldsymbol{\nabla}_{\boldsymbol{Y}} \gamma_p^{(0)}\,, \quad \boldsymbol{Y} \in \partial\Omega_p\left(\boldsymbol{x}\right), \tag{4.5.1.1.8b}$$

$$\gamma_m^{(0)} = \gamma_p^{(0)}\,, \qquad \boldsymbol{Y} \in \partial\Omega_p\left(\boldsymbol{x}\right), \tag{4.5.1.1.8c}$$

$$\gamma_m^{(0)} \to \boldsymbol{\Gamma}^{(0)}(\boldsymbol{x})\,, \qquad |\boldsymbol{Y}| \to \infty, \tag{4.5.1.1.8d}$$

which has the solution that

$$\gamma^{(0)} = \boldsymbol{\Gamma}^{(0)}\left(\boldsymbol{x}\right). \tag{4.5.1.1.9}$$

Moreover, the inner problem is the same for all orders in β but the first, and so we conclude that

$$\gamma_\alpha = \gamma_\alpha^{const.}\left(\boldsymbol{x}\right) + \beta\gamma_\alpha^{(1)}\left(\boldsymbol{x}, \boldsymbol{Y}\right)\,, \quad \boldsymbol{Y} \in \Omega_\alpha, \tag{4.5.1.1.10}$$

where $\gamma_\alpha^{const.}\left(\boldsymbol{x}\right)$ is an undetermined constant.

Now, the purpose of calculating γ is to compute the effective permittivity, which concerns integrating $\boldsymbol{y}$ derivatives of $\boldsymbol{\Gamma}$. Upon differentiating with respect to $\boldsymbol{y}$, the only term that does not vanish is that containing $\gamma^{(1)}$, which is therefore the only term that we need to calculate. This does not even require us to determine $\boldsymbol{\Gamma^{(1)}}(\boldsymbol{x})$, since it

turns out that the requirement that $\gamma^{(1)}$ goes to any constant at infinity is a sufficient matching condition for its calculation. The only problem that needs solving, then, is the first order problem of the inner solution, given by

$$\nabla_{\boldsymbol{Y}} \cdot \left[\varepsilon_\alpha \nabla_{\boldsymbol{Y}} \gamma_\alpha^{(1)}\right] = 0\,, \qquad \boldsymbol{Y} \in \Omega_\alpha(\boldsymbol{x}) \tag{4.5.1.1.11a}$$

$$\varepsilon_p \boldsymbol{n}_0 \cdot \left[\mathcal{I} + \nabla_{\boldsymbol{Y}} \gamma_p^{(1)}\right] = \boldsymbol{n}_0 \cdot \left\{\varepsilon_m \left[\mathcal{I} + \nabla_{\boldsymbol{Y}} \gamma_m^{(1)}\right]\right\}\,, \quad \boldsymbol{Y} \in \partial\Omega_p(\boldsymbol{x}) \tag{4.5.1.1.11b}$$

$$\gamma_m^{(1)} = \gamma_p^{(1)}\,, \qquad \boldsymbol{Y} \in \partial\Omega_p(\boldsymbol{x}) \tag{4.5.1.1.11c}$$

$$\gamma_m^{(1)} \text{ constant as } |\boldsymbol{Y}| \to \infty. \tag{4.5.1.1.11d}$$

Whilst the problem inside the particle is fairly simple, the problem in the medium is significantly more complex. The governing equations present mixed derivatives and the boundary conditions must be satisfied on the variable elliptical particle boundary. We therefore first continue with a simplified case in which we develop a framework to solve the problem.

4.5.1.2 Simplified First Order Matched Asymptotic Problem

We will first consider this problem for an isotropic surrounding medium and particles whose axes lie along the y_1 and y_2 axes. Under these simplifications the cell problem in equations (4.5.1.1.11a)-(4.5.1.1.11d) becomes

$$\nabla_{\boldsymbol{Y}}^2 \gamma_\alpha^{(1)} = 0\,, \qquad \boldsymbol{Y} \in \Omega_\alpha, \tag{4.5.1.2.1a}$$

$$\gamma_m^{(1)} = \gamma_p^{(1)}\,, \qquad \boldsymbol{Y} \in \partial\Omega_p(\boldsymbol{x})\,, \tag{4.5.1.2.1b}$$

$$\gamma_m^{(1)} \text{ constant as } |\boldsymbol{Y}| \to \infty, \tag{4.5.1.2.1c}$$

along with the remaining boundary condition

$$\begin{pmatrix} a^{-2}y_1 \\ b^{-2}y_2 \end{pmatrix} \cdot \left[\varepsilon_p \left(\mathcal{I} + \nabla_{\boldsymbol{Y}} \gamma_p^{(1)}\right)\right] = \begin{pmatrix} a^{-2}y_1 \\ b^{-2}y_2 \end{pmatrix} \cdot \left[\varepsilon_m \left(\mathcal{I} + \nabla_{\boldsymbol{Y}} \gamma_m^{(1)}\right)\right], \quad \boldsymbol{Y} \in \partial\Omega_p, \tag{4.5.1.2.1d}$$

where βa and βb are the ($\boldsymbol{x}$ dependent) major and minor semi-axes of the particle.

We now adopt elliptic coordinates defined by

$$\boldsymbol{Y} = \begin{pmatrix} F\cosh\mu\cos\nu \\ F\sinh\mu\sin\nu \end{pmatrix}, \tag{4.5.1.2.2}$$

where $\mu \geq 0$ and $0 \leq \nu < 2\pi$. In this system, lines of constant μ give ellipses and lines of constant ν give hyperbolas, both of which are orthogonal to one another. The

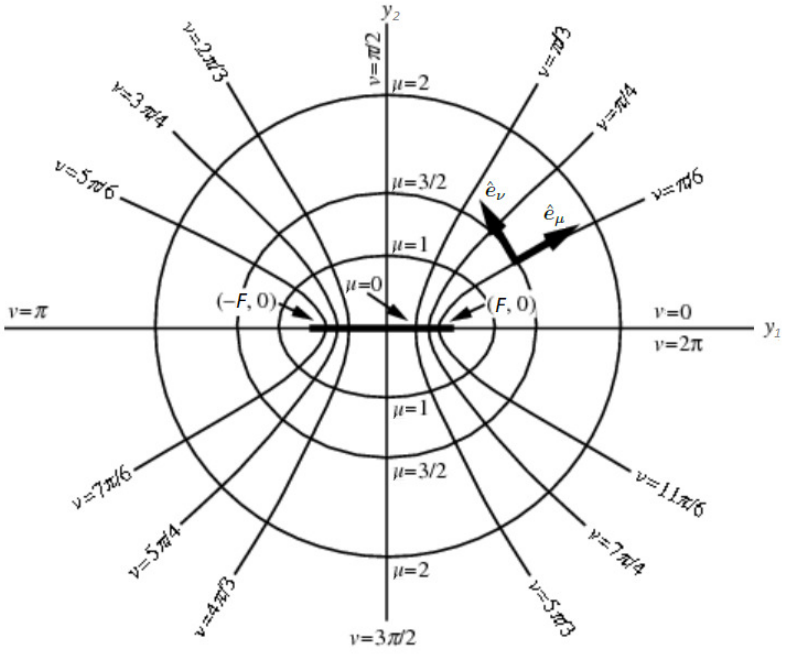


FIGURE 4.9: Depiction of the elliptic coordinate system [58]. The curves of constant μ trace out ellipses with decreasing eccentricity, while the curves of constant ν trace out hyperbolas. F is the focal length of the coordinate system, with a larger F giving more eccentric ellipses. The unit vectors $\hat{e}_\mu$ and $\hat{e}_\nu$ are normal and tangential to the ellipses, respectively.

parameter F is the focal length, i.e., the foci of the ellipses and the hyperbolas lie at $(\pm F, 0)$. A depiction of the coordinate system is shown in figure 4.9. The derivatives in this coordinate system are

$$\boldsymbol{\nabla}_Y = \frac{1}{F\sqrt{\sinh^2\mu + \sin^2\nu}}\left(\frac{\partial}{\partial\mu}, \frac{\partial}{\partial\nu}\right) \tag{4.5.1.2.3}$$

and

$$\nabla_Y^2 = \frac{1}{F^2\left(\sinh^2\mu + \sin^2\nu\right)}\left(\frac{\partial^2}{\partial\mu^2} + \frac{\partial^2}{\partial\nu^2}\right). \tag{4.5.1.2.4}$$

This presents a very natural choice of coordinate system. Not only do the governing equations remain as Laplace's equation, but it greatly simplifies the geometrical aspects of the problem. With the correct choice of F we can paramaterise the particle surface by the line $\mu = \mu_0$, and we also have that the particle normal is simply the basis vector $\boldsymbol{e}_{\mu_0}$. This simplifies matters even further since the normal derivative in the boundary condition is just the derivative with respect to μ.

To describe the problem fully, we start by noting that at the particle boundary

$$\left(\frac{Y_1}{a}\right)^2 + \left(\frac{Y_2}{b}\right)^2 = \left(\frac{F\cosh\mu_0\cos\nu}{a}\right)^2 + \left(\frac{F\sinh\mu_0\sin\nu}{b}\right)^2 = 1. \tag{4.5.1.2.5}$$

Clearly we then require

$$F\cosh\mu_0 = a \tag{4.5.1.2.6}$$

and

$$F\sinh\mu_0 = b. \tag{4.5.1.2.7}$$

This then gives that

$$F = \sqrt{a^2 - b^2} \tag{4.5.1.2.8}$$

and

$$\mu_0\left(\boldsymbol{x}\right) = \tanh^{-1}\left(\frac{b\left(\boldsymbol{x}\right)}{a\left(\boldsymbol{x}\right)}\right) = \frac{1}{2}\ln\left(\frac{a\left(\boldsymbol{x}\right)+b\left(\boldsymbol{x}\right)}{a\left(\boldsymbol{x}\right)-b\left(\boldsymbol{x}\right)}\right). \tag{4.5.1.2.9}$$

Putting this all together, the problem in equations (4.5.1.2.1a)-(4.5.1.2.1d) becomes

$$\left(\frac{\partial^2}{\partial\mu^2}+\frac{\partial^2}{\partial\nu^2}\right)\gamma^{(1)} = 0\,, \qquad \boldsymbol{Y}\in\Omega_\alpha \tag{4.5.1.2.10a}$$

$$\varepsilon_p\frac{\partial\gamma_p^{(1)}}{\partial\mu} - \varepsilon_m\frac{\partial\gamma_m^{(1)}}{\partial\mu} = \left(\varepsilon_m-\varepsilon_p\right)\begin{pmatrix} b\cos\nu \\ a\sin\nu\end{pmatrix}\,, \quad \boldsymbol{Y}\in\partial\Omega_p, \tag{4.5.1.2.10b}$$

$$\gamma_m^{(1)} = \gamma_p^{(1)}\,, \qquad \boldsymbol{Y}\in\partial\Omega_p, \tag{4.5.1.2.10c}$$

after multiplying both sides of the boundary condition by ab.

We now make the ansatz that

$$\gamma_\alpha^{(1)} = \begin{pmatrix} M_{\alpha1}\left(\mu\right)b\cos\nu \\ M_{\alpha2}\left(\mu\right)a\sin\nu\end{pmatrix} \tag{4.5.1.2.11}$$

for $\alpha = m, p$. Substituting into equation (4.5.1.2.10a) yields that

$$M_{\alpha i} = A_{\alpha i}\left(\boldsymbol{x}\right)e^{\mu} + B_{\alpha i}\left(\boldsymbol{x}\right)e^{-\mu} \tag{4.5.1.2.12}$$

for $i = 1, 2$. The leading order solution is constant and so we must have that $A_{mi}\left(\boldsymbol{x}\right) = 0$ for the first order solution to be bounded (and smaller than the leading order). The continuity condition in equation (4.5.1.2.10c) implies that

$$B_{mi}e^{-\mu_0} = A_{pi}e^{\mu_0} + B_{pi}e^{-\mu_0}, \tag{4.5.1.2.13}$$

and the boundary condition in equation (4.5.1.2.10b) that

$$\varepsilon_p\left(A_{pi}e^{\mu_0} - B_{pi}e^{-\mu_0}\right) + \varepsilon_m B_{mi}e^{-\mu_0} = \varepsilon_m - \varepsilon_p. \tag{4.5.1.2.14}$$

Eliminating B_{mi} we find that

$$A_{pi} = \delta\varepsilon\left(1 - B_{pi}e^{-\mu_0}\right)e^{-\mu_0}, \tag{4.5.1.2.15}$$

where we have defined

$$\delta\varepsilon = \frac{\varepsilon_m - \varepsilon_p}{\varepsilon_m + \varepsilon_p}. \tag{4.5.1.2.16}$$

After eliminating A_{pi} we find

$$B_{mi} = \delta\varepsilon e^{\mu_0} + (1 - \delta\varepsilon)\, B_{pi}. \tag{4.5.1.2.17}$$

The solution is then of the form

$$\gamma_{mi}^{(1)} = \left[\delta\varepsilon e^{\mu_0} + (1 - \delta\varepsilon)\, B_{pi}\right] e^{-\mu} n_i \tag{4.5.1.2.18}$$

$$\gamma_{pi}^{(1)} = \left[\delta\varepsilon \left(1 - B_{pi} e^{-\mu_0}\right) e^{\mu - \mu_0} + B_{pi} e^{-\mu}\right] n_i, \tag{4.5.1.2.19}$$

where $\boldsymbol{n} = (b\cos\nu, a\sin\nu)^T$ and the B_{pi} are still to be determined.

With no other boundary conditions we must instead turn to the nature of the elliptic coordinate system to determine these constants. To do so, we start by noting that $\mu = 0$ describes the degenerate ellipse forming a line segment between the foci at $(\pm F, 0)$. The elliptic coordinates $(0, \pm\nu)$ then describe the same point on the line segment, and $\boldsymbol{\gamma}_p^{(1)}$ must have the same value at these points. Since $\sin(-\nu) = -\sin(\nu)$, it follows that we must have

$$M_{p2}|_{\mu=0} = \delta\varepsilon \left(1 - B_{p2} e^{-\mu_0}\right) e^{-\mu_0} + B_{p2} = 0. \tag{4.5.1.2.20}$$

Rearranging this we find that

$$B_{p2} = -\frac{\delta\varepsilon e^{-\mu_0}}{1 - \delta\varepsilon e^{-2\mu_0}}. \tag{4.5.1.2.21}$$

Finding B_{p1} is more difficult owing to the fact that $\cos(-\nu) = \cos(\nu)$ means continuity as $\mu \to 0$ is automatically satisfied for any value of B_{p1}. As a result, we must instead consider the gradient of $\gamma_{p1}^{(1)}$ along the degenerate line segment; specifically, the component $\boldsymbol{e}_\mu \frac{\partial}{\partial\mu}$ and how it must be continuous. If we imagine standing at the point $(0, \nu)$, the unit vector $\boldsymbol{e}_\mu$ has a positive cartesian y_2 component. Conversely, it has a negative y_2 component at the point $(0, -\nu)$. It then follows that for the gradient to be continuous and well defined at $(0, \pm\nu)$ we must have that

$$\left.\frac{\partial \gamma_{p1}^{(1)}}{\partial\mu}\right|_{(\mu,\nu)=(0,\nu)} = -\left.\frac{\partial \gamma_{p1}^{(1)}}{\partial\mu}\right|_{(\mu,\nu)=(0,-\nu)}. \tag{4.5.1.2.22}$$

From this we find that

$$B_{p1} = \frac{\delta\varepsilon e^{-\mu_0}}{1 + \delta\varepsilon e^{-2\mu_0}}, \tag{4.5.1.2.23}$$

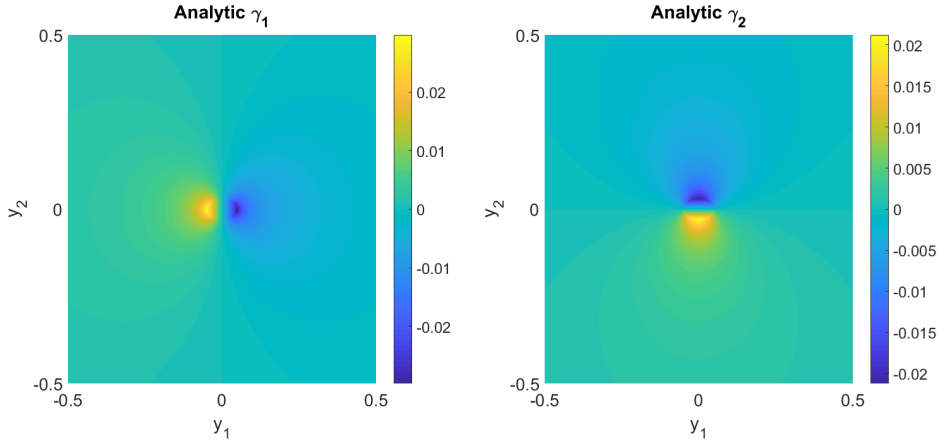


FIGURE 4.10: Analytic solution to the cell problem for non-rotated particles in isotropic medium.

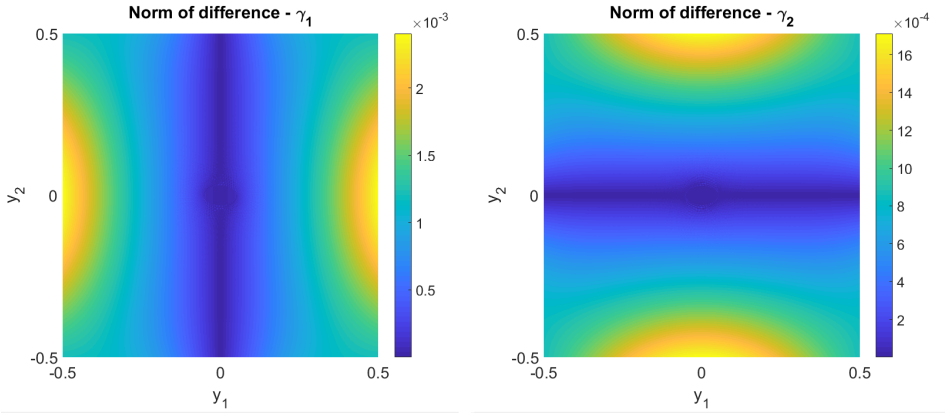


FIGURE 4.11: Comparison of the analytic solution for a non-rotated particle in isotropic medium with the finite elements solution provided by Comsol.

and putting this all together we arrive at

$$\gamma_m^{(1)} = ab\sqrt{\frac{a+b}{a-b}}\left(\varepsilon_m - \varepsilon_p\right) e^{-\mu} \begin{pmatrix} \left(a\varepsilon_m + b\varepsilon_p\right)^{-1} \cos\nu \\ \left(a\varepsilon_p + b\varepsilon_m\right)^{-1} \sin\nu \end{pmatrix} \tag{4.5.1.2.24}$$

$$\gamma_p^{(1)} = \sqrt{a^2 - b^2}\left(\varepsilon_m - \varepsilon_p\right) \begin{pmatrix} b\left(a\varepsilon_m + b\varepsilon_p\right)^{-1} \cosh\mu\cos\nu \\ a\left(a\varepsilon_p + b\varepsilon_m\right)^{-1} \sinh\mu\sin\nu \end{pmatrix}. \tag{4.5.1.2.25}$$

Shown in figure 4.10 is the solution for $\gamma^{(1)}$, and a comparison to the finite elements solution to the cell problem provided by Comsol is shown in figure 4.11. The larger errors at the cell boundary are a result of slightly differing boundary conditions; the analytic solution goes to zero at infinity whilst the solution from Comsol goes to zero at the cell boundaries. This error decreases as the particles become smaller and the approximation that the cell boundaries lie at infinity becomes more accurate. The particles used for these comparisons are necessarily large so that they can be seen clearly in the graphing process. As will be seen later, we only need the solution at the particle boundary to calculate the effective permittivity. Given that the two solutions

are in excellent agreement near the particle, with an error that is at least two orders of magnitude smaller than the solution itself, we conclude that the model provides an accurate representation of the solution, insofar as we are interested.

4.5.2 First Order Problem for Rotating Particles in Isotropic Medium

We continue now by relaxing the simplification on the particle orientations and allow the particle to lie at an angle θ with the Y_1 axis. We rotate our axes anti-clockwise by the angle θ so that the major axis of the particle lies along the first of the new axes. This induces the coordinate change $\tilde{\boldsymbol{Y}} = \mathrm{R}(\theta)\,\boldsymbol{Y}$, where $\mathrm{R}(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$. We can then write

$$\tilde{\boldsymbol{\gamma}}^{(1)} = \mathrm{R}(\theta)\,\boldsymbol{\gamma}^{(1)} \tag{4.5.2.1}$$

$$\tilde{\boldsymbol{n}}_0 = \mathrm{R}(\theta)\,\boldsymbol{n}_0 \tag{4.5.2.2}$$

owing to the transformation law for contravariant vectors, while the gradient transforms covariantly as

$$\frac{\partial}{\partial Y_i}\hat{\boldsymbol{e}}_i = \frac{\partial \tilde{Y}_j}{\partial Y_i}\frac{\partial}{\partial \tilde{Y}_j}\hat{\boldsymbol{e}}_i = \mathrm{R}^T_{ij}\frac{\partial}{\partial \tilde{Y}_j}\hat{\boldsymbol{e}}_i. \tag{4.5.2.3}$$

It follows that

$$\frac{\partial^2 \gamma_j}{\partial Y_i^2} = \mathrm{R}^T_{ij}\frac{\partial}{\partial \tilde{Y}_j}\left[\mathrm{R}^T_{ik}\frac{\partial}{\partial \tilde{Y}_k}\left(\mathrm{R}_{lm}\tilde{\gamma}_m\right)\right] = \delta_{jk}\frac{\partial}{\partial \tilde{Y}_j}\left(\frac{\partial}{\partial \tilde{Y}_k}\tilde{\gamma}_m\right)\mathrm{R}_{lm} = \frac{\partial^2 \tilde{\gamma}_m}{\partial \tilde{Y}_k^2}\mathrm{R}_{lm} \tag{4.5.2.4}$$

and so equation (4.5.1.1.11a) simply becomes

$$\nabla^2_{\tilde{\boldsymbol{Y}}}\tilde{\gamma}^{(1)}_\alpha = 0\,, \qquad \tilde{\boldsymbol{Y}} \in \tilde{\Omega}_\alpha(\boldsymbol{x})\,, \tag{4.5.2.5}$$

after multiplying both sides by $\mathrm{R}^T(\theta)$. The boundary condition becomes

$$\begin{aligned} 0 &= \mathrm{R}_{ik}\tilde{n}_k\left[\left(\varepsilon_p - \varepsilon_m\right)\delta_{ij} + \mathrm{R}_{il}\partial_{\tilde{l}}\left(\mathrm{R}_{jm}\tilde{\gamma}_m\right)\right] \\ &= \tilde{n}_k\left[\left(\varepsilon_p - \varepsilon_m\right)\mathrm{R}_{jk} + \delta_{lk}\mathrm{R}_{jm}\partial_{\tilde{l}}\tilde{\gamma}_m\right] \\ &= \tilde{n}_k\left[\left(\varepsilon_p - \varepsilon_m\right)\delta_{pk} + \partial_{\tilde{k}}\tilde{\gamma}_p\right] \end{aligned} \tag{4.5.2.6}$$

after multiplying by R_{pj}. The problem for $\tilde{\boldsymbol{\gamma}}^{(1)}$ is then the same as for the unrotated particle, and it follows that the solution for a rotated particle is the rotation of the solution for an axis aligned particle:

$$\gamma^{(1)}_m = ab\sqrt{\frac{a+b}{a-b}}\left(\varepsilon_m - \varepsilon_p\right)e^{-\tilde{\mu}}\mathrm{R}^T(\theta)\begin{pmatrix} \left(a\varepsilon_m + b\varepsilon_p\right)^{-1}\cos\tilde{v} \\ \left(a\varepsilon_m + b\varepsilon_p\right)^{-1}\sin\tilde{v} \end{pmatrix} \tag{4.5.2.7}$$

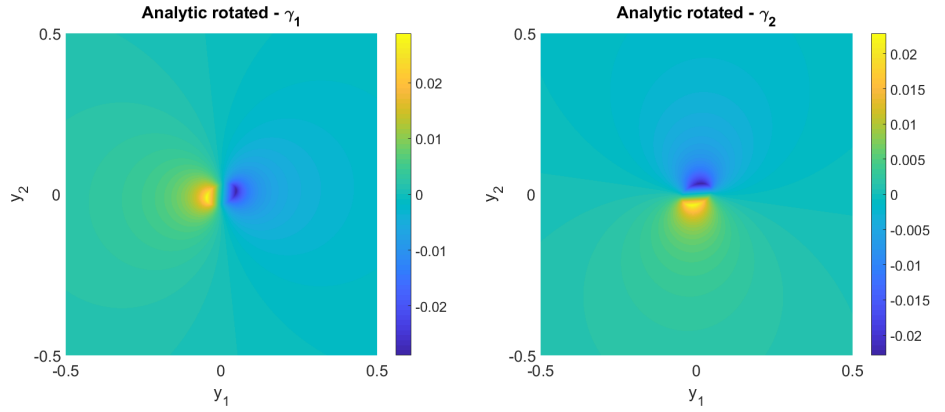


FIGURE 4.12: The analytic solution to the first order cell problem for a particle at an angle of $\pi/8$, $\varepsilon_m = 1$, and $\varepsilon_p = 2$.

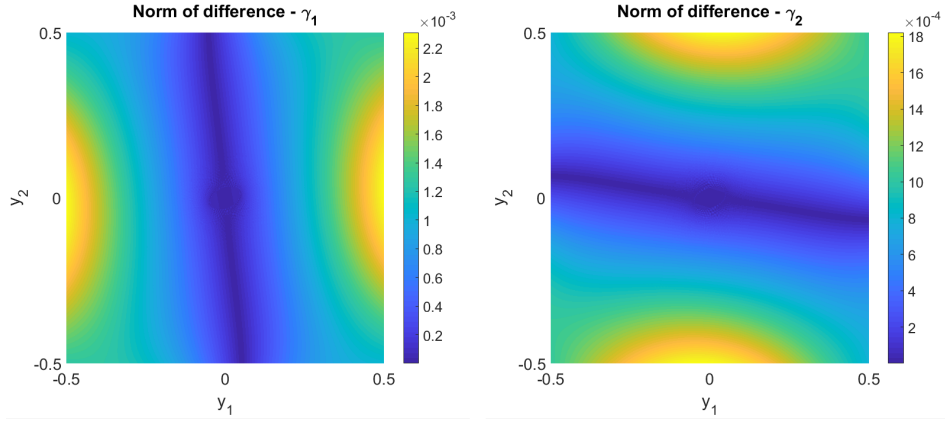


FIGURE 4.13: Norm of the difference between the analytic solution and Comsol solution for a particle at an angle of $\pi/8$, $\varepsilon_m = 1$, and $\varepsilon_p = 2$.

$$\gamma_p^{(1)} = \sqrt{a^2 - b^2}\left(\varepsilon_m - \varepsilon_p\right) \mathrm{R}^T(\theta) \begin{pmatrix} b\left(a\varepsilon_m + b\varepsilon_p\right)^{-1} \cosh\tilde{\mu}\cos\tilde{\nu} \\ a\left(b\varepsilon_m + a\varepsilon_p\right)^{-1} \sinh\tilde{\mu}\sin\tilde{\nu} \end{pmatrix}. \tag{4.5.2.8}$$

The rotated solution to the first order cell problem for a particle at $\pi/8$ is shown in figure 4.12 along with a comparison to the finite elements solution in figure 4.13. As expected for a linear combination of the unrotated solutions, the analytic rotated solutions are again in good agreement with those from finite elements in the region of the particle.

4.5.3 Effective Permittivity for Particles in Isotropic Medium

Now that we have the solution to the first order inner problem, we are able to calculate the effective permittivity, which we redefine for convenience:

$$\varepsilon_{\text{eff}} = \sum_{\alpha=m,p} \int_{\Omega_\alpha(\boldsymbol{x})} \varepsilon_\alpha \left[\mathcal{I} + \boldsymbol{\nabla}_{\boldsymbol{y}}\boldsymbol{\Gamma}\right] dV. \tag{4.5.3.1}$$

Taking the unit cell to have a volume of 1, we immediately obtain that

$$\sum_{\alpha=m,p} \varepsilon_\alpha \int_{\Omega_\alpha} \mathcal{I} dV = \left[\left(\varepsilon_p - \varepsilon_m \right) \left| \Omega_p \right| + \varepsilon_m \right] \mathcal{I}, \tag{4.5.3.2}$$

where $\left|\Omega_p\right|$ is the volume of particles. We now turn our attention to the remaining integrals. The Gradient theorem states that for a scalar field ψ and a closed volume V (bounded by S)

$$\int \boldsymbol{\nabla} \psi dV = \int \psi \boldsymbol{n} dS. \tag{4.5.3.3}$$

We now consider integrating the gradient of a vector field $\boldsymbol{\psi}$. For any constant vector field $\boldsymbol{v}$, we have that

$$\begin{aligned} \int \boldsymbol{\nabla} \boldsymbol{\psi} dV \cdot \boldsymbol{v} &= \int \boldsymbol{\nabla} \left(\boldsymbol{\psi} \cdot \boldsymbol{v} \right) dV \\ &= \int \boldsymbol{n} \left(\boldsymbol{\psi} \cdot \boldsymbol{v} \right) dS \\ &= \int \boldsymbol{n} \otimes \boldsymbol{\psi} dS \cdot \boldsymbol{v}, \end{aligned} \tag{4.5.3.4}$$

where the second line follows from the Gradient theorem. Since this result holds true for all constant vectors $\boldsymbol{v}$, we must have

$$\int \boldsymbol{\nabla} \boldsymbol{\psi} dV = \int \boldsymbol{n} \otimes \boldsymbol{\psi} dS. \tag{4.5.3.5}$$

Applying this result to (4.5.3.1), we find

$$\begin{aligned} \sum_\alpha \varepsilon_\alpha \int_{\Omega_\alpha(\boldsymbol{x})} \boldsymbol{\nabla}_{\boldsymbol{y}} \boldsymbol{\Gamma} dV &= \left(\varepsilon_p - \varepsilon_m \right) \int_{\partial \Omega_p(\boldsymbol{x})} \boldsymbol{n} \otimes \beta \boldsymbol{\gamma}_p^{(1)} dS \\ &= \left(\varepsilon_p - \varepsilon_m \right) \int_{\partial \Omega_p(\boldsymbol{x})} \boldsymbol{\nabla}_{\boldsymbol{y}} \beta \boldsymbol{\gamma}_p^{(1)} dS \end{aligned} \tag{4.5.3.6}$$

(where $\boldsymbol{n}$ is the outward particle normal), owing to the periodicity of the outer solution $\boldsymbol{\Gamma}$ at the cell boundaries and to the continuity of the inner solution, $\gamma^{(1)}$, at the particle boundary. We now note that the particle solution 4.5.2.8 can be written as

$$\begin{aligned} \boldsymbol{\gamma}_p^{(1)} &= \left(\varepsilon_m - \varepsilon_p \right) \mathrm{R}^T(\theta) \begin{pmatrix} b \left(a\varepsilon_m + b\varepsilon_p \right)^{-1} & 0 \\ 0 & a \left(b\varepsilon_m + a\varepsilon_p \right)^{-1} \end{pmatrix} \tilde{\boldsymbol{Y}} \\ &= \left(\varepsilon_m - \varepsilon_p \right) \mathrm{R}^T(\theta) \begin{pmatrix} b \left(a\varepsilon_m + b\varepsilon_p \right)^{-1} & 0 \\ 0 & a \left(b\varepsilon_m + a\varepsilon_p \right)^{-1} \end{pmatrix} \mathrm{R}(\theta) \frac{\boldsymbol{y}}{\beta}. \end{aligned} \tag{4.5.3.7}$$

The term $\boldsymbol{\nabla}_{\boldsymbol{y}} \boldsymbol{y} = \mathcal{I}$, and so the integral simply becomes

$$\sum_\alpha \varepsilon_\alpha \int_{\Omega_\alpha(\boldsymbol{x})} \boldsymbol{\nabla}_{\boldsymbol{y}} \boldsymbol{\Gamma} dV = - \left| \Omega_p \right| \left(\varepsilon_m - \varepsilon_p \right)^2 \mathrm{R}^T(\theta) \begin{pmatrix} b \left(a\varepsilon_m + b\varepsilon_p \right)^{-1} & 0 \\ 0 & a \left(b\varepsilon_m + a\varepsilon_p \right)^{-1} \end{pmatrix} \mathrm{R}(\theta), \tag{4.5.3.8}$$

where $|\Omega_p|$ is the volume fraction of the particles. Writing

$$\hat{\varepsilon}_{11} = -|\Omega_p| \frac{b\left(\varepsilon_m - \varepsilon_p\right)^2}{\left(a\varepsilon_m + b\varepsilon_p\right)} \tag{4.5.3.9}$$

and

$$\hat{\varepsilon}_{22} = -|\Omega_p| \frac{a\left(\varepsilon_m - \varepsilon_p\right)^2}{\left(b\varepsilon_m + a\varepsilon_p\right)}, \tag{4.5.3.10}$$

we find that

$$\varepsilon_{\text{eff}} = \left[\left(\varepsilon_p - \varepsilon_m\right)|\Omega_p| + \varepsilon_m\right]\mathcal{I} + \left\{\frac{\hat{\varepsilon}_{11} + \hat{\varepsilon}_{22}}{2}\mathcal{I} + \frac{\hat{\varepsilon}_{11} - \hat{\varepsilon}_{22}}{2}\begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix}\right\}. \tag{4.5.3.11}$$

The form of the effective permittivity is in excellent agreement with the previous simulations in section 4.3.4. The first term of the effective permittivity is an isotropic background that is the weighted volume average of the component permittivities. This agrees with the first observation of the previous simulations, where the addition of the particles shifted the overall average level of the effective permittivity towards that of the particles'. The second term of the effective permittivity induces the anisotropy that is caused by the particles' geometries. It behaves exactly like the liquid crystal Q-tensor, with its first component being the average value of the induced anisotropy in the y_1 and y_2 directions. The second term causes the different components of the permittivity to oscillate between $\hat{\varepsilon}_{11}$ and $\hat{\varepsilon}_{22}$ about their average value, and accounts for the orientation of the particles. This again agrees with the simulations, with each component achieving its maxima and minima when the particles are aligned parallel or perpendicular to the relevant axis. Moreover, we can immediately see that the magnitude of these particle induced effects is proportional to the volume of the particles as previously observed, and in the limit of zero particle volume we are simply left with ε_m, as you would expect. We can also see how the anisotropy of the particles' shapes directly influences the anisotropy of the permittivity. For a circular particle where $a = b$, we have that $\hat{\varepsilon}_{11} = \hat{\varepsilon}_{22}$ and the permittivity is isotropic, agreeing with the symmetries of circular particles. At the other end of the scale, it can be seen that when one of the semi-axes is much larger than the other, one of the $\hat{\varepsilon}_{ii}$ go to zero and the magnitude of the resulting anisotropy becomes much greater. This is counterbalanced by the dependence on the particle volume, however, and we observe the previous results where a greater aspect ratio results in greater anisotropy, up until the point the particle becomes too thin and has too little volume.

The dependence on the permittivities of the medium and particles is also very evident. Firstly, we can see that the effective permittivity is dependent on the difference between the two permittivities, as expected. In the case of $\varepsilon_m = \varepsilon_p$ (an isotropic medium from an electric fields point of view) the anisotropic part vanishes

and we are simply left with the isotropic permittivity $\varepsilon_m = \varepsilon_p$. The magnitude of the anisotropy then increases as the difference in permittivities increases, however, in the limit $\varepsilon_p \gg \varepsilon_m$ we observe the same asymptotic behaviour as before. In this limit (and just considering $\theta = 0$ for simplicity), we have that

$$\begin{aligned}(\varepsilon_{\text{eff}})_{11} &= \varepsilon_m + |\Omega_p| \left(\varepsilon_p - \varepsilon_m - \frac{b\left(\varepsilon_m - \varepsilon_p\right)^2}{a\varepsilon_m + b\varepsilon_p} \right) \\ &= \varepsilon_m \left[1 + |\Omega_p| \left(1 + \frac{a}{b} \right) \right] + O\left(\frac{1}{\varepsilon_p} \right) \end{aligned} \tag{4.5.3.12}$$

after linearising. We similarly have that

$$(\varepsilon_{\text{eff}})_{22} = \varepsilon_m \left[1 + |\Omega_p| \left(1 + \frac{b}{a} \right) \right] + O\left(\frac{1}{\varepsilon_p} \right) \tag{4.5.3.13}$$

and we find that the permittivity behaves asymptotically. As in the simulations, we have that the saturation permittivity is largest along the axis of the major semi-axis. Since we are in the limit of low volume fractions, both of these values are close to ε_m, also agreeing with intuition and the simulations shown in figure 4.4b.

4.5.4 Particles in Anisotropic Medium

4.5.4.1 The Canonical Problem

We now relax the condition that the surrounding medium is isotropic and, as such, the permittivity of the medium becomes a tensor. We return to equations (4.5.1.1.11a)-(4.5.1.1.11c) where we made this simplification. Written out in full, the first order matched asymptotic inner problem is

$$\nabla_{\boldsymbol{Y}}^2 \gamma_p^{(1)} = 0\,, \qquad \boldsymbol{Y} \in \Omega_p(\boldsymbol{x})\,, \tag{4.5.4.1.1a}$$

$$\boldsymbol{\nabla}_{\boldsymbol{Y}} \cdot \left[\varepsilon_m \boldsymbol{\nabla}_{\boldsymbol{Y}} \gamma_m^{(1)} \right] = 0\,, \qquad \boldsymbol{Y} \in \Omega_m(\boldsymbol{x})\,, \tag{4.5.4.1.1b}$$

$$\varepsilon_p \boldsymbol{n}_0 \cdot \boldsymbol{\nabla}_{\boldsymbol{Y}} \gamma_p^{(1)} - \boldsymbol{n}_0 \cdot \left(\varepsilon_m \boldsymbol{\nabla}_{\boldsymbol{Y}} \gamma_m^{(1)} \right) = \boldsymbol{n}_0 \cdot \left(\varepsilon_m - \varepsilon_p \mathcal{I} \right)\,, \quad \boldsymbol{Y} \in \partial\Omega_p(\boldsymbol{x})\,, \tag{4.5.4.1.1c}$$

$$\gamma_m^{(1)} = \gamma_p^{(1)}\,, \qquad \boldsymbol{Y} \in \partial\Omega_p(\boldsymbol{x})\,, \tag{4.5.4.1.1d}$$

$$\gamma_m^{(1)} = const. \qquad |\boldsymbol{Y}| \to \infty. \tag{4.5.4.1.1e}$$

The governing equation inside the medium is not separable, owing to the cross derivative terms that now arise from the tensorial nature of medium's permittivity. We therefore seek a change of coordinates so as to simplify the equations; specifically, we look to put the equation into canonical form.

For a general second order PDE of the form

$$A\frac{\partial^2 f}{\partial y_1^2} + B\frac{\partial^2 f}{\partial y_1 \partial y_2} + C\frac{\partial^2 f}{\partial y_2^2} + \text{lower order terms} = G, \tag{4.5.4.1.2}$$

the discriminant is defined as

$$\Delta = B^2 - 4AC, \tag{4.5.4.1.3}$$

and its sign is invariant under coordinate transformations. The sign of the discriminant classifies the type of PDE and, in our case, where $B^2 - 4AC = -4\varepsilon_{\parallel}\varepsilon_{\perp} < 0$, the problem is elliptic. For every elliptic PDE there exists a coordinate transformation $\overline{y}_1(y_1, y_2), \overline{y}_2(y_1, y_2)$ such that

$$\left(\frac{\partial^2}{\partial \overline{y}_1^2} + \frac{\partial^2}{\partial \overline{y}_2^2}\right)\overline{f} = \overline{G}, \tag{4.5.4.1.4}$$

and all cross-derivative terms are removed. We therefore look to put our anisotropic problem into canonical form, which we can solve easily using the previous results for isotropic media.

The method of finding this coordinate change starts by substituting the new coordinates $\overline{y}_1$ and $\overline{y}_2$ into the original PDE to find

$$a\frac{\partial^2 f}{\partial \overline{y}_1^2} + b\frac{\partial^2 \overline{f}}{\partial \overline{y}_1 \partial \overline{y}_2} + c\frac{\partial^2 \overline{f}}{\partial \overline{y}_2^2} + \text{lower order terms} = \overline{G}. \tag{4.5.4.1.5}$$

We require $a = c$ and $b = 0$, from which a quadratic relation between the old and new coordinates can be derived. We use this relation to look for characteristics of the new coordinates (lines along which they are constant), expressed as ODEs in terms of the old coordinates. Solving these ODEs then gives the required coordinate change.

Before implementing this method, we can first save ourselves a bit of trouble with a simple rotation of coordinates. For the method of homogenisation to be valid, ε_m must be uniform throughout the unit cell, and therefore has vanishing $\boldsymbol{y}$ derivatives. Considering equation (4.5.4.1.1b) in a reference frame such that $\varepsilon_m = \begin{pmatrix} \varepsilon_{\parallel} & 0 \\ 0 & \varepsilon_{\perp} \end{pmatrix}$ (where $\varepsilon_{\parallel}$ and $\varepsilon_{\perp}$ are the permittivities parallel and perpendicular to the optical axis of the medium), the equation then has the form

$$\left(\varepsilon_{\parallel}\frac{\partial^2}{\partial y_1^2} + \varepsilon_{\perp}\frac{\partial^2}{\partial y_2^2}\right)\gamma_m^{(1)} = 0. \tag{4.5.4.1.6}$$

Denoting $\gamma_m^{(1)}$ by γ and ε_m by ε for brevity, we substitute our change of variables $\overline{y}_1(y_1, y_2), \overline{y}_2(y_1, y_2)$ into equation (4.5.4.1.1b), giving

$$b = 2\varepsilon_{11}\overline{y}_{2,1}\overline{y}_{1,1} + 2\varepsilon_{12}\left(\overline{y}_{2,1}\overline{y}_{1,2} + \overline{y}_{2,2}\overline{y}_{1,1}\right) + 2\varepsilon_{22}\overline{y}_{2,2}\overline{y}_{1,1}, \tag{4.5.4.1.7}$$

$$a = \varepsilon_{11}\overline{y}_{2,1}^2 + 2\varepsilon_{12}\overline{y}_{2,1}\overline{y}_{2,2} + \varepsilon_{22}\overline{y}_{2,2}^2, \tag{4.5.4.1.8}$$

and

$$c = \varepsilon_{11}\overline{y}_{1,1}^2 + 2\varepsilon_{12}\overline{y}_{1,1}\overline{y}_{1,2} + \varepsilon_{22}\overline{y}_{1,2}^2, \tag{4.5.4.1.9}$$

where a subscript comma denotes partial differentiation, i.e., $\overline{y}_{i,j} = \frac{\partial \overline{y}_i}{\partial y_j}$. We now enforce that $b = 0$ and $a = c$. If this holds true then we can write

$$b + (a - c)\, i = 0, \tag{4.5.4.1.10}$$

where i is the imaginary unit. This can be rewritten as

$$\varepsilon_{11}\left(\overline{y}_{2,1} + i\overline{y}_{1,1}\right)^2 + 2\varepsilon_{12}\left(\overline{y}_{2,1} + i\overline{y}_{1,1}\right)\left(\overline{y}_{2,2} + i\overline{y}_{1,2}\right) + \varepsilon_{22}\left(\overline{y}_{2,2} + i\overline{y}_{1,2}\right)^2 = 0, \tag{4.5.4.1.11}$$

or, by writing

$$\alpha = \overline{y}_2 + i\overline{y}_1, \tag{4.5.4.1.12}$$

we have

$$\varepsilon_{11}\alpha_{,1}^2 + 2\varepsilon_{12}\alpha_{,1}\alpha_{,2} + \varepsilon_{22}\alpha_{,2}^2 = 0. \tag{4.5.4.1.13}$$

We solve the quadratic to find the two complex conjugate solutions

$$\begin{aligned}\alpha_{,1} &= \frac{-\varepsilon_{12} \pm i\sqrt{\varepsilon_{11}\varepsilon_{22} - \varepsilon_{12}^2}}{\varepsilon_{11}} \\ &= -\tau \pm i\sigma.\end{aligned} \tag{4.5.4.1.14}$$

Along the coordinate lines $\alpha = const.$ we have

$$d\alpha = \frac{\partial \alpha}{\partial y_1}\partial y_1 + \frac{\partial \alpha}{\partial y_2}\partial y_2 = 0, \tag{4.5.4.1.15}$$

from which it follows that such lines are described by

$$\frac{dy_2}{dy_1} = \frac{\alpha_{,1}}{\alpha_{,2}} = -\tau \pm i\sigma. \tag{4.5.4.1.16}$$

This gives the two families of solutions

$$y_2 = (-\tau \pm i\sigma)\, y_1 + c, \tag{4.5.4.1.17}$$

where c is an integration constant, and along which $\alpha = c$. We obtain our desired coordinate transformation from (4.5.4.1.12) by

$$\overline{y}_2 = \frac{\alpha + \alpha^*}{2} = y_2 - \tau y_1 \tag{4.5.4.1.18a}$$

and

$$\overline{y}_1 = \frac{\alpha - \alpha^*}{2i} = \sigma y_1. \tag{4.5.4.1.18b}$$

To check that this is correct we substitute into the original equation, giving

$$0 = \varepsilon_{11}\left(\tau^2\overline{\gamma}_{ww} - 2\tau\sigma\overline{\gamma}_{wz} + \sigma^2\overline{\gamma}_{zz}\right) + 2\varepsilon_{12}\left(-\tau\overline{\gamma}_{ww} + \sigma\overline{\gamma}_{zw}\right) + \varepsilon_{22}\overline{\gamma}_{ww}. \tag{4.5.4.1.19}$$

After dividing by ε_{11} and collecting like terms this becomes

$$0 = \left(\frac{\varepsilon_{22}}{\varepsilon_{11}} - \tau^2\right)\overline{\gamma}_{ww} + \sigma^2\overline{\gamma}_{zz}, \tag{4.5.4.1.20}$$

which, since $\frac{\varepsilon_{22}}{\varepsilon_{11}} - \tau^2 = \sigma^2$, is the Laplace equation

$$\overline{\nabla}^2\overline{\gamma} = 0, \tag{4.5.4.1.21}$$

as desired. We therefore define our coordinate transformation T matrix as

$$\begin{pmatrix} \overline{y}_1 \\ \overline{y}_2 \end{pmatrix} = \mathrm{T}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \tag{4.5.4.1.22}$$

where

$$\mathrm{T} = \begin{pmatrix} \sigma & 0 \\ -\tau & 1 \end{pmatrix}. \tag{4.5.4.1.23}$$

We now have a coordinate system in which the permittivity of the medium is isotropic and the problem inside the medium is separable. We cannot, however, simply solve the entire problem by operating in this system. Whilst the problem in the particle is isotropic in the original cartesian system, after the coordinate change (4.5.4.1.23) it will be the anisotropic problem

$$\overline{\boldsymbol{\nabla}} \cdot \left(\mathrm{T}^T\overline{\boldsymbol{\nabla}}\gamma_p^{(1)}\right) = 0. \tag{4.5.4.1.24}$$

Due to T having off-diagonal components this problem is then also not separable. As a result, we will need to find a method in which we can solve the problem inside the particle in the original cartesian system, map this solution to the barred coordinate system (in which we can solve the problem in the medium), and then apply the boundary conditions to solve the complete problem. Initially, we do not want to have to deal with this mapping and matching whilst we check the feasibility of using this coordinate change to solve the problem in the medium. The next step, then, is to solve the problem in the limit that ε_p is very large, equivalent to our particle being metallic, and there is no field inside the particle.

4.5.4.2 Metallic Particles in Anisotropic Medium

We consider the cell problem for a metallic particle in an anisotropic medium:

$$\boldsymbol{\nabla} \cdot (\varepsilon_m \boldsymbol{\nabla} \boldsymbol{\gamma}) = 0\,, \quad \boldsymbol{y} \in \Omega_m, \tag{4.5.4.2.1a}$$

$$\boldsymbol{n} \cdot (\varepsilon_m \boldsymbol{\nabla} \boldsymbol{\gamma}) = \boldsymbol{n}\,, \quad \boldsymbol{y} \in \partial\Omega_p, \tag{4.5.4.2.1b}$$

$$\boldsymbol{\gamma} = const. \qquad |\boldsymbol{y}| \to \infty. \tag{4.5.4.2.1c}$$

We know the governing equation becomes (4.5.4.1.21) after making the coordinate change (4.5.4.1.22), and so we now look to find how the boundary condition transforms. From here we denote $\varepsilon_m = \varepsilon$ so as not to confuse indices.

To begin, we note that the left hand side of (4.5.4.2.1b) can be written as $(\varepsilon \boldsymbol{n}) \cdot \boldsymbol{\nabla} \boldsymbol{\gamma} = \boldsymbol{N} \cdot \boldsymbol{\nabla} \boldsymbol{\gamma}$, since ε is symmetric. Next, we use the fact that a tensor contraction is invariant under a coordinate transformation, that is, for two arbitrary tensors W and V being contracted over an index,

$$W^i V_i = \frac{\partial y^i}{\partial \overline{y}^j} \overline{W}^j \frac{\partial \overline{y}^k}{\partial y^i} \overline{V}_k = \delta^k_j \overline{W}^k \overline{V}_j = \overline{W}^k \overline{V}_k. \tag{4.5.4.2.2}$$

From this it follows that

$$\boldsymbol{N} \cdot \boldsymbol{\nabla} \boldsymbol{\gamma} = \overline{\boldsymbol{N}} \cdot \overline{\boldsymbol{\nabla}} \boldsymbol{\gamma} \tag{4.5.4.2.3}$$

and we need to identify $\overline{\boldsymbol{N}}$. By first calculating that

$$\overline{\varepsilon} = \sigma^2 \varepsilon_{11} \mathcal{I} = \epsilon \mathcal{I}, \tag{4.5.4.2.4}$$

we again use the invariance of a contraction to find

$$\overline{N}^i = \overline{\varepsilon}^{ij} n_j = \overline{\varepsilon}^{ij} \overline{n}_j = \epsilon \overline{n}^i; \tag{4.5.4.2.5}$$

and so this now leaves the question as to what is $\overline{\boldsymbol{n}}$. By considering the level set, which is a scalar field and, hence, invariant under coordinate changes, we have that

$$\frac{\partial \overline{\chi}}{\partial \overline{y}^i} = \frac{\partial \chi}{\partial \overline{y}^i} = \frac{\partial y^j}{\partial \overline{y}^i} \frac{\partial \chi}{\partial y^j} = \frac{\partial y^j}{\partial \overline{y}^i} n_j = \overline{n}_i, \tag{4.5.4.2.6}$$

and so $\overline{\boldsymbol{n}}$ and N are in the direction of the normal to the particle boundary after the change of coordinates.

To find the normal, we substitute the coordinate change (4.5.4.1.22) into the equation for the ellipse. We consider, for now, a particle that is aligned along the y_1 axis with major and minor semi-axes $a(\boldsymbol{x})$ and $b(\boldsymbol{x})$. This gives

$$\left(\frac{\overline{y}_1}{a\sigma}\right)^2 + \left(\frac{\tau \overline{y}_1 + \sigma \overline{y}_2}{b\sigma}\right)^2 = 1, \tag{4.5.4.2.7}$$

which is the ellipse equation

$$\begin{aligned} 1 &= \left(\frac{b^2+a^2\tau^2}{(ab\sigma)^2}\right)\overline{y}_1^2 + \frac{2\tau}{\sigma b^2}\overline{y}_1\overline{y}_2 + \frac{\overline{y}_2^2}{b^2} \\ &= (\overline{y}_1, \overline{y}_2)\begin{pmatrix} \frac{b^2+a^2\tau^2}{(ab\sigma)^2} & \frac{\tau}{\sigma b^2} \\ \frac{\tau}{\sigma b^2} & \frac{1}{b^2} \end{pmatrix}\begin{pmatrix} \overline{y}_1 \\ \overline{y}_2 \end{pmatrix} \\ &= (\overline{y}_1, \overline{y}_2)\,\overline{\mathrm{Q}}\begin{pmatrix} \overline{y}_1 \\ \overline{y}_2 \end{pmatrix}. \end{aligned} \tag{4.5.4.2.8}$$

The matrix $\overline{\mathrm{Q}}$ is symmetric and has positive components; by the Principal Axis Theorem it has two distinct, positive eigenvalues, corresponding to two orthogonal eigenvectors. As it is written in (4.5.4.2.8), $\overline{\mathrm{Q}}$ is a linear transformation acting on the unit circle. This means we have two orthogonal directions in which the unit circle is stretched, with the result being an ellipse; and so the eigenvectors of $\overline{\mathrm{Q}}$ must give the directions of the semi-axes. Calling the eigenvectors $\boldsymbol{v}_i$, for the ellipse equation (4.5.4.2.8) to be satisfied it follows that

$$\boldsymbol{v}_i^T\overline{\mathrm{Q}}\boldsymbol{v}_i = \boldsymbol{v}_i^T\lambda_i\boldsymbol{v}_i = \lambda_i\left|\boldsymbol{v}_i\right|^2 = 1, \tag{4.5.4.2.9}$$

where λ_i is eigenvalue corresponding to $\boldsymbol{v}_i$. Thus, for the semi-axis described by $\boldsymbol{v}_i$, the length of the semi-axis is given by $|\boldsymbol{v}_i| = \sqrt{\frac{1}{\lambda_i}}$. The major and minor semi-axes, $\boldsymbol{v}^{(a)}$ and $\boldsymbol{v}^{(b)}$, are then given by

$$\left|\boldsymbol{v}^{(a)}\right| = \overline{a} = \sqrt{\frac{1}{\lambda_a}}, \tag{4.5.4.2.10}$$

$$\left|\boldsymbol{v}^{(b)}\right| = \overline{b} = \sqrt{\frac{1}{\lambda_b}}, \tag{4.5.4.2.11}$$

$$\lambda_a < \lambda_b. \tag{4.5.4.2.12}$$

We choose the orientation of $\boldsymbol{v}^{(a)}$ to have a positive $\overline{y}_1$ component and, denoting the angle made between the particle's major axis and the $\overline{y}_1$ axis as θ, we can then say that $\cos\theta = \sqrt{\lambda_a}v_1^{(a)}$ and $\sin\theta = \sqrt{\lambda_a}v_2^{(a)}$. This is all shown in figure (4.14).

We can now define a second change of coordinates

$$\begin{pmatrix} \tilde{y}_1 \\ \tilde{y}_2 \end{pmatrix} = \sqrt{\lambda_a}\begin{pmatrix} v_1^{(a)} & v_2^{(a)} \\ -v_2^{(a)} & v_1^{(a)} \end{pmatrix}\begin{pmatrix} \overline{y}_1 \\ \overline{y}_2 \end{pmatrix} = \mathrm{R}^T(\theta)\begin{pmatrix} \overline{y}_1 \\ \overline{y}_2 \end{pmatrix}, \tag{4.5.4.2.13}$$

such that the particle is then lying along the $\tilde{y}_1$ axis. The major and minor semi-axes in this new coordinate system are $\tilde{a} = \overline{a}$ and $\tilde{b} = \overline{b}$, and we can introduce an elliptic

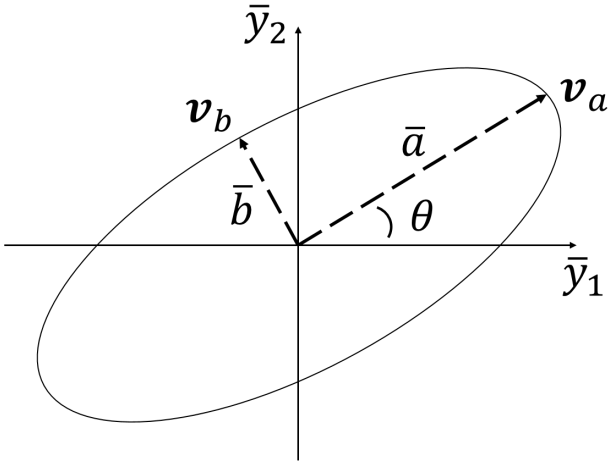


FIGURE 4.14: A diagram of the particle shape in the $\overline{\boldsymbol{y}}$ coordinate system. The major and minor semi-axes, $\boldsymbol{v}^{(a)}$ and $\boldsymbol{v}^{(b)}$, have lengths $\overline{a}$ and $\overline{b}$, and the particle angle is parameterised by θ.

coordinate system

$$\begin{pmatrix} \tilde{y}_1 \\ \tilde{y}_2 \end{pmatrix} = \begin{pmatrix} \tilde{F}\cosh\tilde{\mu}\cos\tilde{\nu} \\ \tilde{F}\sinh\tilde{\mu}\sin\tilde{\nu} \end{pmatrix}, \tag{4.5.4.2.14}$$

where

$$\tilde{F} = \sqrt{\overline{a}^2 - \overline{b}^2} \tag{4.5.4.2.15}$$

and the particle boundary is given by

$$\tilde{\mu} = \tilde{\mu}_0 = \frac{1}{2}\ln\left(\frac{\overline{a}+\overline{b}}{\overline{a}-\overline{b}}\right). \tag{4.5.4.2.16}$$

We are almost in a position to solve the problem, we just need to transform the right hand side of the boundary condition into the $\tilde{\mu}$, $\tilde{\nu}$ coordinate system. The normal has components $\frac{\partial\chi}{\partial y^i} = n_i$, and so these must transform covariantly. That is, we express the right hand side of the boundary condition as $n_i = \frac{\partial\tilde{y}^j}{\partial y^i}\tilde{n}_j$. This is a linear combination of the components of the normal to the transformed particle boundary, with coefficients given by the components of the forward coordinate transformation. The final form of the problem that we must solve is then

$$\tilde{\nabla}^2\boldsymbol{\gamma} = 0\,, \qquad \tilde{\boldsymbol{y}} \in \overline{\Omega}_m, \tag{4.5.4.2.17a}$$

$$\epsilon\frac{\partial\boldsymbol{\gamma}}{\partial\tilde{\mu}} = S^T\begin{pmatrix} \overline{b}\cos\tilde{\nu} \\ \overline{a}\sin\tilde{\nu} \end{pmatrix}, \qquad \tilde{\boldsymbol{y}} \in \partial\overline{\Omega}_p, \tag{4.5.4.2.17b}$$

$$\boldsymbol{\gamma} = const. \qquad |\tilde{\boldsymbol{y}}| \to \infty. \tag{4.5.4.2.17c}$$

Here we have defined the composite, forward coordinate transformation

$$\mathrm{S} = \mathrm{R}\,(\theta)\,\mathrm{T} \tag{4.5.4.2.18}$$

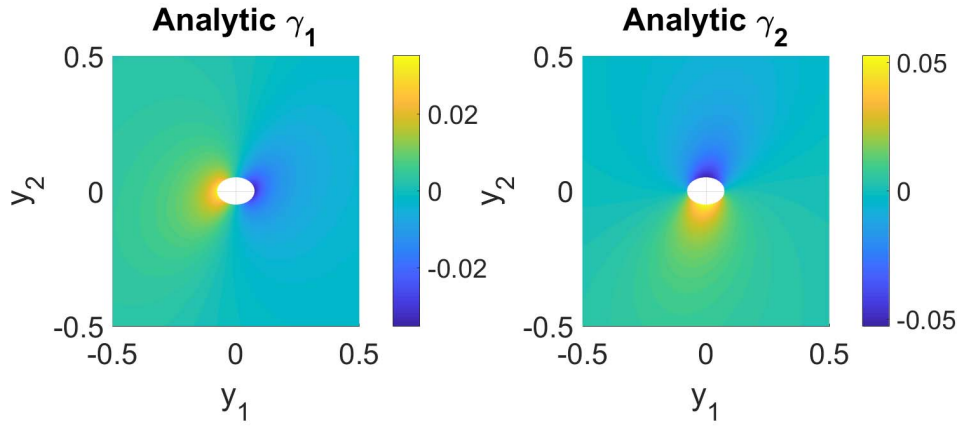


FIGURE 4.15: Analytic solution for a metallic particle in an anisotropic medium with axis at $\pi/3$ to the y_1 axis. $a = 0.075$, $b = 0.05$, and $\varepsilon_m = \mathrm{R}(\pi/3) \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \mathrm{R}^T(\pi/3)$.

that takes us from the $\boldsymbol{y}$ to $\tilde{\boldsymbol{y}}$ systems.

This problem is very simple to solve, as before we must have that the $\tilde{\mu}$ dependence is exponential and the $\tilde{v}$ dependence is 2π periodic. Given the form of the right hand side of the boundary condition, we make the ansatz

$$\gamma_i = \frac{1}{\epsilon}\left(A_i e^{\tilde{\mu}} + B_i e^{-\tilde{\mu}}\right) \mathrm{S}_{ji}\tilde{n}_j. \tag{4.5.4.2.19}$$

Matching at infinity, we must have that $A_i = 0$, or else the leading order solution cannot be constant. Applying the boundary condition (4.5.4.2.17b), the $\tilde{v}$ dependence cancels out from both sides and it follows that

$$B_i = -e^{\tilde{\mu}_0} = -\sqrt{\frac{\overline{a}+\overline{b}}{\overline{a}-\overline{b}}}. \tag{4.5.4.2.20}$$

Putting these pieces together, the solution for a metallic particle in an anisotropic medium is then given by

$$\gamma = -\frac{1}{\epsilon}\sqrt{\frac{\overline{a}+\overline{b}}{\overline{a}-\overline{b}}} e^{-\tilde{\mu}} \mathrm{S}^T \begin{pmatrix} \overline{b}\cos\tilde{v} \\ \overline{a}\sin\tilde{v} \end{pmatrix}. \tag{4.5.4.2.21}$$

Figures 4.15 and 4.16 show the analytic solutions and their comparison with the finite elements solution provided by Comsol. We can safely conclude that the coordinate transformations and method that has been employed are effective in providing an accurate analytic solution for the low volume fraction approximation. Not only is it accurate, but, in the case of metallic particles at least, the method actually makes the problem rather simple to solve. However, the problem that we are trying to solve involves dielectric particles, and we are still left with the aforementioned problem of

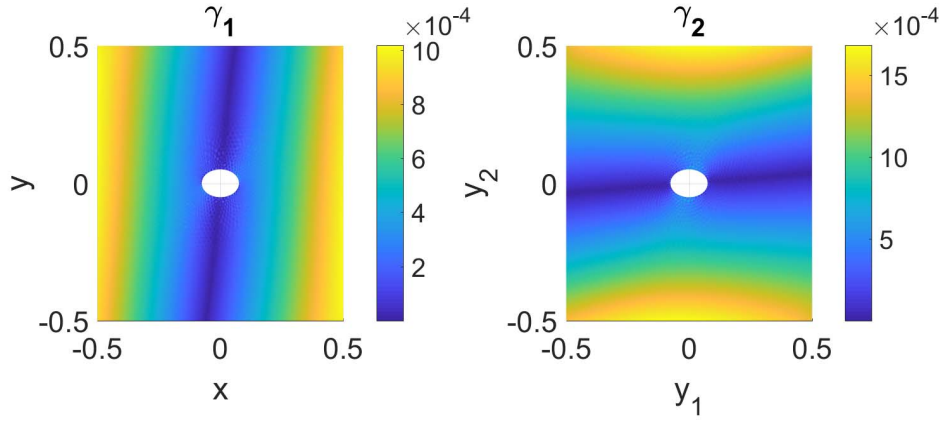


FIGURE 4.16: Norm of the difference between the analytic solution and the finite elements solution from Comsol. Solution is for a metallic particle in an anisotropic medium with axis at $\pi/3$ to the y_1 axis. $a = 0.075$, $b = 0.05$, and $\varepsilon_m = \mathrm{R}(\pi/3)\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\mathrm{R}^T(\pi/3)$.

matching the solutions inside and outside of the particle. As it turns out, we do not need to solve this problem. We can, in effect, avoid it entirely, using these metallic results for the easy solution of the full dielectric problem.

4.5.4.3 Dielectric Particles in Anisotropic Medium

The simple solution of the dielectric problem arises from the simplicity of the solution inside the particle. We first solve the governing equation inside the particle in the elliptic coordinate system (equation (4.5.1.2.10a)). We must have that the general solution is of the form

$$\left(\gamma_p^{(1)}\right)_i = A\cosh\mu\cos\nu + B\sinh\mu\sin\nu + C\cosh\mu\sin\nu + D\sinh\mu\cos\nu, \quad (4.5.4.3.1)$$

where we assume from the boundary condition (4.5.1.2.10b) that there is no dependence on higher harmonics of ν. If we again apply our regularity conditions

$$\left(\gamma_p^{(1)}\right)_i\Big|_{(\mu,\nu)=(0,\nu)} = \left(\gamma_p^{(1)}\right)_i\Big|_{(\mu,\nu)=(0,-\nu)} \quad (4.5.4.3.2)$$

and

$$\left.\frac{\partial\gamma_{p1}^{(1)}}{\partial\mu}\right|_{(\mu,\nu)=(0,\nu)} = -\left.\frac{\partial\gamma_{p1}^{(1)}}{\partial\mu}\right|_{(\mu,\nu)=(0,-\nu)}, \quad (4.5.4.3.3)$$

we find that we must have $C = D = 0$. We therefore make the ansatz that

$$\gamma_p^{(1)} = \begin{pmatrix} A_{p1} & B_{p1} \\ A_{p2} & B_{p2} \end{pmatrix}\begin{pmatrix} F\cosh\mu\cos\nu \\ F\sinh\mu\sin\nu \end{pmatrix}$$

$$= \mathrm{C}\boldsymbol{y}. \tag{4.5.4.3.4}$$

The normal derivative of $\gamma_p^{(1)}$ is then $n_i \partial_i \mathrm{C}_{jk} y_k = n_i \mathrm{C}_{ji}$, and we use this to rewrite the dielectric problem in equations (4.5.1.1.11a)-(4.5.1.1.11c) as

$$\boldsymbol{\nabla}_{\boldsymbol{y}} \cdot \left(\varepsilon_m \boldsymbol{\nabla}_{\boldsymbol{Y}} \gamma_m^{(1)} \right) = 0 \,, \qquad \boldsymbol{y} \in \Omega_m \left(\boldsymbol{x} \right) \tag{4.5.4.3.5}$$

$$\boldsymbol{n} \cdot \left(\varepsilon_m \boldsymbol{\nabla} \gamma_m^{(1)} \right) = \boldsymbol{n} \cdot \left[\varepsilon_p \left(\mathcal{I} + \mathrm{C}^T \right) - \varepsilon_m \right] \,, \quad \boldsymbol{y} \in \partial\Omega_p \left(\boldsymbol{x} \right) \tag{4.5.4.3.6}$$

$$\gamma_m^{(1)} = \gamma_p^{(1)} \,, \qquad \boldsymbol{y} \in \partial\Omega_p \left(\boldsymbol{x} \right). \tag{4.5.4.3.7}$$

$$\gamma_m^{(1)} = const. \text{ as } |\boldsymbol{y}| \to \infty \tag{4.5.4.3.8}$$

This is very similar to the problem for metallic particles, with the right hand side of the boundary condition (4.5.4.3.6) being a linear combination of the components of $\boldsymbol{n}$, as opposed to just $\boldsymbol{n}$. If we therefore take a matching linear combination of the solutions given in (4.5.4.2.21), we will have a solution that satisfies both the governing equation and the new boundary condition, as desired. Defining

$$\mathrm{D} = \left[\varepsilon_p \left(\mathcal{I} + \mathrm{C}^T \right) - \varepsilon_m \right], \tag{4.5.4.3.9}$$

we then have that the solution for a dielectric particle in an anisotropic medium is given by

$$\gamma_m^{(1)} = \mathrm{D}^T \gamma, \tag{4.5.4.3.10}$$

where γ is the solution for a metallic particle in an anisotropic medium from the previous section. This solution is now dependent only on the components of C, which are fixed using the continuity condition (4.5.4.3.7). We rewrite the solution in the particle as

$$\gamma_p^{(1)} = \mathrm{C}\mathrm{S}^{-1} \tilde{\boldsymbol{y}} \tag{4.5.4.3.11}$$

and the solution in the medium as

$$\gamma_m^{(1)} = -\frac{1}{\epsilon} \sqrt{\frac{\overline{a} + \overline{b}}{\overline{a} - \overline{b}}} e^{-\tilde{\mu}} \mathrm{D}^T \mathrm{S}^T \begin{pmatrix} \overline{b} \cos \tilde{\nu} \\ \overline{a} \sin \tilde{\nu} \end{pmatrix} \tag{4.5.4.3.12}$$

$$= -\frac{1}{\epsilon} \sqrt{\frac{\overline{a} + \overline{b}}{\overline{a} - \overline{b}}} e^{-\tilde{\mu}} \mathrm{D}^T \mathrm{S}^T \begin{pmatrix} \frac{\overline{b}}{\overline{a}} & 0 \\ 0 & \frac{\overline{a}}{\overline{b}} \end{pmatrix} \begin{pmatrix} \overline{a} \cos \tilde{\nu} \\ \overline{b} \sin \tilde{\nu} \end{pmatrix} \tag{4.5.4.3.13}$$

$$= -\frac{1}{\epsilon} \sqrt{\frac{\overline{a} + \overline{b}}{\overline{a} - \overline{b}}} e^{-\tilde{\mu}} \mathrm{D}^T \mathrm{S}^T \mathrm{K} \tilde{\boldsymbol{y}}_0, \tag{4.5.4.3.14}$$

where $\tilde{\boldsymbol{y}}_0$ is $\tilde{\boldsymbol{y}}$ evaluated at $\tilde{\mu} = \tilde{\mu}_0$. Imposing continuity at the particle boundary then gives

$$\mathrm{C}\mathrm{S}^{-1} = -\frac{1}{\epsilon} \mathrm{D}^T \mathrm{S}^T \mathrm{K}. \tag{4.5.4.3.15}$$

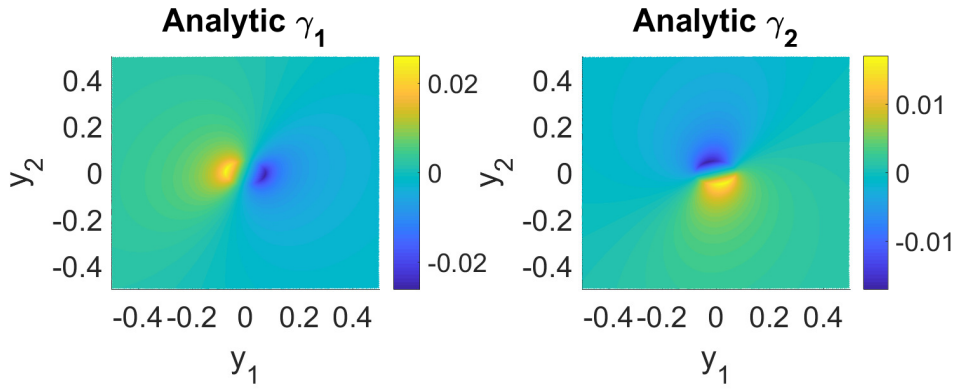


FIGURE 4.17: Analytic solution for a dielectric particle in an anisotropic medium with axis at $\pi/3$ to the y_1 axis. $a = 0.075$, $b = 0.05$, and $\varepsilon_m = \mathrm{R}(\pi/3)\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\mathrm{R}^T(\pi/3)$.

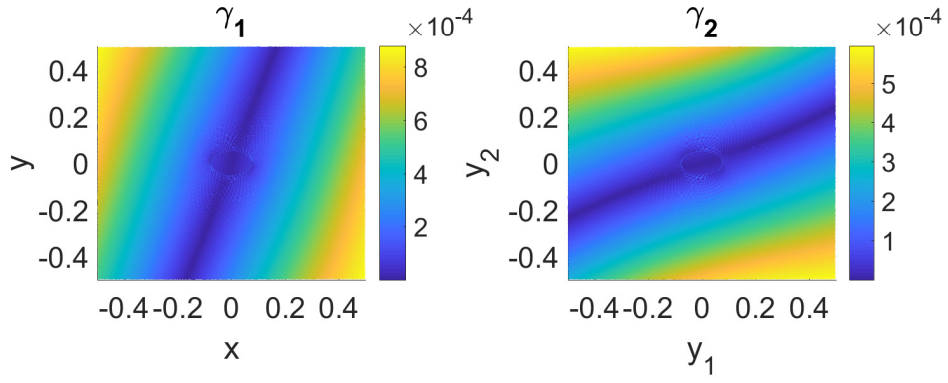


FIGURE 4.18: Norm of the difference between the analytic solution and the finite elements solution from Comsol. Solution is for a dielectric particle in an anisotropic medium with axis at $\pi/3$ to the y_1 axis. $a = 0.075$, $b = 0.05$, and $\varepsilon_m = \mathrm{R}(\pi/3)\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\mathrm{R}^T(\pi/3)$.

The matrix D is a function of C and so we rearrange this equation to find

$$\mathrm{C} = \left(\varepsilon_m - \varepsilon_p\mathcal{I}\right)\left(\epsilon\mathrm{S}^{-1}\mathrm{K}^{-1}\mathrm{S}^{-T} + \varepsilon_p\mathcal{I}\right)^{-1}. \tag{4.5.4.3.16}$$

This is highly reminiscent of the factors of the form $\left(\varepsilon_m - \varepsilon_p\right)\left(a\varepsilon_m + b\varepsilon_p\right)^{-1}$ from the case of an isotropic medium (given that the components of K are ratios of $\bar{a}$ and $\bar{b}$). In fact, in the case that ε_m is isotropic, we would have that $\epsilon = \varepsilon_m$ and $\mathrm{S} = \mathcal{I}$, and the C_1 component then becomes $\left(\varepsilon_m - \varepsilon_p\right)\left(\bar{a}\varepsilon_m + \bar{b}\varepsilon_p\right)^{-1}$. This is in exact agreement with the isotropic solution since $a = \bar{a}$ and $b = \bar{b}$ when S $=\mathcal{I}$. The analytic solution for a particle in an anisotropic medium is shown in figures 4.17 and 4.18. We also again see

excellent agreement between the solutions in the neighbourhood of the particle, which is to be expected given that it is just a linear combination of the earlier solutions.

4.5.5 Effective Permittivity of Dielectric Particles in Anistropic Medium

We now have a solution for $\boldsymbol{\gamma}_p^{(1)}$ given a dielectric particle that is aligned along the y_1 axis and surrounded by an anisotropic medium with an arbitrary permittivity tensor. What remains is to use this solution to calculate the laboratory frame effective permittivity for a particle at an angle θ in a surrounding medium with optical axis making an angle ψ to the x_1 axis. Calling the extraordinary and ordinary permittivities of the medium $\varepsilon_{\parallel}$ and $\varepsilon_{\perp}$, it follows that

$$\varepsilon_m = \frac{\varepsilon_{\parallel} + \varepsilon_{\perp}}{2}\mathcal{I} + \frac{\varepsilon_{\parallel} - \varepsilon_{\perp}}{2}\begin{pmatrix} \cos 2\psi & \sin 2\psi \\ \sin 2\psi & -\cos 2\psi \end{pmatrix}. \tag{4.5.5.1}$$

To compute the effective permittivity, we use a rotated system (denoted by a $\check{}$) in which the particle is aligned along the first axis ($\check{\theta} = 0$). The medium will then be described by

$$\check{\varepsilon}_m = \mathrm{R}(\theta)\,\varepsilon_m \mathrm{R}^T(\theta). \tag{4.5.5.2}$$

In this system we know how to calculate the solution, it is given by

$$\check{\varepsilon}_{\text{eff}} = \sum_{\alpha=m,p} \check{\varepsilon}_\alpha \int_{\Omega_\alpha} \mathcal{I} dV + \left(\varepsilon_p \mathcal{I} - \check{\varepsilon}_m\right) \int_{\partial\check{\Omega}_p(\boldsymbol{x})} \check{\boldsymbol{n}} \otimes \check{\boldsymbol{\gamma}}_p^{(1)} dS, \tag{4.5.5.3}$$

where

$$\check{\boldsymbol{\gamma}}_p^{(1)} = \left(\check{\varepsilon}_m - \varepsilon_p \mathcal{I}\right)\left(\check{\epsilon}\check{\mathrm{S}}^{-1}\check{\mathrm{K}}^{-1}\check{\mathrm{S}}^{-T} + \varepsilon_p \mathcal{I}\right)^{-1} \check{\boldsymbol{y}}. \tag{4.5.5.4}$$

Here, $\check{\epsilon}$, $\check{\mathrm{S}}$, and $\check{\mathrm{K}}$ are derived from $\check{\varepsilon}_m$ in the exact same manner as their un-accented counterparts in the previous section. The second term of the effective permittivity can be written as

$$-\left(\varepsilon_p \mathcal{I} - \check{\varepsilon}_m\right)^2 \left(\tilde{\epsilon}\tilde{\mathrm{S}}^{-1}\tilde{\mathrm{K}}^{-1}\tilde{\mathrm{S}}^{-T} + \varepsilon_p \mathcal{I}\right)^{-1} \left[\int_0^{2\pi} \begin{pmatrix} \overline{b}\cos\nu \\ \overline{a}\sin\nu \end{pmatrix} \otimes \begin{pmatrix} \overline{a}\cos\nu \\ \overline{b}\sin\nu \end{pmatrix} d\nu\right]^T, \tag{4.5.5.5}$$

which, after completing the integral, becomes

$$-\left|\Omega_p\right|\left(\varepsilon_p \mathcal{I} - \check{\varepsilon}_m\right)^2 \left(\check{\epsilon}\check{\mathrm{S}}^{-1}\check{\mathrm{K}}^{-1}\check{\mathrm{S}}^{-T} + \varepsilon_p \mathcal{I}\right)^{-1}. \tag{4.5.5.6}$$

We can now put the solution into the laboratory frame and bring all the neccessary definitions into one place. The final solution for the effective permittivity of spatially varying ellipsoidal particles (with major and minor semi-axes $a(\boldsymbol{x})$ and $b(\boldsymbol{x})$, and an angle θ between the major and x_1 axes) in a uniaxial, anisotropic medium (with an optical axis at an angle ψ to the x_1 axis, and principal permittivities $\varepsilon_{\parallel}$ and $\varepsilon_{\perp}$) is given

by:

$$\varepsilon_{\text{eff}} = \mathrm{R}^{T}(\theta)\,\check{\varepsilon}_{\text{eff}}\mathrm{R}(\theta), \tag{4.5.5.7a}$$

where

$$\check{\varepsilon}_{\text{eff}} = \left[\left(\varepsilon_p\mathcal{I} - \check{\varepsilon}_m\right)\left|\Omega_p\right| + \check{\varepsilon}_m\right] - \left|\Omega_p\right|\left(\varepsilon_p\mathcal{I} - \check{\varepsilon}_m\right)^2\left(\check{\epsilon}\check{\mathrm{S}}^{-1}\check{\mathrm{K}}^{-1}\check{\mathrm{S}}^{-T} + \varepsilon_p\mathcal{I}\right)^{-1}, \tag{4.5.5.7b}$$

$$\check{\varepsilon}_m = \mathrm{R}^{-1}(\theta)\,\varepsilon_m\mathrm{R}^{-T}(\theta), \tag{4.5.5.7c}$$

$$\varepsilon_m = \frac{\varepsilon_{\parallel} + \varepsilon_{\perp}}{2}\mathcal{I} + \frac{\varepsilon_{\parallel} - \varepsilon_{\perp}}{2}\begin{pmatrix} \cos 2\psi & \sin 2\psi \\ \sin 2\psi & -\cos 2\psi \end{pmatrix}, \tag{4.5.5.7d}$$

$$\check{\epsilon} = \sigma^2\left(\check{\varepsilon}_m\right)_{11}, \tag{4.5.5.7e}$$

$$\sigma = \frac{\sqrt{\left(\check{\varepsilon}_m\right)_{11}\left(\check{\varepsilon}_m\right)_{22} - \left(\check{\varepsilon}_m\right)_{12}^2}}{\left(\check{\varepsilon}_m\right)_{11}}, \tag{4.5.5.7f}$$

$$\tau = \left(\check{\varepsilon}_m\right)_{12}\left(\check{\varepsilon}_m\right)_{11}^{-1}, \tag{4.5.5.7g}$$

$$\check{\mathrm{K}} = \begin{pmatrix} \overline{b}\overline{a}^{-1} & 0 \\ 0 & \overline{a}\overline{b}^{-1} \end{pmatrix}, \tag{4.5.5.7h}$$

$$\overline{a} = \sqrt{\frac{1}{\lambda_1}}, \tag{4.5.5.7i}$$

$$\overline{b} = \sqrt{\frac{1}{\lambda_2}}, \tag{4.5.5.7j}$$

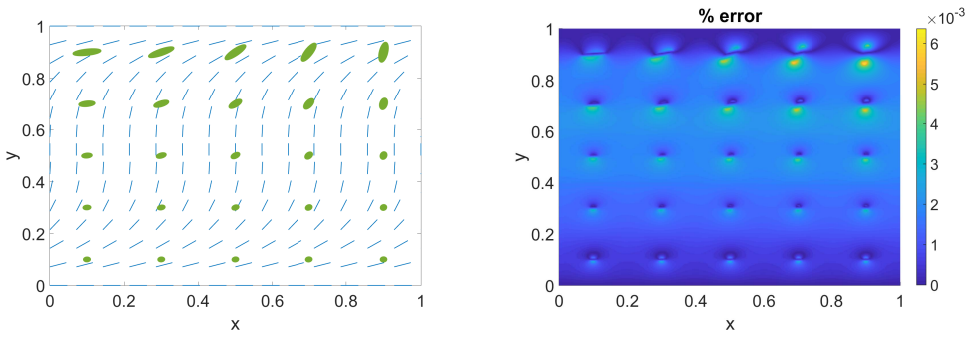
λ_1 and λ_2 are the smallest and largest eigenvalues, respectively, of the matrix

$$\overline{\mathrm{Q}} = \begin{pmatrix} \frac{b^2 + a^2\tau^2}{(ab\sigma)^2} & \frac{\tau}{\sigma b^2} \\ \frac{\tau}{\sigma b^2} & \frac{1}{b^2} \end{pmatrix}, \tag{4.5.5.7k}$$

and

$$\check{\mathrm{S}} = \begin{pmatrix} \sigma v_1 + \tau v_2 & -v_2 \\ \sigma v_2 - \tau v_1 & v_1 \end{pmatrix}. \tag{4.5.5.7l}$$

Figure 4.19 shows a comparison of the homogenised electric potential against the full simulation of the composite, along with the system setup. The optical axis of the medium is described by its angle $\psi = \pi\sin(\pi y)/2$, typical of planar liquid crystal cell under an external voltage, with $\varepsilon_{\parallel} = 1.6^2$ and $\varepsilon_{\perp} = 1.5^2$. The particles have permittivity $\varepsilon_p = 10$ and vary in volume fraction from 0.025 at the top of the cell, to 0.005 at the bottom. Their angle varies as $\theta = \pi x/2$, and the ratio of the minor to major semi-axes changes according to $1 - 0.8y$, going from ellipses at the top to circles at the bottom. The full microscopic solution, ϕ_{micro}, for this system is computed in Comsol, subject to a voltage of 2 and 1 at the top and bottom of the cell, respectively, and periodic boundary conditions at the sides. The homogenised potential, ϕ_{hom} is also solved for in Comsol, using the spatially varying permittivity 4.5.5.7a throughout the equivalent effective medium. The percentage error between the two solutions



(A) (B)

FIGURE 4.19: (a) Diagram of the particle geometries (green ellipses) and the optical axis of the host medium (blue lines). (b) The percentage error between the homogenised and Comsol solutions for the electric potential.

(calculated as $|\phi_{\text{hom}} - \phi_{\text{micro}}| / \phi_{\text{micro}}$) is shown in figure 4.19. The agreement between the two is very good, with less than 1% error through the cell. As you would expect, the largest errors occur at the particle boundaries, where the microscopic solution sees the full scattered field (before it has decayed) and the homogenised solution sees only the average of the scattering.

4.6 3D particles in Isotropic Medium

4.6.1 Traditional Effect Medium Theories

Before looking at the homogenised effective permittivity for 3D particles, we first have a brief look at the existing effective medium theories that will be used for comparison purposes [53]. There are a very large number of different models in use, ranging from power-law models of the form $\varepsilon_{\text{eff}}^{\beta} = f\varepsilon_m^{\beta} + (1-f)\,\varepsilon_p^{\beta}$, $\beta > 0$, to models for different particle shapes and distributions, and for continuously varying random media. While many of these lesser used models were developed with a specific application in mind, the most commonly used theories are those of Maxwell-Garnett, Bruggeman, and the Coherent Potential Approximation.

The Coherent Potential method is rooted heavily in the theory of Green's functions, considering the average potential of electrons within the system and their scattering from the different constituent materials [12]. The physical system is viewed as embedded in an effective host material, and the energy of the scattering propagator is adjusted in such a way that the average of the total scattering is zero; as it would be for an isotropic material. This ties in neatly with the most widely used theories of Maxwell-Garnett and Bruggeman, derived in a very different manner, through the

unified mixing rule for spherical particles [53]

$$\frac{\varepsilon_{\text{eff}} - \varepsilon_m}{\varepsilon_{\text{eff}} + 2\varepsilon_m + \nu\left(\varepsilon_{\text{eff}} - \varepsilon_m\right)} = f\frac{\varepsilon_p - \varepsilon_m}{\varepsilon_p + 2\varepsilon_m + \nu\left(\varepsilon_{\text{eff}} - \varepsilon_m\right)}, \tag{4.6.1.1}$$

where f is the volume fraction of particles. The parameter ν is dimensionless, with $\nu = 0$, 2, and 3 giving the Maxwell-Garnett, Bruggeman, and Coherent Potential theories, respectively. For later interest, this formula can be written as a power series in the volume fraction as

$$\varepsilon_{\text{eff}} = \varepsilon_m + 3\varepsilon_m\frac{\varepsilon_p - \varepsilon_m}{\varepsilon_p + 2\varepsilon_m}f + 3\varepsilon_m\left(\frac{\varepsilon_p - \varepsilon_m}{\varepsilon_p + 2\varepsilon_m}\right)\left(1 + \nu\frac{\varepsilon_p - \varepsilon_m}{\varepsilon_p + 2\varepsilon_m}\right)f^2 + \ldots. \tag{4.6.1.2}$$

Traditionally, the derivation of Maxwell-Garnett [10] relies on the idea of fictitiously removing a spherical portion of the medium (creating the so-called Lorentz cavity), and then writing the local field in the sphere as the sum of the external field, the polarisation field inside sphere, and the Lorentz field. The Lorentz field is the field that would arise from the charge on the surface of the fictitious cavity, due to the polarisation of the rest of the material. The particles in the sphere are then said to acquire a dipole moment dependent on the local field, which, upon assuming the material can be described by an isotropic permittivity ε_{eff}, allows the macroscopic potential to be related linearly to the polarisability of the particles.

More recent derivations of the model [34], however, make the physical modelling assumptions much clearer. To begin, we consider a system of spherical particles and the dipole field arising from each particle in the system in the presence of an external field, $\boldsymbol{E}_{ext}$. The dipole potential is

$$\phi = \frac{1}{4\pi\varepsilon}\frac{\boldsymbol{d}\cdot\boldsymbol{r}}{r^3}, \tag{4.6.1.3}$$

where the dipole moment is

$$\boldsymbol{d} = \alpha\boldsymbol{E}_{ext}, \tag{4.6.1.4}$$

and α is polarisability of the particle. The resulting field is

$$\boldsymbol{E}_d = -\boldsymbol{\nabla}\phi = \frac{1}{4\pi}\frac{3\hat{\boldsymbol{r}}\left(\boldsymbol{d}\cdot\hat{\boldsymbol{r}}\right) - \boldsymbol{d}}{r^3} - \frac{1}{3}\delta\left(\boldsymbol{r}\right)\boldsymbol{d}. \tag{4.6.1.5}$$

For a given volume of the composite, V, the total microscopic field at any point is then

$$\begin{aligned}\boldsymbol{E} &= \boldsymbol{E}_{ext} + \sum_n \boldsymbol{E}_d^{(n)} \\ &= \boldsymbol{E}_{ext} + \sum_n \boldsymbol{E}_d\left(\boldsymbol{r} - \boldsymbol{r}_n\right),\end{aligned} \tag{4.6.1.6}$$

where $\boldsymbol{E}_d^{(n)}$ is the field arising from the n'th dipole at position $\boldsymbol{r}_n$. Assuming a large enough volume and that all particles have the same dipole, the macroscopic field is

then assumed to be the average of the fast-fluctuating microscopic field as

$$\begin{aligned}\boldsymbol{E} &= \langle \boldsymbol{E} \rangle \\ &= \frac{1}{V}\int \boldsymbol{E} dV \\ &= \boldsymbol{E}_{ext} - \frac{1}{3}\frac{N}{V}\boldsymbol{d} \\ &= \left(1 - \frac{1}{3}\frac{N}{V}\alpha\right)\boldsymbol{E}_{ext},\end{aligned} \tag{4.6.1.7}$$

where N is total number of particles in the volume V. Only the contributions from the $\delta(\boldsymbol{r}_n)'$s are retained in the averaging, since the regular part of the dipole field is assumed to average to zero. This is because, for a large enough volume, we can assume that for every particle at $\boldsymbol{r}_n$, there is a particle at $-\boldsymbol{r}_n$ to cancel with it.

Now, using this result, and assuming that the composite material can be modelled with the single isotropic permittivity ε_{eff}, we have that the total dipole field is given by

$$\begin{aligned}\boldsymbol{d}_{tot} &= V\boldsymbol{P} \\ &= V\left(\varepsilon_{\text{eff}} - 1\right)\boldsymbol{E}.\end{aligned} \tag{4.6.1.8}$$

Moreover, we must also have that

$$\boldsymbol{d}_{tot} = N\alpha\boldsymbol{E}_{ext}. \tag{4.6.1.9}$$

Putting these together we arrive at

$$\varepsilon_{\text{eff}} = \frac{1 + 2\frac{N\alpha}{3V}}{1 - \frac{N\alpha}{3V}}, \tag{4.6.1.10}$$

or,

$$\varepsilon_{\text{eff}} = \frac{1 + \frac{2f}{4\pi R^3}\alpha}{1 - \frac{f}{4\pi R^3}\alpha}, \tag{4.6.1.11}$$

by noting that the volume fraction f is equal to $\frac{4\pi}{3}R^3\frac{N}{V}$. In this way the macroscopic permittivity is linked to behaviour of the particles at the microscopic scale through the polarisability α. This is found by considering the problem of a single particle of radius R in an external field, $\boldsymbol{E}_0$, which we take to lie parallel to the x axis:

$$\nabla^2\phi = 0\,, \qquad \forall\, \boldsymbol{x}, \tag{4.6.1.12a}$$

$$\varepsilon_m\boldsymbol{n}_0\cdot\boldsymbol{\nabla}\phi = \varepsilon_p\boldsymbol{n}_0\cdot\boldsymbol{\nabla}\phi\,, \quad |\boldsymbol{x}| = R, \tag{4.6.1.12b}$$

$$\phi_m = \phi_p\,, \qquad |\boldsymbol{x}| = R, \tag{4.6.1.12c}$$

$$\phi_m = -E_0 x, \qquad \boldsymbol{x} \to \infty. \tag{4.6.1.12d}$$

The problem is solved in spherical coordinates

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} r\cos\theta \\ r\sin\theta\sin\psi \\ r\sin\theta\cos\psi \end{pmatrix}, \tag{4.6.1.13}$$

but the rotational symmetries of the problem mean that we can say there is no ψ dependence. The possible solutions are then given by $r^l P_l(\cos\theta)$ and $r^{-(l+1)} P_l(\cos\theta)$, where the P_l are the Legendre polynomials, as

$$\phi_p = \sum_{l=0}^{\infty} r^l P_l(\cos\theta) \tag{4.6.1.14}$$

$$\phi_m = -rP_1(\cos\theta) + \sum_{l=0}^{\infty} r^{-(l+1)} P_l(\cos\theta); \tag{4.6.1.15}$$

this follows from the fact that the solution inside the particle must be regular at the origin, and that the potential must go to the external potential as $r \to \infty$. Applying the boundary conditions it is found that

$$\phi_p = \frac{3\varepsilon_m}{\varepsilon_p + 2\varepsilon_m} E_0 r\cos\theta \tag{4.6.1.16}$$

and

$$\phi_m = \left(-r\cos\theta + R^3 \frac{\varepsilon_p - \varepsilon_m}{\varepsilon_p + 2\varepsilon_m} \frac{\cos\theta}{r^2}\right) E_0. \tag{4.6.1.17}$$

From a macroscopic point of view, the charge induced by the depolarisation field on the surface of the microscopic particles will create a field resembling a dipole. We compare, then, the second term of the potential outside the particle to that of a dipole (4.6.1.3):

$$\frac{1}{4\pi\varepsilon_m} \frac{\boldsymbol{d}\cdot\boldsymbol{r}}{r^3} = \frac{1}{4\pi\varepsilon_m} \frac{\alpha \boldsymbol{E}_0 \cdot \boldsymbol{r}}{r^3} = E_0 R^3 \frac{\varepsilon_p - \varepsilon_m}{\varepsilon_p + 2\varepsilon_m} \frac{r\cos\theta}{r^3}. \tag{4.6.1.18}$$

It follows that

$$\alpha = 4\pi R^3 \varepsilon_m \frac{\varepsilon_p - \varepsilon_m}{\varepsilon_p + 2\varepsilon_m}, \tag{4.6.1.19}$$

since $\boldsymbol{E}_0 \cdot \boldsymbol{r} = E_0 r\cos\theta$. Plugging this into equation (4.6.1.11) we find the Maxwell-Garnett formula

$$\varepsilon_{\mathrm{MG}} = \varepsilon_m + 3f\varepsilon_m \frac{\varepsilon_p - \varepsilon_m}{\varepsilon_p + 2\varepsilon_m - f\left(\varepsilon_p - \varepsilon_m\right)} \tag{4.6.1.20}$$

or, equivalently after rearranging,

$$\frac{\varepsilon_{\mathrm{MG}} - \varepsilon_m}{\varepsilon_{\mathrm{MG}} + 2\varepsilon_m} = f \frac{\varepsilon_p - \varepsilon_m}{\varepsilon_p + 2\varepsilon_m}. \tag{4.6.1.21}$$

This can be generalised to systems with more than one type of inclusion by [53]

$$\frac{\varepsilon_{\mathrm{MG}} - \varepsilon_h}{\varepsilon_{\mathrm{MG}} + 2\varepsilon_h} = \sum_k f_k \frac{\varepsilon_k - \varepsilon_h}{\varepsilon_k + 2\varepsilon_h}, \tag{4.6.1.22}$$

where f_k and ε_k are the volume fraction and permittivity of the k'th type of inclusion.

Perhaps the most striking drawback to this formula is its asymmetry in how it treats the constituent components of the system. The model is derived assuming that there is a definite host medium and particle inclusions. However, in the case that the two are of similar volume fractions, the notion of host and inclusion is not so clear. In fact, you would expect that by interchanging the roles of the host and inclusion, and switching the volume fractions correspondingly, the effective permittivity should be the same. This is, in fact, not the case, with the Maxwell-Garnett formula giving very different answers in the two cases. To combat this, Bruggeman developed a symmetric theory based on that of Maxwell-Garnett. His solution [10] was to imagine that all of the physical composite system existed within an imaginary effective host with permittivity ε_{B}, and that the multi-component Maxwell-Garnett permittivity (4.6.1.22) should be equal to ε_{B}. In this way, a sample of the composite system would be indistinguishable from the isotropic host medium (from the point of view of an electric field), and so this should be the effective, isotropic permittivity of the composite. This is achieved simply by setting $\varepsilon_{\mathrm{MG}} = \varepsilon_h = \varepsilon_{\mathrm{B}}$ in (4.6.1.22) and gives the Bruggeman forumula as

$$(1-f)\,\frac{\varepsilon_m - \varepsilon_{\mathrm{B}}}{\varepsilon_m + 2\varepsilon_{\mathrm{B}}} + f\,\frac{\varepsilon_p - \varepsilon_{\mathrm{B}}}{\varepsilon_p + 2\varepsilon_{\mathrm{B}}} = 0, \tag{4.6.1.23}$$

which treats all of the components symmetrically. The effective permittivity is found by rearranging this equation and solving the resulting quadratic.

Many extensions have been made to these methods over the years to account for different particle geometries, orientations, and anisotropy in the particles and host etc. Here we discuss briefly the case of spheroidal particles in an isotropic medium, the formulas for which will be used for comparison purposes with the homogenised solutions. It can be shown [53] that, for an ellipsoidal particle aligned with one of the coordinate axes, the polarisability tensor has components

$$\alpha_i = \frac{4\pi}{3} r_x r_y r_z \frac{\varepsilon_m}{\varepsilon_p + L_i\left(\varepsilon_p - \varepsilon_m\right)}, \tag{4.6.1.24}$$

where r_x is the length of the x semi-axis etc., and L_i is the depolarisation factor of the spheroid in the i'th coordinate direction. The depolarisation factors describe the depolarising field that arises in the particles when subject to an external field. We will only consider spheroidal particles, in which two of the semi-axes are the same. Assuming that the particles' symmetry axis lies along z, we have two possible cases.

There are prolate spheroids, in which $r_z > r_y = r_x$, and oblate spheroids, in which $r_z < r_y = r_x$. The depolarisation factors are then given in terms of the particle eccentricities, e, as [39]

$$L_z = \begin{cases} \frac{1-e^2}{e^3}\left(\tanh^{-1} e - e\right), & \text{prolate,} \\ \frac{1}{e^2}\left(1 - \frac{\sqrt{1-e^2}}{e}\sin^{-1} e\right), & \text{oblate,} \end{cases} \tag{4.6.1.25a}$$

where

$$e = \begin{cases} \sqrt{1 - \frac{r_x^2}{r_z^2}}, & \text{prolate,} \\ \sqrt{1 - \frac{r_x^2}{r_z^2}}, & \text{oblate.} \end{cases} \tag{4.6.1.25b}$$

The other factors can be found from the fact that

$$L_x = L_y = \frac{1 - L_z}{2}. \tag{4.6.1.25c}$$

The Maxwell-Garnett principal permittivity in the i'th cartesian direction is then found by putting the corresponding polarisability into equation (4.6.1.11) to find

$$(\varepsilon_{\mathrm{MG}})_i = \varepsilon_m + f\varepsilon_m \frac{\varepsilon_p - \varepsilon_m}{\varepsilon_m + L_i\left(1-f\right)\left(\varepsilon_p - \varepsilon_m\right)}. \tag{4.6.1.26}$$

The Bruggeman formula is obtained analogously to the spherical case and, written explicitly for ε_{B}, is found to be

$$(\varepsilon_{\mathrm{B}})_i = \frac{-b_i + \sqrt{b_i^2 + 4\varepsilon_p\varepsilon_m\left(1-L_i\right)L_i}}{2\left(1-L_i\right)} \tag{4.6.1.27a}$$

$$b_i = f\left[\left(1-f\right)\varepsilon_p - L_i\varepsilon_m\right] + \left(1-f\right)\left[\left(1-L_i\right)\varepsilon_m - L_i\varepsilon_p\right]. \tag{4.6.1.27b}$$

We note that a sphere is just an ellipsoid in which the major and minor semi-axes are equal, all three depolarisation factors are equal to $\frac{1}{3}$, and the earlier results are recovered exactly.

4.6.2 Spherical Particles

We now consider the homogenised solution to spherical particles in an isotropic medium. Once again, we consider the limit of low volume fraction and the cell problem is given by equations (4.5.1.1.11a)-(4.5.1.1.11d). By making the change of variables

$$\boldsymbol{\gamma}^{(1)} = E_0^{-1}\tilde{\boldsymbol{\gamma}} - \boldsymbol{Y}, \tag{4.6.2.1}$$

the inner cell problem becomes

$$\nabla^2\tilde{\boldsymbol{\gamma}}_\alpha = 0, \qquad \boldsymbol{Y} \in \Omega_\alpha\left(\boldsymbol{x}\right), \tag{4.6.2.2}$$

$$\boldsymbol{n}_0 \cdot \left(\varepsilon_p \boldsymbol{\nabla} \tilde{\gamma}_p\right) = \boldsymbol{n}_0 \cdot \left(\varepsilon_m \boldsymbol{\nabla} \tilde{\gamma}_m\right) , \quad \boldsymbol{Y} \in \partial\Omega_p\left(\boldsymbol{x}\right), \tag{4.6.2.3}$$

$$\tilde{\gamma}_m = \tilde{\gamma}_p , \qquad \boldsymbol{Y} \in \partial\Omega_p\left(\boldsymbol{x}\right), \tag{4.6.2.4}$$

$$\tilde{\gamma}_m = E_0 \boldsymbol{Y}, \qquad |\boldsymbol{Y}| \to \infty. \tag{4.6.2.5}$$

This is just three copies of the polarisability problem used in the Maxwell-Garnett derivation, with the i'th component of $\tilde{\gamma}$ being the solution when the external field is applied along the i'th coordinate axis. The low volume fraction cell problem, then, is just the polarisability problem with the source moved from an external field at infinity to the boundary of the particle. We also see explicitly how the key assumption in Maxwell-Garnett is that we can neglect the interactions between particles (since this is really the only assumption we have made by considering the homogenised problem in the low volume fraction limit).

Using the result (4.6.1.16) from the previous section, the field inside the particle is given by

$$\tilde{\gamma}_p = \frac{3\varepsilon_m}{\varepsilon_p + 2\varepsilon_m} E_0 \boldsymbol{Y}. \tag{4.6.2.6}$$

Changing back to our original variables $\gamma^{(1)}$ gives

$$\begin{aligned} \gamma^{(1)} &= \left(\frac{3\varepsilon_m}{\varepsilon_p + 2\varepsilon_m} - 1\right) \boldsymbol{Y} \\ &= \frac{\varepsilon_m - \varepsilon_p}{\varepsilon_p + 2\varepsilon_m} \boldsymbol{Y}, \end{aligned} \tag{4.6.2.7}$$

which is very similar to the polarisability (4.6.1.19). Using this solution to calculate the effective permittivity (4.3.2.0.21), we have

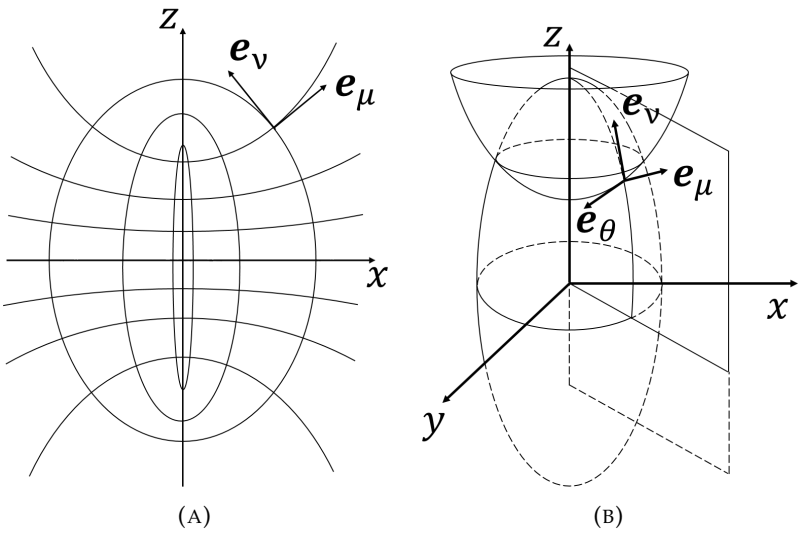
$$\varepsilon_{\text{eff}} = \left[\left(\varepsilon_p - \varepsilon_m\right) f + \varepsilon_m\right] \mathcal{I} + \left(\varepsilon_p - \varepsilon_m\right) \int_{\partial\Omega_p} \hat{\boldsymbol{n}} \otimes \gamma_p dS, \tag{4.6.2.8}$$

where we again have transformed the integral of the gradient into a surface integral, used that the cell boundary integral vanishes due to periodicity, and that $\gamma_m = \gamma_p$ at the particle boundary. The surface integral is

$$\int_{\partial\Omega_p} \hat{\boldsymbol{n}} \otimes \gamma_p dS = -\frac{\left(\varepsilon_p - \varepsilon_m\right)^2}{\varepsilon_p + 2\varepsilon_m} R^3 \int_0^{2\pi} \int_0^{\pi} \hat{\boldsymbol{r}} \otimes \hat{\boldsymbol{r}} \sin\phi d\phi d\theta$$

where R is the radius of the spheres relative to the unit cell. The off diagonal integrals are zero and the diagonal components are all $4\pi/3$. This gives

$$\begin{aligned} \varepsilon_{\text{eff}} &= \left[\left(\varepsilon_p - \varepsilon_m\right) f + \varepsilon_m - f\frac{\left(\varepsilon_p - \varepsilon_m\right)^2}{\varepsilon_p + 2\varepsilon_m}\right] \mathcal{I} \\ &= \left[\varepsilon_m + 3\varepsilon_m f \frac{\left(\varepsilon_p - \varepsilon_m\right)}{\varepsilon_p + 2\varepsilon_m}\right] \mathcal{I}, \end{aligned} \tag{4.6.2.9}$$



(A) (B)

FIGURE 4.20: Diagram depicting the prolate spheroidal coordinate system (b), formed by rotating the elliptic coordinate system (a) about the z axis.

owing to the fact that $4\pi R^3/3$ is the volume fraction of particle per unit cell, and hence the volume fraction of the overall system. The effective permittivity is, as expected, isotropic, owing to the symmetries of the sphere. Comparing to equation (4.6.1.2), the homogenised solution in the limit of low volume fraction is actually exactly the unified mixing formula 4.6.1.2 to first order.

4.6.3 Spheroidal Particles

4.6.3.1 The Cell Problem - Metallic Prolate Spheroids

We now consider the case of prolate spheroidal particles. These are formed by taking an ellipse with major and minor semi-axes of lengths a and b, respectively, and rotating the ellipse around the major semi-axis. If we take the long, symmetry axis to lie along z, a natural coordinate system to consider is the prolate spheroidal coordinates, formed by rotating the elliptic coordinate system around the z axis ($\nu = 0$):

$$\begin{aligned} x &= F \sinh\mu \sin\nu \cos\theta \\ y &= F \sinh\mu \sin\nu \sin\theta \\ z &= F \cosh\mu \cos\nu. \end{aligned} \tag{4.6.3.1.1}$$

This time surfaces of constant μ give ellipsoids, constant ν give two-sheeted hyperboloids, and constant θ give planes. The coordinate system is shown in figure 4.20. The scale factors are

$$h_\mu = h_\nu = F\sqrt{\sinh^2\mu + \sin^2\nu} \tag{4.6.3.1.2}$$

and

$$h_\theta = F\sinh\mu\sin\nu, \tag{4.6.3.1.3}$$

and the Laplacian is given by

$$\nabla^2\phi = \frac{1}{F^2\left(\sinh^2\mu + \sin^2\nu\right)}\left\{\frac{\partial^2\phi}{\partial\mu^2} + \coth\mu\frac{\partial\phi}{\partial\mu} + \frac{\partial^2\phi}{\partial\nu^2} + \cot\nu\frac{\partial\phi}{\partial\nu}\right\}\ldots + \frac{1}{F^2\sinh^2\mu\sin^2\nu}\frac{\partial^2\phi}{\partial\theta^2}. \tag{4.6.3.1.4}$$

The prolate spheroidal harmonics are then given by [38]

$$\phi = \sum_{m,n}\left[A_{mn}P_n^m(\cosh\mu) + B_{mn}Q_n^m(\cosh\mu)\right]\left[C_{mn}P_n^m(\cos\nu) + D_{mn}Q_n^m(\cos\nu)\right]\ldots \times\left[E_m e^{im\theta} + F_m e^{-im\theta}\right], \tag{4.6.3.1.5}$$

where A_{mn} to F_{mn} are undetermined coefficients, and P_n^m and Q_n^m are the associated Legendre polynomials of the first and second kind, of degree n and order m. Both P_n^m and Q_n^m satisfy relations of the form

$$P_n^m(z) = \begin{cases}\left(z^2-1\right)^{\frac{m}{2}}\frac{d^m P_n(z)}{dz^m}, & |z|>1\\ \left(1-z^2\right)^{\frac{m}{2}}\frac{d^m P_n(z)}{dz^m}, & |z|<1\end{cases} \tag{4.6.3.1.6}$$

where P_n are the regular Legendre polynomials.

As in the 2D case, we first fix the geometry of our problem. Substituting (4.6.3.1.1) into the ellipsoid equation

$$\frac{x^2+y^2}{b^2} + \frac{z^2}{a^2} = 1 \tag{4.6.3.1.7}$$

we find

$$a = F\cosh\mu_0 \tag{4.6.3.1.8}$$

and

$$b = F\sinh\mu_0, \tag{4.6.3.1.9}$$

where μ_0 is the surface describing the particle boundary. It follows that

$$F = \sqrt{a^2-b^2}, \tag{4.6.3.1.10}$$

$$\mu_0 = \tanh^{-1}\frac{b}{a} = \frac{1}{2}\ln\frac{b+a}{b-a}, \tag{4.6.3.1.11}$$

and the normal to the particle is given by

$$\boldsymbol{n} = \boldsymbol{e}_\mu\big|_{\mu=\mu_0} = \left.\frac{\partial \boldsymbol{r}}{\partial \mu}\right|_{\mu=\mu_0} = \begin{pmatrix} a\sin\nu\cos\theta \\ a\sin\nu\sin\theta \\ b\cos\nu \end{pmatrix}. \tag{4.6.3.1.12}$$

We also define the particle eccentricity as

$$e = \sqrt{1-\frac{b^2}{a^2}}, \quad 0 \le e < 1, \tag{4.6.3.1.13}$$

with 0 corresponding to a sphere and $e \to 1$ tending toward infinitely thin, needle-like particles.

We will now take a similar approach as we did for particles in an anisotropic medium by first solving the problem where there is no field inside the particles. Still operating in the low volume fraction limit, the inner cell problem we consider is

$$\nabla^2\boldsymbol{\gamma} = 0\,, \qquad \boldsymbol{Y} \in \Omega_m, \tag{4.6.3.1.14}$$

$$\boldsymbol{n}\cdot\boldsymbol{\nabla}\boldsymbol{\gamma} = \boldsymbol{n}\,, \quad \boldsymbol{Y} \in \partial\Omega_p, \tag{4.6.3.1.15}$$

$$\boldsymbol{\gamma} = const. \qquad |\boldsymbol{Y}| \to \infty. \tag{4.6.3.1.16}$$

Looking first at γ_3, since the boundary condition suggests it will be the easiest to solve, the normal derivative is

$$\boldsymbol{n}\cdot\boldsymbol{\nabla} = \boldsymbol{e}_\mu\cdot\boldsymbol{e}_\mu\frac{1}{h_\mu}\frac{\partial}{\partial u} = \frac{\partial}{\partial\mu} \tag{4.6.3.1.17}$$

and the boundary condition 4.6.3.1.15 becomes

$$\frac{\partial\gamma_3}{\partial\mu} = b\cos\nu. \tag{4.6.3.1.18}$$

There is no θ dependence, and so the order of the Legendre solutions is $m = 0$. All of the $P_n(x)$, except $P_0 = 1$, go to infinity as $x \to \infty$, and so cannot be a solution since $\boldsymbol{\gamma}$ must be bounded far from the particle. The solution can then only contain $Q_1^0(\cosh\mu)$. Now, $Q_0(x)$ (and all other Q_n, since they contain a factor of Q_0) has singularities at $z = \pm 1$, and is given by [38]

$$Q_0(z) = \begin{cases} \tanh^{-1} z = \frac{1}{2}\ln\frac{1+z}{1-z}, & |z| < 1, \\ \coth^{-1} z = \frac{1}{2}\ln\frac{z+1}{z-1}, & |z| > 1. \end{cases} \tag{4.6.3.1.19}$$

Since

$$\cosh\mu \ge \cosh\mu_0 = \frac{a}{F} = \frac{1}{e} > 1, \tag{4.6.3.1.20}$$

we need the $\coth^{-1}$ branch for our solution. To check that $Q_1(x) = x\coth^{-1}x - 1$

meets the matching condition at infinity, we expand around $t = 0$, where $t = \cosh \mu^{-1}$,to find

$$\begin{aligned}\lim_{\mu\to\infty} Q_1(\cosh\mu) &= \lim_{t\to 0}\frac{1}{2t}\ln\frac{1+t}{1-t} - 1 \\ &= \lim_{t\to 0}\frac{1}{2t}\left[\left(t - \frac{t^2}{2} + \frac{t^3}{3} + \ldots\right) - \left(-t - \frac{t^2}{2} - \frac{t^3}{3}\ldots\right)\right] - 1 \\ &= 0, \end{aligned} \tag{4.6.3.1.21}$$

as desired. This gives us that

$$\gamma_3 = AQ_1(\cosh\mu)\cos\nu \tag{4.6.3.1.22}$$

and (4.6.3.1.18) becomes

$$A = \frac{b}{\left[\frac{\partial}{\partial\mu}Q_1(\cosh\mu)\right]_{\mu=\mu_0}}. \tag{4.6.3.1.23}$$

The derivative is computed to be

$$\frac{\partial}{\partial\mu}Q_1(\cosh\mu) = \sinh\mu\left(\coth^{-1}\cosh\mu - \frac{\cosh\mu}{\sinh^2\mu}\right), \tag{4.6.3.1.24}$$

and, using

$$\cosh\mu_0 = e^{-1}, \tag{4.6.3.1.25}$$

$$\sinh\mu_0 = e^{-1}\sqrt{1-e^2}, \tag{4.6.3.1.26}$$

we find

$$A = a\left(e^{-2} - 1\right)\left(L_z - 1\right)^{-1}, \tag{4.6.3.1.27}$$

where L_z is the depolarisation factor (4.6.1.25a).

Looking now at γ_1, the boundary condition is

$$\frac{\partial\gamma_1}{\partial\mu} = a\sin\nu\cos\theta \tag{4.6.3.1.28}$$

$$= -aP_1^1(\cos\nu)\,P_1(\cos\theta). \tag{4.6.3.1.29}$$

We make the ansatz that the μ dependence of γ_1 cannot be $P_1^1(\cosh\mu) = \sinh\mu$ since this diverges for large μ. Using (4.6.3.1.6), we check that

$$\begin{aligned} Q_1^1(\cosh\mu) &= \sinh\mu\left.\frac{dQ_1(x)}{dx}\right|_{x=\cosh\mu} \\ &= \sinh\mu\coth^{-1}\cosh\mu - \coth\mu \qquad (4.6.3.1.30) \\ &= \tanh\mu Q_1(\cosh\mu) + \tanh\mu - \coth\mu \qquad (4.6.3.1.31) \end{aligned}$$

has the limit

$$\lim_{\mu\to\infty} Q_1^1(\cosh\mu) \to 1\times 0+1-1=0, \tag{4.6.3.1.32}$$

(using 4.6.3.1.21) which satisfies the matching condition. We therefore make the ansatz that

$$\gamma_1 = -BQ_1^1(\cosh\mu)\,P_1^1(\cos\nu)\,P_1(\cos\theta) \tag{4.6.3.1.33}$$

and the γ_1 boundary condition becomes

$$B = \frac{a}{\left[\frac{\partial}{\partial\mu}Q_1^1(\cosh\mu)\right]_{\mu=\mu_0}} \tag{4.6.3.1.34}$$

$$= a\left(e^{-2}-1\right)(L_z+1)^{-1}, \tag{4.6.3.1.35}$$

using (4.6.3.1.30), (4.6.3.1.25), (4.6.3.1.26), and the definition of the depolarisation factors (4.6.1.25a).

Repeating the almost identical analysis for γ_2, we find the solution to the problem to be

$$\boldsymbol{\gamma} = a\left(e^{-2}-1\right)\begin{pmatrix} (L_z+1)^{-1}Q_1^1(\cosh\mu)\sin\nu\cos\theta \\ (L_z+1)^{-1}Q_1^1(\cosh\mu)\sin\nu\sin\theta \\ (L_z-1)^{-1}Q_1^0(\cosh\mu)\cos\nu \end{pmatrix}. \tag{4.6.3.1.36}$$

Figures 4.21, 4.22, and 4.23 show a comparison of $\boldsymbol{\gamma}$ against the numerical solution, $\boldsymbol{\gamma}^{(N)}$, from Comsol, with the normalised error calculated as $\left|\boldsymbol{\gamma}-\boldsymbol{\gamma}^{(N)}\right|\max(\boldsymbol{\gamma})^{-1}$. The γ_1 solution is shown in the xy plane in figure 4.21; γ_2 is omitted since it is just a rotation of γ_1, and γ_3 is just zero to numerical error (since the xz plane is given by $\nu=0$). The γ_1 and γ_3 solutions are shown in the xz plane in figures 4.22 and 4.23. Again, γ_2 is omitted since it is zero in the xz plane ($\theta=0$), but, when viewed in the yz plane, is exactly the same as figure 4.22.

We see much the same as in the 2D case, with the solution being accurate at the particle surface and the errors growing towards the edge of the cell. The errors are due to the mismatch in boundary conditions; the analytic being a decay towards infinity, and the numerical being a decay towards the cell boundary. The analytic approximation will therefore always decay too slowly. Since the γ_i component has a source flux given by the i'th component of the particle normal, the errors in γ_i will always be worst in the i'th coordinate direction, where γ_i has its greatest magnitude. Moreover, the errors will also be greater where the particle boundary comes closer to the edge of the cell, since the discrepancy in the decay rates will be larger. We see, as expected, that the greatest error is in γ_3, where the source and major semi-axis are aligned. Even at this worst case point, and for the rather large particles considered, the error at the particle boundary is still only around 6%.

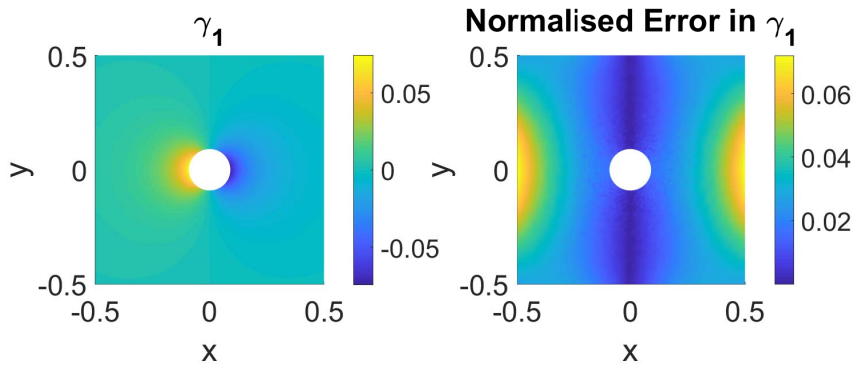


FIGURE 4.21: The x-component of γ for prolate spheroidal, metallic particles aligned along the z axis, and a comparison with the numerical solution from Comsol, shown in the xy plane. The system is described by $\varepsilon_m = 1$, $e = 0.95$, and a volume fraction of 0.02.

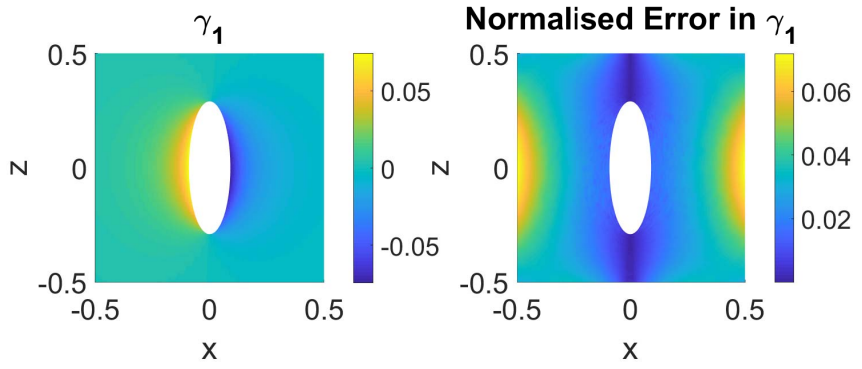


FIGURE 4.22: The x-component of γ for prolate spheroidal, metallic particles aligned along the z axis, and a comparison with the numerical solution from Comsol, shown in the xz plane. The system is described by $\varepsilon_m = 1$, $e = 0.95$, and a volume fraction of 0.02.

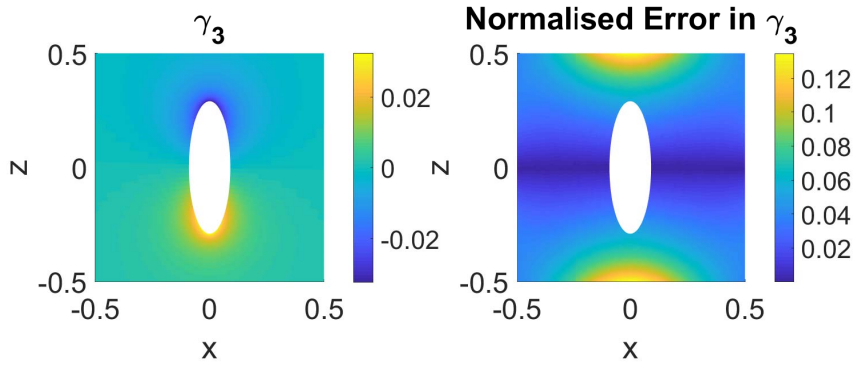
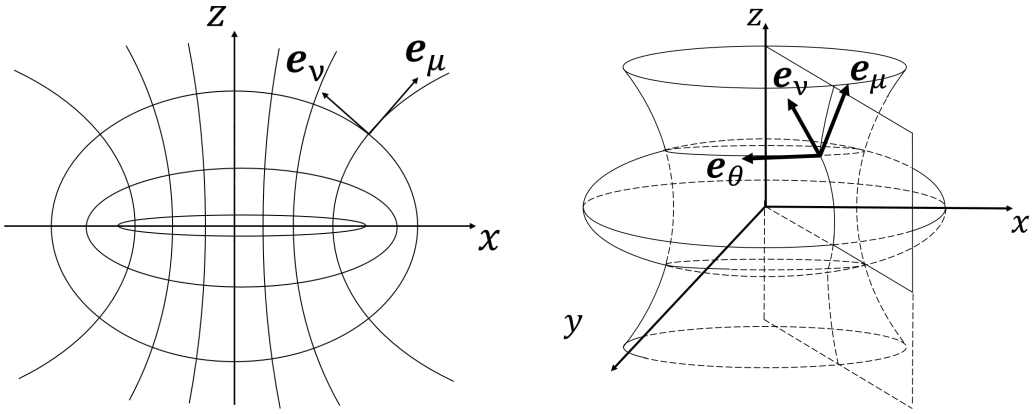


FIGURE 4.23: The z-component of γ for prolate spheroidal, metallic particles aligned along the z axis, and a comparison with the numerical solution from Comsol, shown in the xz plane. The system is described by $\varepsilon_m = 1$, $e = 0.95$, and a volume fraction of 0.02.



(A) (B)

FIGURE 4.24: Diagram depicting the oblate spheroidal coordinate system (b), formed by rotating the elliptic coordinate system (a) about the z axis.

4.6.3.2 The Cell Problem - Metallic Oblate Spheroids

We now repeat the metallic problem for oblate spheroidal particles. These are formed by taking an ellipse with major and minor semi-axes of lengths a and b, respectively, and rotating the ellipse around the minor semi-axis. If we take the short, symmetry axis to lie along z, a natural coordinate system to consider is the oblate spheroidal coordinates, formed by rotating the elliptic coordinate system around the $\nu = \pi/2$ line:

$$\begin{aligned} x &= F \cosh\mu \cos\nu \cos\theta \\ y &= F \cosh\mu \cos\nu \sin\theta \\ z &= F \sinh\mu \sin\nu. \end{aligned} \tag{4.6.3.2.1}$$

This time surfaces of constant μ give ellipsoids, constant ν give one-sheeted hyperboloids, and constant θ give planes. The coordinate system is shown in figure 4.24. The geometry parameters are the same as the prolate case, with the exception of the normal, which is given by

$$\boldsymbol{n} = \boldsymbol{e}_\mu\big|_{\mu=\mu_0} = \frac{\partial \boldsymbol{r}}{\partial \mu}\bigg|_{\mu=\mu_0} = \begin{pmatrix} b\cos\nu\cos\theta \\ b\cos\nu\sin\theta \\ a\sin\nu \end{pmatrix}. \tag{4.6.3.2.2}$$

The scale factors are [38]

$$h_\mu = h_\nu = F\sqrt{\sinh^2\mu + \sin^2\nu}, \tag{4.6.3.2.3}$$

as in the prolate case, but this time

$$h_\theta = F \cosh\mu \cos\nu. \tag{4.6.3.2.4}$$

The Laplacian is then given by

$$\begin{aligned}\nabla^2\phi &= \frac{1}{F^2\left(\sinh^2\mu + \sin^2\nu\right)}\left\{\operatorname{sech}\mu\frac{\partial}{\partial\mu}\left[\cosh\mu\frac{\partial\phi}{\partial\mu}\right] + \sec\nu\frac{\partial}{\nu}\left[\cos\nu\frac{\partial\phi}{\partial\nu}\right]\right\} \\ &= \frac{1}{F^2\cosh^2\mu\cos^2\nu}\frac{\partial^2\phi}{\partial\theta^2},\end{aligned} \tag{4.6.3.2.5}$$

and the harmonics are given by

$$\begin{aligned}\phi = \sum_{m,n}&\left[A_{mn}P_n^m\left(i\sinh\mu\right) + B_{mn}Q_n^m\left(i\sinh\mu\right)\right]\left[C_{mn}P_n^m\left(\sin\nu\right) + D_{mn}Q_n^m\left(\sin\nu\right)\right]\ldots \\ &\times\left[E_m e^{im\theta} + F_m e^{-im\theta}\right].\end{aligned} \tag{4.6.3.2.6}$$

It is the $Q_n^m\left(i\sinh\mu\right)$ that are of interest as the P_n^m diverge as $\sinh\mu$ goes to infinity. By equation (4.6.3.1.26)

$$\sinh\mu_0 \geq \sqrt{e^{-2}-1} \geq 0, \tag{4.6.3.2.7}$$

and so we need to consider both branches of Q_0, as defined in (4.6.3.1.19), for the cases where $\sinh\mu$ is less than and greater than one (and also the resulting discontinuity at $\sinh\mu$=1). However, by writing

$$\coth^{-1} ix = \tanh^{-1} ix + \begin{cases}\frac{i\pi}{2}P_0\left(ix\right), & x > 0, \\ -\frac{i\pi}{2}P_0\left(ix\right), & x < 0,\end{cases} \tag{4.6.3.2.8}$$

we see that, along the imaginary axis, the two branches of Q_0 are related by multiples of P_0. We can therefore use $\coth^{-1}\left(i\sinh\mu\right)$ for all μ, not just for $\sinh\mu > 1$, since this will only have the effect of adding a multiple of $P_0\left(i\sinh\mu\right)$, which is also a harmonic. Furthermore, the discontinuity that would have been present when switching from $\tanh^{-1}$ to $\coth^{-1}$ at $\sinh\mu = 1$ has now shifted to the origin (which can be seen from equation (4.6.3.2.8)), and we avoid the discontinuity completely (by equation (4.6.3.2.7)).

In the same manner as (4.6.3.1.21) and (4.6.3.1.32), we find that $Q_0\left(i\sinh\mu\right)$ and $Q_1^1\left(i\sinh\mu\right)$ go to zero as μ goes to infinity, and so they satisfy the matching condition. We make the ansatz that

$$\gamma = \begin{pmatrix} AQ_1^1\left(i\sinh\mu\right)P_1^1\left(\sin\nu\right)\cos\theta \\ BQ_1^1\left(i\sinh\mu\right)P_1^1\left(\sin\nu\right)\sin\theta \\ CQ_1^0\left(i\sinh\mu\right)P_1^0\left(\cos\nu\right)\end{pmatrix}, \tag{4.6.3.2.9}$$

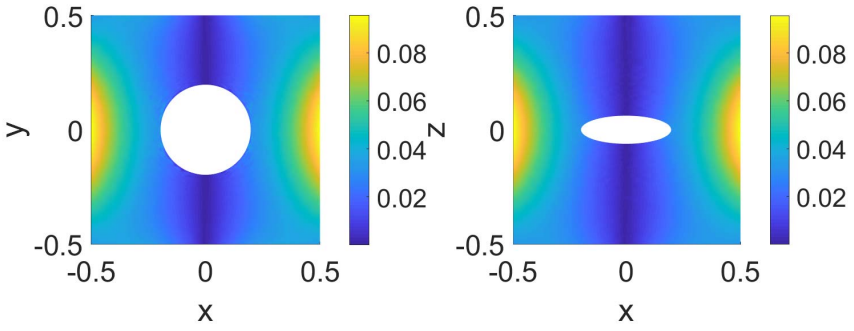


FIGURE 4.25: The normalised error in γ_1 for metallic particles between the analytic, low volume fraction approximation and the numerical solution from Comsol, shown in the xy and xz planes. The system is described by $\varepsilon_m = 1$, $e = 0.95$, and a volume fraction of 0.02.

given that the right hand side of the boundary condition (4.6.3.1.15) is given by the particle normal (4.6.3.2.2). It follows that the solution is given by

$$\gamma = -a\frac{\sqrt{1-e^2}}{e^2}\begin{pmatrix} (L_z+1)^{-1} Q_1^1(i\sinh\mu)\cos\nu\cos\theta \\ (L_z+1)^{-1} Q_1^1(i\sinh\mu)\cos\nu\sin\theta \\ (L_z-1)^{-1} Q_1^0(i\sinh\mu)\sin\nu \end{pmatrix}. \tag{4.6.3.2.10}$$

The accuracy of the oblate solution is slightly better than the prolate solution for a given volume fraction. This is a result of the volume of the spheroids scaling as a^2b for oblate particles and ab^2 for prolate particles. The symmetry axis is then smaller for oblate particles, and the mismatch in the boundary conditions of the analytic, low volume fraction approximation and the numerical calculation is smaller. The largest error is in γ_1 and γ_2 (which are rotations of one another around the z axis), being the components with a source term aligned along a major axis. These errors are shown in figure 4.25, and have a maximum of approximately 3% at the particle boundary.

4.6.3.3 The Cell Problem - Dielectric Spheroids

As before we use a linear combination of the components of the metallic solution to solve the full dielectric problem. For prolate particles, the μ dependence inside the particles can only be in terms of the $P_n^m(\cosh\mu)$, owing to the fact that the Q_n^m diverge at $\cosh\mu = 1$. Comparison to the expression for the normal in prolate coordinates tells us the solution inside the particle must be of the form

$$\gamma_p \propto \begin{pmatrix} P_1^1(\cosh\mu)\,P_1^1(\cos\nu)\cos\theta \\ P_1^1(\cosh\mu)\,P_1^1(\cos\nu)\sin\theta \\ P_1^0(\cosh\mu)\,P_1^0(\cos\nu) \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \tag{4.6.3.3.1}$$

We therefore write the solution in the particle as

$$\boldsymbol{\gamma}_p = \mathrm{C}\boldsymbol{x}, \tag{4.6.3.3.2}$$

where C is a matrix of coefficients. The first boundary condition (4.5.1.1.11b) of the full dielectric cell problem becomes

$$\begin{aligned}\varepsilon_m \boldsymbol{n} \cdot \boldsymbol{\nabla}\boldsymbol{\gamma}_m &= \left[\varepsilon_p \left(\mathcal{I} + \mathrm{C}\right) - \varepsilon_m \mathcal{I}\right] \boldsymbol{n} \\ &= \mathrm{D}\boldsymbol{n}.\end{aligned} \tag{4.6.3.3.3}$$

We know this has the solution

$$\boldsymbol{\gamma}_m = \frac{1}{\varepsilon_m}\mathrm{D}\boldsymbol{\gamma}, \tag{4.6.3.3.4}$$

where $\boldsymbol{\gamma}$ is the metallic prolate solution (4.6.3.1.36). We now look to apply the second cell problem boundary condition (4.5.1.1.11c) (continuity of $\boldsymbol{\gamma}$ at the particle boundary). On the prolate particle boundary

$$\boldsymbol{x} = a \begin{pmatrix} \sqrt{1-e^2}\sin\nu\cos\theta \\ \sqrt{1-e^2}\sin\nu\sin\theta \\ \cos\nu \end{pmatrix} \tag{4.6.3.3.5}$$

and

$$\boldsymbol{\gamma}_m\big|_{\mu=\mu_0} = \frac{a}{\varepsilon_m}\mathrm{D}\begin{pmatrix} \sqrt{1-e^2}\left(L_z-1\right)\left(L_z+1\right)^{-1}\sin\nu\cos\theta \\ \sqrt{1-e^2}\left(L_z-1\right)\left(L_z+1\right)^{-1}\sin\nu\sin\theta \\ L_z\left(L_z-1\right)^{-1}\cos\nu \end{pmatrix}. \tag{4.6.3.3.6}$$

Equating $\boldsymbol{\gamma}_m$ and $\boldsymbol{\gamma}_p$ at the boundary and then rearranging gives

$$\mathrm{C} = \left(\varepsilon_m - \varepsilon_p\right)\begin{pmatrix} \left[\frac{1+L_z}{1-L_z}\varepsilon_m + \varepsilon_p\right]^{-1} & 0 & 0 \\ 0 & \left[\frac{1+L_z}{1-L_z}\varepsilon_m + \varepsilon_p\right]^{-1} & 0 \\ 0 & 0 & \left[\left(L_z^{-1}-1\right)\varepsilon_m + \varepsilon_p\right]^{-1} \end{pmatrix}. \tag{4.6.3.3.7}$$

Repeating the same steps for oblate particles we arrive at the same answer, we just need to make sure that we use the correct depolarisation factor, as defined in (4.6.1.25a).

4.6.4 3D Effective Permittivity

4.6.4.1 The Principal Permittivity Tensor

With the solutions to the cell problem for a spheroidal particle aligned along the z axis, we are now in a position to calculate the homogenised, principal permittivity tensor

for spheroidal particles in an isotropic medium. We want to compute the usual surface integrals over the particle boundary in the spheroidal coordinate systems. However, since γ_p is proportional to $\boldsymbol{x}$, we can actually switch back to a volume integral to find

$$\begin{aligned}\int_{\partial\Omega_p} \hat{\boldsymbol{n}} \otimes \gamma_p dS &= \int_{\Omega_p} \boldsymbol{\nabla}\gamma_p dV \\ &= \int \boldsymbol{\nabla}\boldsymbol{x} dV \mathrm{C}^T \\ &= \int \mathcal{I} dV \mathrm{C}^T \\ &= \left|\Omega_p\right| \mathrm{C}. \end{aligned} \tag{4.6.4.1.1}$$

Substituting this into the effective permittivity 4.5.3.1 gives

$$\varepsilon = \left[\left(\varepsilon_p - \varepsilon_m\right) f + \varepsilon_m\right] \mathcal{I} - f \left(\varepsilon_p - \varepsilon_m\right)^2 \mathrm{C}, \tag{4.6.4.1.2}$$

where

$$f = \begin{cases} \frac{4\pi}{3} ab^2, & \text{prolate,} \\ \frac{4\pi}{3} a^2 b, & \text{oblate,} \end{cases} \tag{4.6.4.1.3}$$

$$= \left|\Omega_p\right| \tag{4.6.4.1.4}$$

is the volume fraction/volume of the particle inside the unit cell.

4.6.4.2 Arbitrary Particle Orientation

We now parameterise the particle orientation (in the lab frame) by its major (optical) axis,$\boldsymbol{a}$, by

$$\boldsymbol{a} = \begin{pmatrix} \sin\phi\cos\theta \\ \sin\phi\sin\theta \\ \cos\phi \end{pmatrix}, \tag{4.6.4.2.1}$$

where ϕ is the angle made with the z axis and θ is the angle of the projection in the xy plane with the x axis, shown in figure 4.26. The transformation that maps the lab frame to the principal frame is a rotation around z by $-\theta$, $\mathrm{R}_z\left(-\theta\right)$, followed by a rotation around the new y axis of $-\phi$, $\mathrm{R}_y\left(-\phi\right)$. Denoting the composite rotation $\mathrm{R} = \mathrm{R}_y\left(-\phi\right)\mathrm{R}_z\left(-\theta\right)$, it follows that the permittivity tensor in the lab frame is

$$\begin{aligned}\varepsilon_{\text{eff}} &= \mathrm{R}^T \tilde{\varepsilon}_{\text{eff}} \mathrm{R} \\ &= \left[\left(\varepsilon_p - \varepsilon_m\right) f + \varepsilon_m\right] \mathcal{I} + f \left(\varepsilon_p - \varepsilon_m\right) \mathrm{R}^T \mathrm{C} \mathrm{R} \\ &= \varepsilon_i \mathcal{I} - \mathrm{R}^T \hat{\varepsilon} \mathrm{R} \end{aligned} \tag{4.6.4.2.2}$$

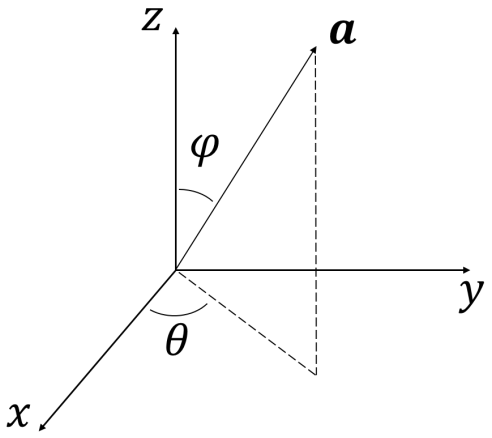


FIGURE 4.26: Diagram of the angles used to parameterise the major/optical axis, $\boldsymbol{a}$, of the spheroids.

where $\tilde{\varepsilon}_{\text{eff}}$ is the effective permittivity in the principal frame, (4.6.4.1.2). The first term, ε_i, is the isotropic volume average. The second term, $\hat{\varepsilon}$, is the particle induced anisotropy, as in the 2D case. We will denote by $\hat{\varepsilon}_\perp$ and $\hat{\varepsilon}_\parallel$ the components of $\hat{\varepsilon}$ that are perpendicular and parallel to the optical axis in the principal frame, i.e., the anisotropic components of the ordinary and extraordinary permittivities.

We now study the behaviour of the permittivity. Performing the ϕ rotations we find

$$\varepsilon_{\text{eff}} = \left[\varepsilon_i + \frac{\hat{\varepsilon}_\perp + \hat{\varepsilon}_\parallel}{2}\right]\mathcal{I} + \frac{\hat{\varepsilon}_\perp - \hat{\varepsilon}_\parallel}{2}\mathrm{R}_z(\theta)\begin{pmatrix}\cos 2\phi & 0 & \sin 2\phi \\ 0 & 1 & 0 \\ \sin 2\phi & 0 & -\cos 2\phi\end{pmatrix}\mathrm{R}_z(-\theta). \quad (4.6.4.2.3)$$

The zz component is unchanged by the θ rotations, giving

$$(\varepsilon_{\text{eff}})_{zz} = \varepsilon_i + \frac{\hat{\varepsilon}_\perp + \hat{\varepsilon}_\parallel}{2} - \frac{\hat{\varepsilon}_\perp - \hat{\varepsilon}_\parallel}{2}\cos 2\phi. \quad (4.6.4.2.4)$$

We see that the permittivity in the z direction, as you may expect, oscillates between $\varepsilon_i + \varepsilon_\parallel$ and $\varepsilon_i + \varepsilon_\perp$ as $\cos 2\phi$, achieved when the particle is aligned parallel ($\phi = \pi$) and perpendicular ($\phi = \pi/2$) to the z axis. The z/xy components, describing the permittivity in directions at 45° to the z/optical axis, vary as

$$(\varepsilon_{\text{eff}})_{z/xy} = \frac{\hat{\varepsilon}_\perp - \hat{\varepsilon}_\parallel}{2}\sin 2\phi\begin{pmatrix}0 & 0 & \cos\theta \\ 0 & 0 & \sin\theta \\ \cos\theta & \sin\theta & 0\end{pmatrix}, \quad (4.6.4.2.5)$$

also behaving exactly as expected. Their maximum modulation is the average of perpendicular and parallel permittivities, and is limited by $\sin 2\phi$, achieving maxima and minima at $\phi = \pi/4$ and $\phi = 0,\ \pi/2$, respectively. They also then oscillate with θ,

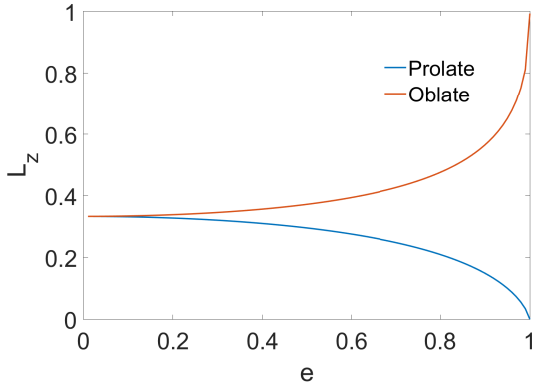


FIGURE 4.27: Graph showing the behaviour of the depolarisation factors, L_z, with the particle eccentricity, e.

with the xz component achieving its maximum when $\theta = 0$ (optical axis in the xz plane) , and at $\theta = \pi/2$ for the yz component. Finally we have the x/y components, given by

$$(\varepsilon_{\text{eff}})_{x/y} = \left[\varepsilon_i + \frac{\left(1+\cos^2\phi\right)\hat{\varepsilon}_\perp + \sin^2\phi\hat{\varepsilon}_\parallel}{2}\right]\mathcal{I} - \frac{\hat{\varepsilon}_\perp - \hat{\varepsilon}_\parallel}{2}\sin^2\phi\begin{pmatrix}\cos 2\theta & \sin 2\theta & 0\\ \sin 2\theta & -\cos 2\theta & 0\\ 0 & 0 & 0\end{pmatrix}. \tag{4.6.4.2.6}$$

For $\phi = 0$ and π we see that these components are correctly fixed at $\hat{\varepsilon}_\perp$, independently of θ. Otherwise, they oscillate with θ, with period π. The minimum xx component occurs at $\theta = \pi/2$ (and the minimum yy component at $\theta = 0$), where the above formula gives $\hat{\varepsilon}_\perp$ independently of ϕ, as desired. The maximum xx component occurs when $\theta = 0$ (and the maximum yy component when $\theta = \pi/2$) and is given by $\cos^2\phi\hat{\varepsilon}_\perp + \sin^2\phi\hat{\varepsilon}_\parallel$, correctly giving $\hat{\varepsilon}_\parallel$ when $\phi = \pi/2$. We also note that the arising factors of $\sin^2\phi$ and $\cos^2\phi$, instead of $\sin 2\phi$ and $\cos 2\phi$, mean that the modulation amplitude is symmetric for optical axis alignments above and below the xy plane, as it should be.

We can also study the behaviour with particle shape. The dependence of the depolarisation factors on eccentricity is shown in figure 4.27. For both oblate and prolate particles, $e = 0$ gives spheres, which have a depolarisation factor of $1/3$. Substituting this into (4.6.3.3.7) and (4.6.4.1.2) gives

$$\mathrm{C} = \frac{\varepsilon_m - \varepsilon_p}{\varepsilon_p + 2\varepsilon_m}\mathcal{I} \tag{4.6.4.2.7}$$

and

$$\varepsilon = \left(\varepsilon_p - \varepsilon_m\right)f + \varepsilon_m - f\frac{\left(\varepsilon_p - \varepsilon_m\right)^2}{\varepsilon_p + 2\varepsilon_m}, \tag{4.6.4.2.8}$$

which is exactly the solution obtained in (4.6.2.9). We can also consider the limit $e \to 1$,

in which prolate particles go to needles and oblate particles go to flat discs. Supposing we fix the major axis a, the volume fraction

$$f = \frac{4\pi}{3}ab^2 = \frac{4\pi}{3}a^3\frac{b^2}{a^2} = \frac{4\pi}{3}a^3\left(1-e^2\right) \to 0 \tag{4.6.4.2.9}$$

for prolate particles and

$$f = \frac{4\pi}{3}a^2b = \frac{4\pi}{3}a^3\sqrt{1-e^2} \to 0 \tag{4.6.4.2.10}$$

for oblate particles. Thus, in the limiting case that the particles become increasingly one-dimensional, their volume fraction vanishes and the effective permittivity simply goes to ε_m.

We can also fix f as $e \to 1$. The needles must stretch in length, becoming infinitely long cylinders, and the permittivity is a constant function in the direction of the major axis. Considering the principal frame, $L_z \to 0$,

$$\left(\varepsilon_{\text{eff}}\right)_{zz} \to \varepsilon_p f + \varepsilon_m\left(1-f\right), \tag{4.6.4.2.11}$$

and

$$\left(\varepsilon_{\text{eff}}\right)_{xx} = \left(\varepsilon_{\text{eff}}\right)_{yy} \to \left(\varepsilon_p - \varepsilon_m\right) f + \varepsilon_m - f\frac{\left(\varepsilon_p - \varepsilon_m\right)^2}{\varepsilon_p + \varepsilon_m}. \tag{4.6.4.2.12}$$

If each line $z = const.$ is considered like a capacitor with the same area and thickness, the zz component is exactly what you would expect by considering that capacitors in parallel simply sum. The xx and yy components are also exactly the solution to the 2D ellipsoidal problem 4.5.3.11 when the major and minor semi-axes are set equal. For the oblate particles, fixing f and letting $e \to 1$ means they become infinitely large sheets. $L_z \to 1$ and

$$\left(\varepsilon_{\text{eff}}\right)_{xx} = \left(\varepsilon_{\text{eff}}\right)_{yy} \to \varepsilon_p f + \varepsilon_m\left(1-f\right), \tag{4.6.4.2.13}$$

and

$$\left(\varepsilon_{\text{eff}}\right)_{zz} \to \left(\varepsilon_p - \varepsilon_m\right) f + \varepsilon_m - f\frac{\left(\varepsilon_p - \varepsilon_m\right)^2}{\varepsilon_p}. \tag{4.6.4.2.14}$$

The xx and yy components are once again the same as the capacitors in parallel analogy, since we are looking down the sheets from the side. The zz component should be the same as the 1D homogenised problem considered in 4.3.3.0.6, equivalent to capacitors in series. The 1D solution is given by

$$\varepsilon_{\text{eff}}^{1D} = \frac{\varepsilon_m\varepsilon_p}{\varepsilon_m f + \varepsilon_p\left(1-f\right)}, \tag{4.6.4.2.15}$$

which is clearly not the same as (4.6.4.2.14). However, if we linearise $\varepsilon_{\text{eff}}^{1D}$ in f, which is valid in the low volume fraction limit, we recover (4.6.4.2.14) exactly, as desired.

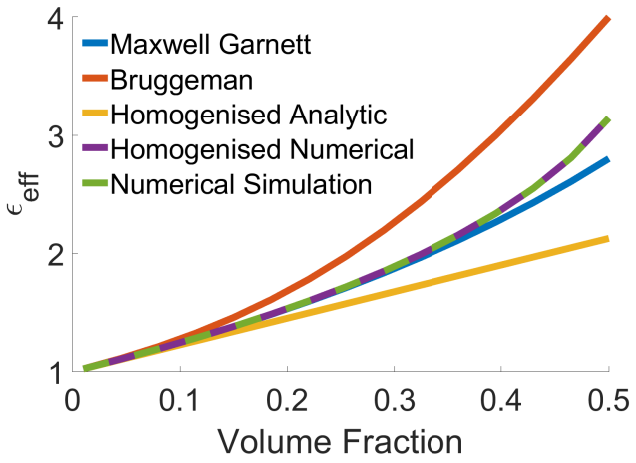


FIGURE 4.28: A comparison of the various effective medium theories for spherical particles in an isotropic medium

4.6.5 Comparison of Homogenisation with Traditional Theories

We now make a comparison of the homogenised effective permittivity with the Maxwell-Garnett and Bruggeman theories, along with the results of a full numerical simulation of the composite system, ε_{num}. This is found in Comsol by considering a stack of equispaced particles in medium (with periodic boundary conditions), applying a voltage down the stack, and then calculating the electric potential throughout the system. By calculating the total electromagnetic energy

$$W = \frac{1}{2}\int \boldsymbol{E}\cdot\boldsymbol{D}dV \tag{4.6.5.1}$$

we can then find ε_{num} by comparing this infinitely extending system of particles to a standard parallel plate capacitor model. This means we have

$$W = \frac{1}{2}CV^2, \tag{4.6.5.2}$$

and

$$C = \frac{\varepsilon_{\text{num}}A}{d} \tag{4.6.5.3}$$

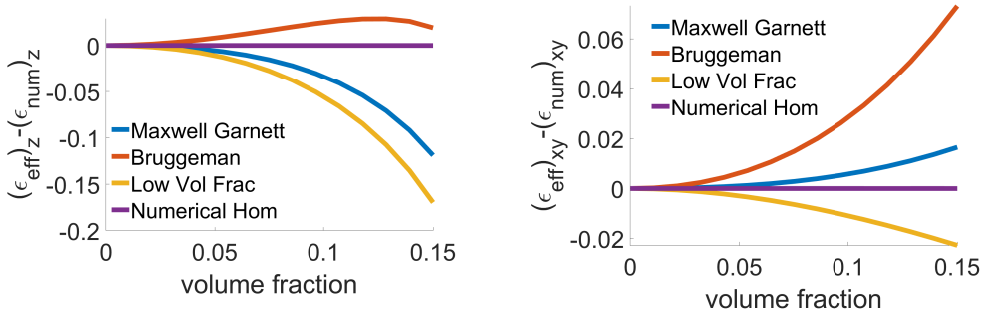
(where C is the capacitance, V is the voltage, A is the area, and d is the thickness), from which we can then calculate ε_{num}. This is, in theory, the capacitance that we would measure if we were to experimentally apply a voltage across a slab of composite.

First we consider the case of particles in an isotropic medium. The various permittivities for spherical particles are shown in figure 4.28, for $\varepsilon_m = 1$ and $\varepsilon_p = 10$. The volume fraction stops at 0.5 as this is the point just before the particles touch. The homogenised analytic solution is linear in volume fraction, so it is no surprise that it becomes inaccurate a higher volume fractions (as opposed to the homogenised

solution, which is computed entirely numerically, including the cell problem, under no other assumptions than periodicity). Bruggeman is also significantly different from the others, especially at larger volume fractions. This seems surprising at first, given that it was formulated specifically for the case where the host/inclusion relationship is more symmetric. However, in the formulation of Bruggeman the host is considered as an inclusion in exactly the same way as the particles are considered. We are therefore implicitly assuming that the host is also composed of spheres, which is clearly not an applicable approximation in this case. On the other hand, Maxwell-Garnett proves itself to be accurate to (perhaps surprisingly) large volume fractions. This can be explained, though, given that this system of spherical particles on a cubic lattice matches exactly with the system considered in the derivation. The assumption that the dipole fields from the nearby particles sum to zero is, in fact, not an approximation in this highly symmetric situation, but completely true. Thus, ignoring the inter-particle interactions, which would usually introduce errors, is completely valid in this case. Finally, the accuracy of the homogenised solution proves to be virtually identical to the experimental result. This is maybe also unsurprising given that it, too, is derived on a periodic lattice; and also given the results of section 4.6.2, where we saw that the cell problem is equivalent to the electrostatic problem, up to a change of variables. If we made the same change of variables for the full cell problem (not the low volume fraction approximation), we would be solving the periodic, electrostatic problem for an external potential applied in each of the three cartesian directions. This is exactly the experimental problem described above, just repeated for the three particle orientations that correspond to applied fields of the cell problem.

These ideas are confirmed when we consider the results of the various permittivities for ellipsoidal particles. In this instance, the cubic symmetry of the system is broken now that one of the three semi-axes differs from the others. This can be seen in figure 4.29, where the eccentricity is fixed at 0.8, and the difference between the various permittivies and the experimental solution is shown as a function of volume fraction. Noting that the volume fraction at which the particles touch is much smaller than for spheres, stopping at a little over 0.15, and comparing to figure 4.28, we see just how much the particle shape anisotropy has decreased the accuracy of Maxwell-Garnett, which was indistinguishable from the homogenised and experimental permittivities at a volume fraction of 0.15 for spherical particles. It does, though, fair much better in the directions of the minor-semi-axes, in which the field sees the circular, cross-sectional symmetry of the particles.

We also see increased errors in the z permittivity of the analytic homogenised solution, despite the volume fraction being smaller. We must note that calling it a low volume fraction approximation is perhaps somewhat of a misnomer. The approximation we actually made was that the size of the particle was small compared with the unit cell. The major axis of the particle grows to the size of the unit cell much quicker than the



(A) (B)

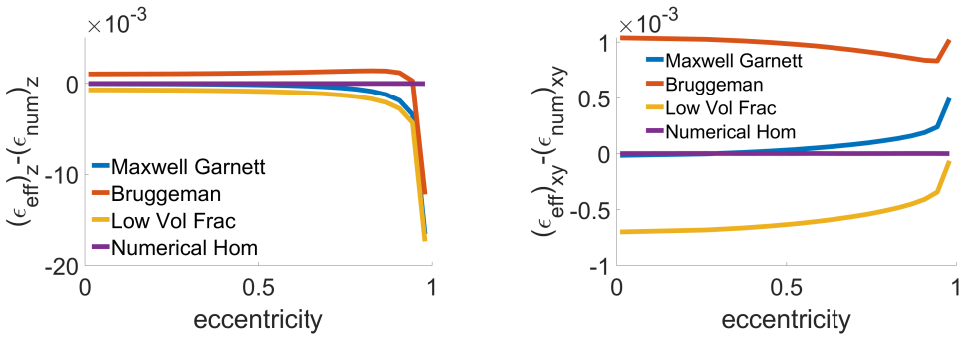
FIGURE 4.29: A comparison of the various permittivities as a function of volume fraction for prolate particles. The eccentricity is fixed at 0.8, along with $\varepsilon_p = 10$ and $\varepsilon_m = 1$.

minor axis, and so we see the errors in the z direction growing much faster than in the xy directions.

On the other hand, Bruggeman does far better in its approximation of the z permittivity, especially for larger volume fractions. This is likely due to the system approaching the capacitors in parallel analogy, where there is much greater symmetry between the host and particles. The xy permittivities, the directions in which the system is much more asymmetric (with respect to host/particle), see similar errors as the spherical case. As expected from the earlier discussion, the results of numerical homogenisation (which only requires the assumption of periodicity) and the experimental permittivites are virtually identical.

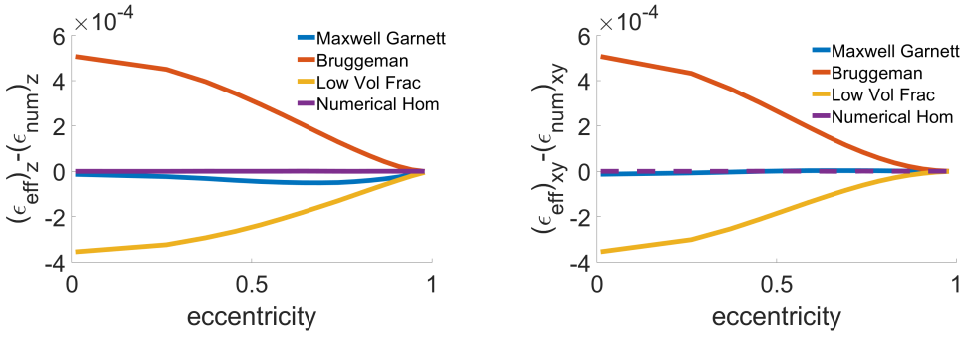
We can also look at how the permittivites vary with eccentricity. Figure 4.30 shows the behaviour with growing eccentricity of the various permittivities for a fixed volume fraction of 0.02 (enabling the eccentricity to range up to 0.98 before the particles touch). In this case the major axis of the particle grows to the width of the cell while the minor axis shrinks to zero. As expected, the largest errors in the analytic formulae are in the z direction, and we see these errors growing with eccentricity as the particles become closer together and the particle interactions that are neglected become stronger. Figure 4.31 instead shows the case where the major axis is fixed at 0.15. The minor axis and volume fraction then go to zero as the eccentricity increases, and all of the different permittivities consequently converge to the same value.

It is clear that, regardless of size and shape, numerical homogenisation provides the most accuracy in the calculation of the effective permittivity. Maxwell-Garnett is also, in general, rather accurate, provided that the volume fraction and particle shape anisotropy are low. This is due to the assumption that the particle interactions average to zero; which falls apart when the particle lattice loses its cubic symmetry, and also when the particles become closer together, meaning the scattered field seen by its neighbours cannot be compared to a point source dipole. The analytic, low volume



(A) (B)

FIGURE 4.30: A comparison of the various permittivities as a function of eccentricity for prolate particles. The volume fraction is fixed at 0.02, along with $\varepsilon_p = 10$ and $\varepsilon_m = 1$.



(A) (B)

FIGURE 4.31: A comparison of the various permittivities as a function of eccentricity for prolate particles. The major semi-axis is fixed at 0.15, along with $\varepsilon_p = 10$ and $\varepsilon_m = 1$.

fraction approximation to the homogenised solution makes a slightly different approximation, that there are no particle interactions at all. It seems that the errors introduced by this assumption are, while not enormous, still more costly than assuming zero-average interactions. For an improvement to the analytic homogenised solution to be made, a better solution to the periodic cell problem is necessary.

4.6.6 A Periodic Approximation to the Cell Problem

It is not clear if finding an analytic solution to the periodic cell problem is even possible. The boundary conditions at the particle surface necessitate the use of spheroidal coordinates, and the resulting solutions are in no way easily transformed back into cartesian coordinates to enforce periodicity on the cell boundaries. However, it turns out that we can find an approximate periodic solution that greatly increases the accuracy of the analytic solution. Returning to section 4.6.3.1, where we considered small, metallic, prolate particles, we found that the solution in the medium could only

have a dependence on the Legendre polynomials of the second kind, in the form

$$\boldsymbol{\gamma}^{(Q)} \propto \begin{pmatrix} Q_1^1(\cosh\mu)\sin\nu\cos\theta \\ Q_1^1(\cosh\mu)\sin\nu\sin\theta \\ Q_1^0(\cosh\mu)\cos\nu \end{pmatrix}. \tag{4.6.6.1}$$

This was because the other types of solution matching the boundary condition at the particle were of the form

$$\boldsymbol{\gamma}^{(P)} \propto \begin{pmatrix} P_1^1(\cosh\mu)\sin\nu\cos\theta \\ P_1^1(\cosh\mu)\sin\nu\sin\theta \\ P_1^0(\cosh\mu)\cos\nu \end{pmatrix} = \frac{\boldsymbol{x}}{F}, \tag{4.6.6.2}$$

which are linear in the cartesian coordinates and diverge at infinity. If, however, we enforce our boundary conditions on the sides of the unit square, rather than at infinity, these linear solutions are now feasible, and can be used to obtain a much better approximation than before. We can add an external potential at infinity that we choose in such a way as to minimise the error from the periodic solution.

Consider, for instance, the γ_1 component of the low volume fraction approximation, depicted in figures 4.21 and 4.22. The solution is even in both the y and z directions (and so periodic), but is odd in the x direction. We can therefore add a linear term in x to force the solution to decay quicker in the x direction, which is where we saw the largest error. As mentioned, evaluating the $\boldsymbol{\gamma}^{(Q)}$ solutions along the entirety of the sides $x = \pm 1/2$ is not at all simple (or even possible), but we can easily do it at the points $\boldsymbol{x} = (\pm 1/2, 0, 0)$, where we get

$$\gamma_1^{(Q)} = \pm Q_1^1(\cosh\mu_x), \tag{4.6.6.3}$$

where

$$\mu_x = \sinh^{-1}\frac{1}{2F} \tag{4.6.6.4}$$

is the coordinate of the ellipsoid whose minor semi-axis is equal to $1/2$. We can then add the boundary condition that the solution must decay to zero at this point, with the assumption that the solution along the rest of the $x = \pm 1/2$ sides will be approximately periodic.

We can do the same for other components, adding a linear correction in y to γ_2 and a correction in z to γ_3, leaving us with the approximation that

$$\boldsymbol{\gamma} = \begin{pmatrix} AQ_1^1(\cosh\mu)\sin\nu\cos\theta + Bx \\ CQ_1^1(\cosh\mu)\sin\nu\sin\theta + Dy \\ EQ_1^0(\cosh\mu)\cos\nu + Gz \end{pmatrix}. \tag{4.6.6.5}$$

The usual metallic problem boundary condition $\boldsymbol{n} \cdot \boldsymbol{\nabla}\gamma = \boldsymbol{n}$ gives

$$\frac{\partial}{\partial \mu}\begin{pmatrix} AQ_1^1(\cosh\mu) \\ CQ_1^1(\cosh\mu) \\ EQ_1^0(\cosh\mu) \end{pmatrix}_{\mu=\mu_0} = \begin{pmatrix} (1-B)\,a \\ (1-D)\,a \\ (1-G)\,b \end{pmatrix}. \tag{4.6.6.6}$$

We then add the periodic approximation boundary condition

$$\begin{pmatrix} AQ_1^1(\cosh\mu_x) \\ CQ_1^1(\cosh\mu_y) \\ EQ_1^0(\cosh\mu_z) \end{pmatrix} = -\frac{1}{2}\begin{pmatrix} B \\ D \\ G \end{pmatrix}, \tag{4.6.6.7}$$

where $\mu_y = \mu_x$, as defined above, and

$$\cosh\mu_x = \sqrt{(4F^2)^{-1} + 1}, \tag{4.6.6.8}$$

$$\cosh\mu_z = \frac{1}{2F}. \tag{4.6.6.9}$$

Using the results of section 4.6.3.1 for the Q terms, we solve this system of equations to find

$$A = C = a\frac{1-e^2}{e^2}\left[(L_z+1) - \frac{3}{2\pi F^2}\left|\Omega_p\right| Q_1^1(\cosh\mu_x)\right]^{-1} \tag{4.6.6.10}$$

and

$$E = a\frac{1-e^2}{e^2}\left[(L_z-1) - \frac{3}{2\pi F^2}\left|\Omega_p\right| Q_1^0(\cosh\mu_z)\right]^{-1}. \tag{4.6.6.11}$$

This is almost the same as the low volume fraction case, but we have now added the terms proportional to $\left|\Omega_p\right|$ that are due to the linear corrections. From equation 4.6.6.6 we find

$$B = D = -2Q_1^1(\cosh\mu_x)\,A \tag{4.6.6.12}$$

and

$$G = -2Q_1^0(\cosh\mu_z)\,E. \tag{4.6.6.13}$$

We apply the usual methodology to find the field inside the particle, and so we need to evaluate $\boldsymbol{\gamma}_p = \mathrm{C}\boldsymbol{x}$ and $\boldsymbol{\gamma}_m = \varepsilon_m^{-1}\mathrm{D}\boldsymbol{\gamma}$ at the particle boundary. As before, we have

$$\boldsymbol{x}|_{\mu=\mu_0} = a\begin{pmatrix} \sqrt{1-e^2}\sin\nu\cos\theta \\ \sqrt{1-e^2}\sin\nu\sin\theta \\ \cos\nu \end{pmatrix}, \tag{4.6.6.14}$$

and so we write

$$\boldsymbol{\gamma}_m|_{\mu=\mu_0} = \frac{1}{\varepsilon_m}\mathrm{D}\left[\mathrm{K} + \begin{pmatrix} B & 0 & 0 \\ 0 & B & 0 \\ 0 & 0 & G \end{pmatrix}\right]\boldsymbol{x}, \tag{4.6.6.15}$$

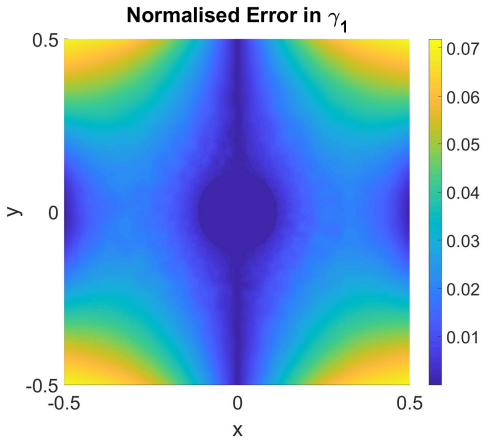


FIGURE 4.32: The normalised error between the linearly shifted γ_1 solution and the periodic solution from Comsol. As in figure 4.21, the system is described by $\varepsilon_p = 10$, $\varepsilon_m = 1$, $e = 0.95$, and a volume fraction of 0.02.

where

$$\mathrm{K} = \begin{pmatrix} Aa^{-1}\sqrt{1-e^2}^{-1} Q_1^1(\cosh\mu) & 0 & 0 \\ 0 & Aa^{-1}\sqrt{1-e^2}^{-1} Q_1^1(\cosh\mu) & 0 \\ 0 & 0 & Ea^{-1} Q_1^0(\cosh\mu) \end{pmatrix}. \tag{4.6.6.16}$$

Equating γ_m and γ_p and rearranging, we find that C is a diagonal matrix with components

$$\mathrm{C}_{11} = \mathrm{C}_{22} = (\varepsilon_p - \varepsilon_m)\left[\frac{L_z + 1 - \frac{3}{2\pi F^2}|\Omega_p| Q_1^1(\cosh\mu_x)}{L_z - 1 - \frac{3}{2\pi F^2}|\Omega_p| Q_1^1(\cosh\mu_x)}\varepsilon_m - \varepsilon_p\right]^{-1} \tag{4.6.6.17}$$

and

$$\mathrm{C}_{33} = (\varepsilon_p - \varepsilon_m)\left[\frac{L_z - 1 - \frac{3}{2\pi F^2}|\Omega_p| Q_1^0(\cosh\mu_z)}{L_z - \frac{3}{2\pi F^2}|\Omega_p| Q_1^0(\cosh\mu_z)}\varepsilon_m - \varepsilon_p\right]^{-1}. \tag{4.6.6.18}$$

It turns out that this solution is not much of an improvement on the low volume fraction approximation. While the low volume fraction approximation decayed too slowly away from the particle, resulting in a field inside the particle that was too small, this new approximation decays too quickly, resulting in an field that is too large. Now, we added a linear correction to, say, γ_1, that fixed the error at$(\pm 1/2, 0, 0)$, to zero. This was the point with the largest error, and so, by shifting the entire solution along the side $x = 1/2$, we have simply moved this maximum error out towards the corners, as depicted in figure 4.32. It follows that a much better approximation would be the average of these two solutions, the total error at the boundaries will be better minimised and the field inside the particle will be better approximated. Under this

approximation the x and y components of the particle field are described by

$$\hat{\mathrm{C}}_{11} = \hat{\mathrm{C}}_{22} = \left(\varepsilon_p - \varepsilon_m\right) \frac{L_z - 1 - \left[1 + \left(L_z - 1\right)\left(\varepsilon_p - \varepsilon_m\right)\right] \tilde{Q}_x}{\left(L_z + 1\right)\varepsilon_m - \left(L_z - 1\right)\varepsilon_p + \left(\varepsilon_p - \varepsilon_m\right) \tilde{Q}_x} \tag{4.6.6.19}$$

and the z component by

$$\hat{\mathrm{C}}_{33} = \left(\varepsilon_p - \varepsilon_m\right) \frac{L_z - \left[1 + L_z\left(\varepsilon_p - \varepsilon_m\right)\right] \tilde{Q}_z}{\left(L_z - 1\right)\varepsilon_m - L_z\varepsilon_p + \left(\varepsilon_p - \varepsilon_m\right) \tilde{Q}_z}, \tag{4.6.6.20}$$

where

$$\tilde{Q}_x = \frac{3}{2\pi F^2} \left|\Omega_p\right| Q_1^1 \left(\cosh \mu_x\right), \tag{4.6.6.21}$$

and

$$\tilde{Q}_z = \frac{3}{2\pi F^2} \left|\Omega_p\right| Q_1^0 \left(\cosh \mu_z\right). \tag{4.6.6.22}$$

As before the effective permittivity is given by

$$\varepsilon_{\text{eff}} = \left[\left(\varepsilon_p - \varepsilon_m\right) f + \varepsilon_m\right] \mathcal{I} + f \left(\varepsilon_p - \varepsilon_m\right) \hat{\mathrm{C}}. \tag{4.6.6.23}$$

Repeating the above analysis for oblate particles, we find that the effective permittivity is found simply by setting

$$\tilde{Q}_x = -\frac{3}{2\pi F^2} \left|\Omega_p\right| Q_1^1 \left(i \sinh \mu_x\right) \tag{4.6.6.24}$$

and

$$\tilde{Q}_z = -\frac{3}{2\pi F^2} \left|\Omega_p\right| Q_1^0 \left(i \sinh \mu_z\right), \tag{4.6.6.25}$$

where

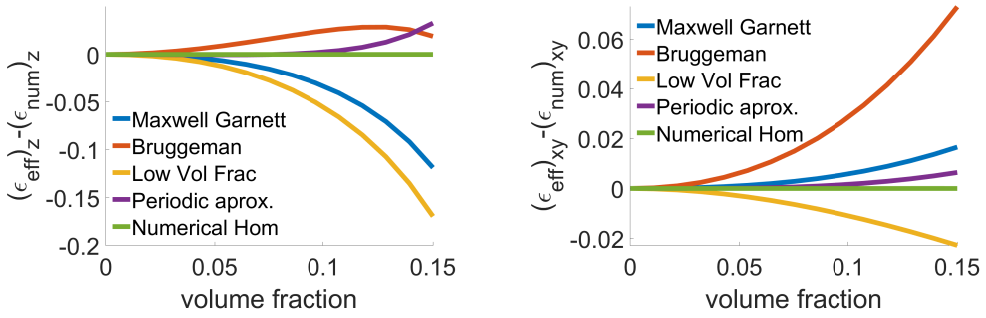
$$\sinh \mu_x = \sinh\left(\cosh^{-1} \frac{1}{2F}\right) = \sqrt{\frac{1}{4F^2} - 1} \tag{4.6.6.26}$$

and

$$\sinh \mu_z = \frac{1}{2F}. \tag{4.6.6.27}$$

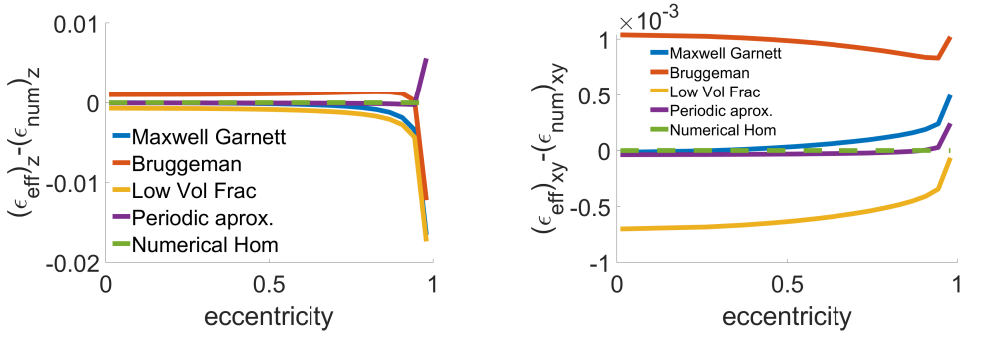
4.6.7 Comparison of Homogenisation with Traditional Theories Cont.

We now reproduce figures 4.29, 4.30, and 4.31, including the new periodic approximation, in figures 4.33, 4.34, and 4.35, respectively. We see that, in general, the periodic approximation offers much greater accuracy than the other analytic alternatives. It scales far better with volume fraction/particle size, as shown in figure 4.33, maintaining excellent agreement with numerical homogenisation and the experimental permittivity up to a volume fraction of around 0.1. This is for a particle with an eccentricity of 0.8, corresponding to a major semi-axis slightly larger than 0.4, almost covering the entire cell width in the z direction. For reference the z permittivity



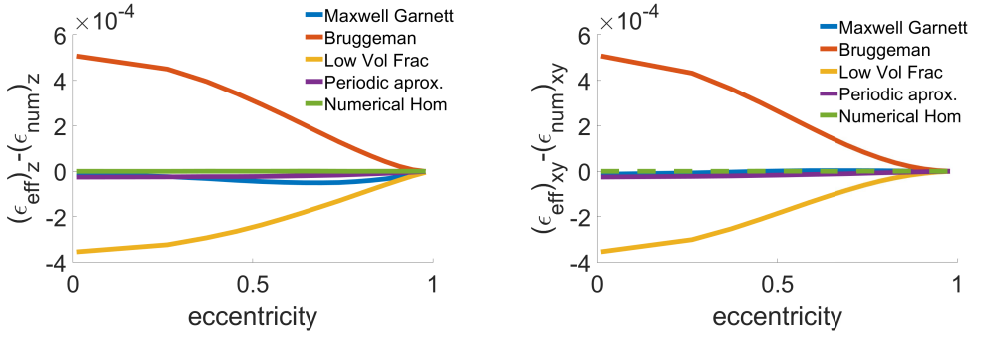
(A) (B)

FIGURE 4.33: A comparison of the various effective permittivities for prolate particles in the principal frame of the effective medium (particles major axis along z), with e =0.8, $\varepsilon_p = 10$, and $\varepsilon_m = 1$.



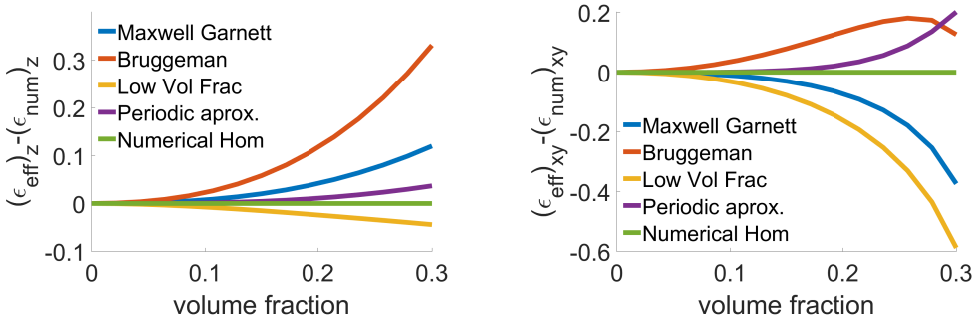
(A) (B)

FIGURE 4.34: A comparison of the various permittivities as a function of eccentricity for prolate particles. The volume fraction is fixed at 0.02, along with $\varepsilon_p = 10$ and $\varepsilon_m = 1$.



(A) (B)

FIGURE 4.35: A comparison of the various permittivities as a function of eccentricity for prolate particles. The major semi-axis is fixed at 0.15, along with $\varepsilon_p = 10$ and $\varepsilon_m = 1$.

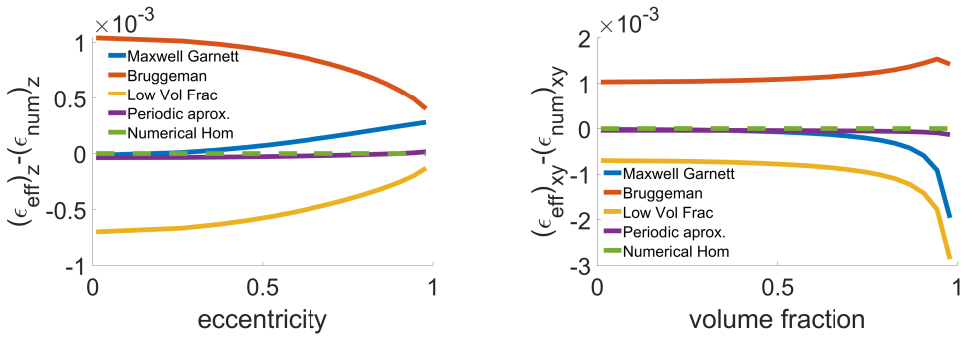


(A) (B)

FIGURE 4.36: A comparison of the various effective permittivities for oblate particles in the principal frame of the effective medium (particles major axis along z), with e =0.8, $\varepsilon_p = 10$, and $\varepsilon_m = 1$.

of the experimental solution ranges between 1 and approx. 1.6, and the xy permittivity between 1 and approx. 1.3. Even the largest errors, where the volume fraction is 0.15 and the major semi-axis is 0.46, are less than 5%. Figure 4.34 shows much the same results. The accuracy is excellent up to around an eccentricity of 0.97, corresponding to a major semi-axis of 0.4 for the fixed volume fraction of 0.02. We see that its accuracy decreases rapidly past this point, though, as the particles become ever closer together. Figure 4.35 shows the other end of the scale. The major semi-axis is fixed at 0.15 and the particle volume shrinks with the eccentricity. The accuracy of Maxwell-Garnett is much more comparable in this case, though the periodic approximation still looks to be a little more accurate. Figures 4.36 and 4.37 show equivalent graphs for oblate particles. The errors behave more or less identically up to a volume fraction of 0.15. However, since oblate particles have two major semi-axes, the volume fraction is able to reach 0.3 before the particles touch, and the errors keep growing accordingly. However, the oblate particles actually induce a greater change in the permittivity than the prolate, and the percentage errors at the maximum volume fraction of 0.3 are 8% in xy directions and 2.5% in the z.

On the whole, the periodic approximation to the homogenised effective permittivity looks to be an excellent alternative to traditional theories. It provides an appreciable increase of accuracy for a large range of particle sizes and shapes, provided that the particles are not especially closely packed. It also manages this without being a particularly complicated formula (one of the major advantages of Maxwell-Garnett). If greater accuracy is desired, or the particles become too large and densely packed, we have seen that numerical homogenisation will provide (theoretically) perfect accuracy, under the assumption of periodicity. Indeed, we can also conclude that numerical homogenisation will be the method of choice for particles of arbitrary shape, and should always provide the same level of accuracy with little effort, since the homogenised equations are invariant with respect to dimension and geometry. With a working finite elements model for the cell problem, only the particle geometry needs



(A) (B)

FIGURE 4.37: A comparison of the various permittivities as a function of eccentricity for oblate particles. The volume fraction is fixed at 0.02, along with $\varepsilon_p = 10$ and $\varepsilon_m = 1$.

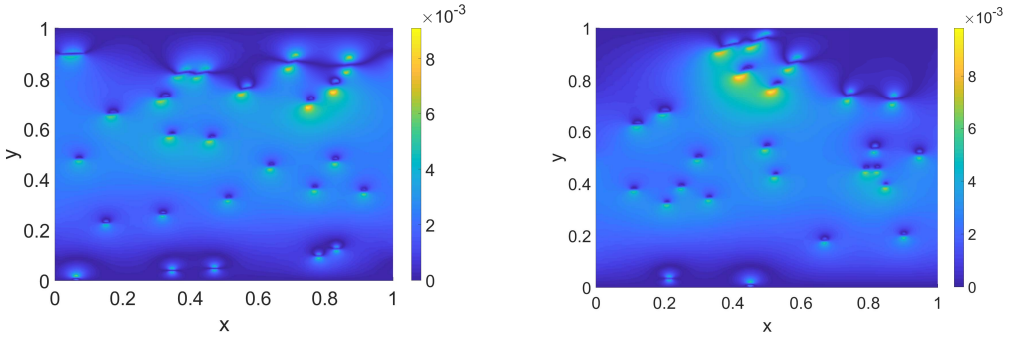
to be changed and the model rerun. In contrast, Maxwell-Garnett (and Bruggeman following on) would be in no way as simple. While the method of Maxwell-Garnett could be applied to an arbitrary particle shape, the electrostatic problem (equivalent to the cell problem) would still need to be solved numerically in order to find the polarisability. There is then complication with the choice of the Lorentz cavity, and how its shape affects the final effective permittivity, and how it should match with the shape of the particle [10], and the correct implementation of Maxwell-Garnett becomes much more unclear.

We note, however, that this is all for particles arranged on a periodic lattice. It is known that Maxwell-Garnett can still provide a good approximation to the effective permittivity for disordered particles, provided that they are distributed, on average over large volumes, homogeneously [34]. This raises clear questions for the validity of homogenisation in such situations, where the key requirement is periodicity. We will, therefore, make a last comparison studying the effective permittivity of disordered particles in 2D (since the number and size of the simulations required mean that it is infeasible in 3D).

4.7 Disordered 2D Particles

4.7.1 Homogenisation Vs Finite Elements

In light of our work on the homogenised low volume fraction approximation, and in [7] studying diffusion in disordered, porous media, we expect that homogenisation will still be feasible at small volume fractions. We could approximate a random system by keeping our grid of unit cells, and placing a single particle at random inside of each cell. Since the solution to the cell problem decays quickly away from the particles, and is small to begin with for small particles, we expect that neighbouring



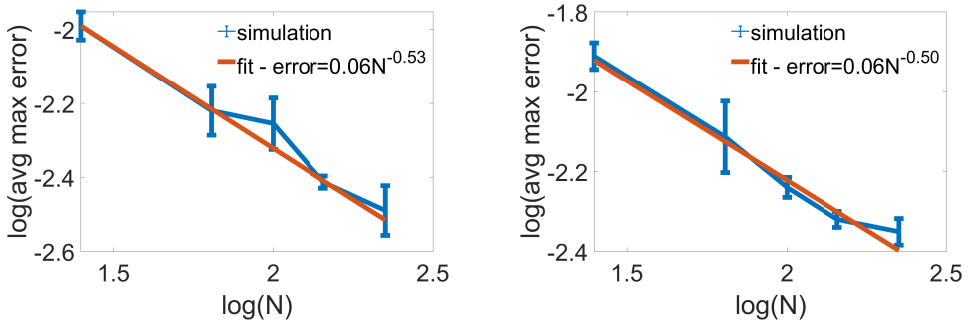
(A) (B)

FIGURE 4.38: The percentage error between the homogenised potential and the microscopic potential for (a) disordered and (b) random particles. The spatial dependence of the system parameters is the same as in figure (4.19), though the particles are not on an ordered lattice this time around.

particles will have little effect on one another, regardless of their positions in their cells. If this is the case, they may as well all be placed in the centre of their cells, and arranged on a periodic lattice, and the homogenised solution should still be valid. Such a setup will be referred to as disordered. In contrast, we will also test fully random systems where the particles are not restricted to individual unit cells.

First we make a direct comparison of the potential, as calculated using the analytic results of section 4.5.5, and a full microscopic simulation using Comsol. Figure 4.38 shows the normalised error between the homogenised and microscopic solutions, calculated as $|\phi_{\text{hom}} - \phi_{\text{micro}}|\,\phi_{\text{micro}}^{-1}$, for disordered and random particles. The spatial dependence of the medium and particles is the same as in figure 4.19, which showed the particles arranged periodically. For periodic particles the maximum error was a little over 6×10^{-3}. In general, the errors grew with increasing volume fraction (increasing y) as the particles became larger and closer together. They also grew with increasing tilt angle (increasing x) as the major axes aligned with the applied field and the γ_3 solution was seen more (which is the component where the largest errors were seen). The disordered and random situations are much the same, with maximum errors around 9×10^{-3} and 10^{-2}. The errors are also, in general, greater for larger y and x. We do, however, note larger than expected errors near $y = 0$ in both graphs, this is where the particles are either more clustered or more isolated than in a periodic setup. This effect can also be observed particularly clearly in the disordered graph around $(0.8, 0.7)$ and the random graph around $(0.5,0.8)$, where there are clusters. The lower of the particles experiences a much larger error at their lower end adjacent to the large void in particles. The errors at all of the particles are not hugely different, though, and so looking at the maximum error seems a suitable measure for the accuracy of the homogenised solution.

The variation in the errors due to the changes in the particle geometries follows as a consequence of the accuracy of the solution to the cell problem and is already



(A) (B)

FIGURE 4.39: Graph showing the average of the maximum normalised error between the homogenised and microscopic solutions for the potential as a function of N, the number of particles, for (a) disordered and (b) random distributions. The error bars show one standard deviation.

understood. We remove these variations to better test the effects of the particle positions, and we study the behaviour of the maximum error as a function of the volume fraction and number of particles. Intuitively, you would expect that having a greater number of particles for a fixed volume fraction would lead to a more homogeneous distribution of particles that can be better approximated by a periodic lattice. We test this idea by simulating circular particles in an isotropic medium. The volume fraction is fixed at 0.01 and the number of particles is ranged from 25 to 400. For each number of particles tested, five different setups of particles are generated (for random and disordered) and the microscopic solution found in Comsol. We then compare this with the homogenised solution and find the average of the maximum normalised error (calculated as $(\phi_{\text{hom}} - \phi_{\text{micro}})\,\phi_{\text{micro}}^{-1}$). The results of this are shown in figure 4.39. We see that both disordered and random systems behave pretty much the same, though we see a larger variance in the random systems, especially for systems with few particles. This is likely due to the greater potential for clusters and voids with fewer particles, and is likely validation of our earlier assumption. As in the periodic case we see that the error decreases as the number of particles increases and the particle distribution becomes more homogenous. Even for systems with as few as 25 particles, the maximum error is still only around 1%. Finally, we also note that the error decreases approximately as $N^{-\frac{1}{2}}$. This is actually the width of each unit cell, and the small parameter η used in the multiple scales expansion of the homogenisation method. We see explicitly, then, how the homogenised solution converges linearly to the true solution as $\eta = N^{-\frac{1}{2}} \to 0$. We conclude that homogenisation, at least at low volume fractions, still provides an excellent approximation for non-periodic arrangments of particles.

We also look at the case where the number of particles is fixed and the volume fraction is varied. The volume fraction is allowed to range up to 0.74, which is just below the maximum volume fraction for packed circles, and we look at the accuracy of

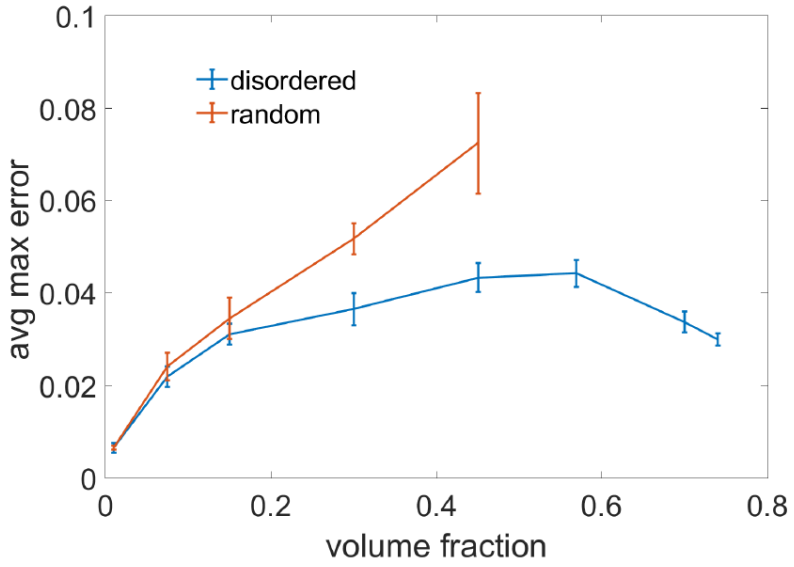
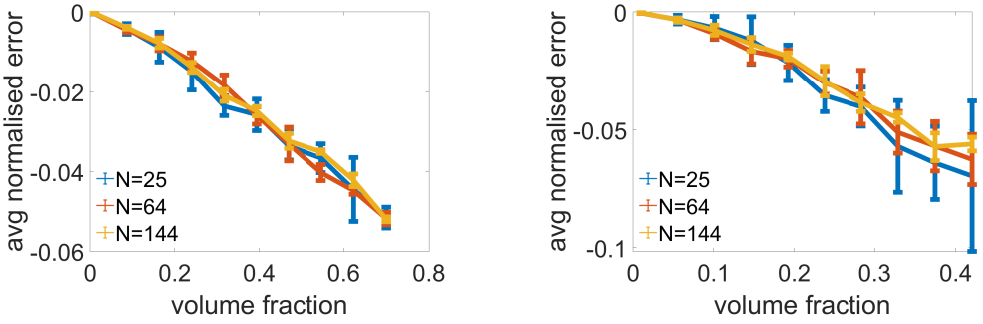


FIGURE 4.40: The average of the maximum normalised error between the homogenised and microscopic potentials as a function of volume fraction. The number of particles is fixed at 64.

numerical homogenisation. The results of this are shown in figure 4.40. As the volume fraction becomes larger and larger, the disordered particles grow to fill the full width of their unit cell, and the system is forced into a packed periodic state. This means that, after an initial decrease in accuracy, the errors in numerical homogenisation begin decreasing with volume fraction. As expected, the random particles do not fare as well, with both larger errors and variances. It also became too time consuming trying to generate random graphs past a volume fraction of 0.5 (with N fixed at 64), since this essentially entails randomly generating a very ordered setup. However, if we were able to generate a random setup to such high volume fractions, they would necessarily have to be more periodic, and we would again see the accuracy increase.

It is important to note that the errors with volume fraction here are not due to any approximations made in solving the cell problem. Indeed, we have used the numerical homogenised solution which solves the full periodic cell problem under no approximations. What we are seeing here is a combination of larger, more densely packed particles interacting (non-periodically), and the errors made in the initial separation of scales in the derivation of the homogenised equations. As the particles become larger, the small parameter η becomes more comparable with the length scale of the macroscopic system. Resultingly, the variations of the averaged homogenised solution away from the microscopic solution become much larger. For larger volume fractions, then, it is also beneficial to consider if homogenisation still captures the average behaviour of the system correctly. We can do this by repeating the results of the previous section, and comparing to an experimental permittivity.

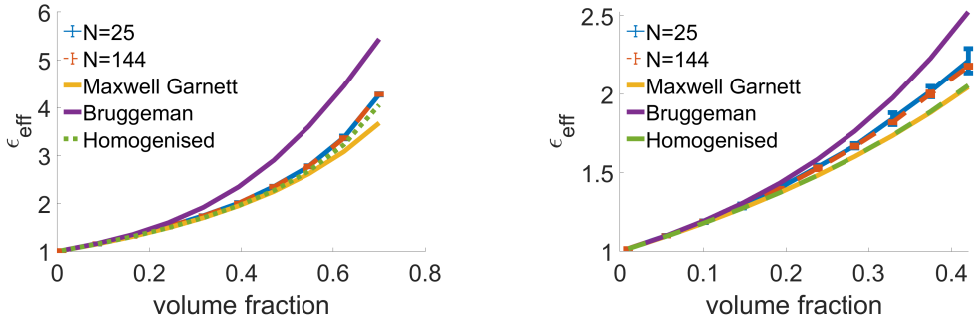


(A) (B)

FIGURE 4.41: Graphs showing the average, normalised error between the homogenised and experimental permittivities for N (a) disordered and (b) random, circular particles in an isotropic medium. The error bars show one standard deviation.

4.7.2 Comparison of Homogenisation with Traditional Theories

Finally, we make a comparison of numerical homogenisation and traditional theories against the theoretical, 2D, experimental result from Comsol. The experimental permittivity will again be found by solving fully for the potential in the presence of particles, calculating the total electromagnetic energy, and then comparing to a parallel plate capacitor. This is done for multiple iterations at a range of volume fractions. Figure 4.41 first shows a comparison of the homogenised and experimental permittivities for circular particles in an isotropic medium. The average normalised error (calculated as$(\varepsilon_{\text{eff}} - \varepsilon_{\text{num}})\,\varepsilon_{\text{num}}^{-1}$) seems to be independent of the number of particles, N. Whilst the number of particles affects the errors in the potential, the errors in the average behaviour of the system (described by ε_{eff}) seems to be dependent only on the volume fraction. What we do see once again, however, is a far larger variance in the errors for smaller numbers of particles. This is again due to the greater chances of larger particles forming clusters and voids, making the system less homogeneous (and less like the periodic homogenised setup). Finally, figure 4.42 shows a comparison of the various effective permittivities, calculated for disordered and random circular particles in an isotropic medium. For periodic particles, we know that the homogenised and experimental solutions are identical. What we see from figure 4.42, then, is that disorder in the particles arrangement actually leads to an increase in the permittivity of the system. It is likely that this must be due to the formation of clusters, since the increase in permittivity is larger for random particles, where clustering is more prominent. For volume fractions smaller than 0.4, the homogenised and Maxwell-Garnett permittivities are essentially identical. Any differences in the two formulations are smaller than the variance in errors in the experimental solution, and so neither can be said to offer better accuracy. On the other hand, as the volume fraction becomes much larger and the particles are forced to become more ordered, we see that the inclusion of the particle interactions in the



(A) (B)

FIGURE 4.42: Graphs showing a comparison of the various effective permittivities for (a) disordered, and (b) random, circular particles in an isotropic medium. The error bars show one standard deviation in the experimental permittivity as calculated for N particles.

homogenised model does offer a marked increase in accuracy. This is, however, for the numerical homogenised permittivity, and so the simplicity of the Maxwell-Garnett formula is definitely preferable for lower volume fractions.

Over all we see that homogenisation can still be used to successfully model the effective permittivity for non-periodic arrangements of particle inclusions. While we have seen that homogenisation offers much greater accuracy in ordered systems, and can do so with analytic forumulae of comparable simplicity to traditional methods, it offers little or no added accuracy in simple systems (circular particles in isotropic media) at lower volume fractions. It could be of interest for further work, however, to explore more complex disordered systems, by introducing anisotropy to the host environment or particle geometries. We have seen that homogenisation, for periodic arrangements, performs much better than traditional theories for non-circular particles. We also have seen that adding anisotropy to the host permittivity is mathematically equivalent to adding anisotropy to the particle shape. It would be of interest, then, to see if the added accuracy of homogenisation in these cases carries over to a measurable improvement once disorder is also added to the system. This could be done in 2D, both numerically, and by applying the 3D periodic approximation methodology to the analytic solution. It could also be done in 3D, however significant time and computer resources would be needed to carry out random simulations of such large numbers of particles. In both cases, generalisations of Maxwell-Garnett and Bruggeman for anisotropic media have been developed (found in, e.g., [10, 53]) and can be used for comparison.

Chapter 5

Conclusions

The work presented in this thesis has covered a wide range of electromagnetism, studying various areas of the interaction of light and matter, both analytically and numerically. The first main area of work presented set out with the goal of determining the photoalignment of thin PAAD films upon exposure to polarised light. While the behaviour of thicker layers on the micron scale is well understood, thinner layers on the nanometer scale, such as those used in liquid crystal alignment layers, were demonstrated in previous experiments to behave differently. New experiments studying the diffraction efficiency of photoaligned gratings formed in these thin PAAD layers were performed by our research group, and the work in this thesis presented a theoretical study of the complex experimental system to try and deduce what we could about the system by comparison with the experimental results. The problems to overcome were varied, requiring an understanding of light propagation in both isotropic and anisotropic media, propagation in layered structures, and diffraction effects. Although each separate topic was based in well-established textbook theory and scientific literature, the final model necessitated the integration of these different effects into a single cohesive model that could be implemented efficiently and quickly for the purposes of minimisation routines.

This lead to the developement of a highly successful new technique, the Iterated Ray method, emerging as a powerful new method for the computation of multilayer propagation. Whilst maintaining the accuracy and stability of the S-matrix formulation, the current leading standard method for multilayer propagation, the Iterated Ray method displays much greater computational efficiency. This proved invaluable while running minimisation routines to obtain system parameters from the experimental results. The method was also extended to multi-mode systems and, coupled with the existing theory on anisotropic propagation and electromagnetic behaviour at a single isotropic/anistropic interface, proved highly accurate in the calculation of transmission across an anistropic layer. Being a purely analytical technique, it also presents a distinct advantage over numerical studies of anisotrpic

layers; it does not suffer from non-physical resonance effects at high angles of incidence, and maintains its accuracy for the full range of incident angles. The Iterated Ray method and its results have been submitted to the Journal of the Optical Society of America A.

The Iterated Ray method enabled us to account for the multiple layers of the system, but the diffraction process itself still needed to be modelled. A non-paraxial scalar diffraction theory based in Fourier and signal analysis was used to effectively describe the diffraction process for a known field emerging from a structure, and to provide an analytic description of the power diverted into the different diffracted orders. The results of the transmission across isotropic and anisotropic layers was used to approximate the transmission of the PAAD grating as a square wave, to be passed to the diffraction theory to model the resulting diffracted orders. While both the extension of the Iterated Ray method to multi-mode propagation and the scalar diffraction theory proved robust against numerical simulations, it seems that the modelling approximation of describing the diffraction grating behaving as a square wave was not able to properly capture the photoalignment behaviour of the thin PAAD layers. As a result, the final model was not able to provide good quantitative agreement with the experimental results. Despite this, it seems that a reasonable estimate of the photoalignment behaviour was found, with some good qualitative agreement being found, and with system parameters agreeing across different experiments, and with known bounds and theory.

The second area of work presented used the method of homogenisation to describe the electromagnetic behaviour of composite systems of host and particle inclusions. Whilst the composite system possesses both microscopic and macroscopic behaviour, the method of homogenisation models the system purely on the macroscopic scale, with the effects of the inclusions accounted for through an effective permittivity, a principle used in many traditional effective medium theories. For particles on an ordered lattice, it is found that the numerical calculation of the homogenised effective permittivity offers excellent results, regardless of particle size or geometry, or anisotropy in the host medium. While the homogenised solution may not be exact pointwise, since it does not account for individual particles and their scattering, it is found to describe the average behaviour of the system at the macroscopic level excellently. This does still require numerical calculation of the cell problem, however, and so analytic approximations were found in the limit of low volume fraction. This was done for 2D spatially varying elliptical particles in anisotropic media, and for 3D spheroidal particles in isotropic media. The analytic results, and also the numerical, were compared to traditional effective medium theories used extensively throughout the literature in such host/inclusion problems. For periodic, ordered systems, homogenisation was found to give an appreciable increase in accuracy over

traditional theories, especially in the presence of shape and material anisotropy, and with analytic formulae of similar complexity to traditional theories.

Finally, the applicability of homogenisation in real world systems where there is disorder in the particles arrangements was studied. Large simulations of the full disordered particle microstructure were carried out, allowing the average behaviour of such systems to be studied. The results were compared against homogenisation and traditional theories for spherical particles in isotropic media. At low volume fractions the various methods were found to be of equal accuracy, with the difference in the analytic solutions being smaller than the variance in the random simulations. In such situations the simplicity of traditional theories is more beneficial. At large volume fractions, as the particles become forced into a more ordered alignment, homogenisation proves itself preferable again, though the numerical solution of the cell problem is required in this case. Beneficial future work could include a study of disordered anisotropic systems, it would be interesting to see if the added accuracy in the periodic case carries over to disordered systems. A possible continuation of this work that may be undertaken by our research group is to use this method to study the optical and alignment properties of surface structures to better control liquid alignment.

References

[1] Rajratan Basu. Soft memory in a ferroelectric nanoparticle-doped liquid crystal. *Physical Review E*, 89(2):022508, 2014.

[2] Jeroen Beeckman, Kristiaan Neyts, and Pieter J. M. Vanbrabant. Liquid-crystal photonic applications. *Optical Engineering*, 50(8):081202–081202–17, 2011. . URL `http://dx.doi.org/10.1117/1.3565046`.

[3] Thomas Paul Bennett. *Multiscale modelling and experimental estimation of liquid crystals parameters*. PhD thesis, Mathematical Sciences, 2016.

[4] P-A Blanche, Ph C Lemaire, M Dumont, and M Fischer. Photoinduced orientation of azo dye in various polymer matrices. *Optics letters*, 24(19):1349–1351, 1999.

[5] Max Born and Emil Wolf. *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*. Elsevier, 2013.

[6] DAG Bruggeman. Calculation of different physical constants of heterogen substances i dielectric constants and conductibility of mixtures from isotrop substances. *Annalen der Physik*, 24(8):665–679, 1935.

[7] Maria Bruna and S Jonathan Chapman. Diffusion in spatially varying porous media. *SIAM Journal on Applied Mathematics*, 75(4):1648–1674, 2015.

[8] SV Burylov and Yu L Raikher. Macroscopic properties of ferronematics caused by orientational interactions on the particle surfaces. i. extended continuum model. *Molecular Crystals and Liquid Crystals Science and Technology. Section A. Molecular Crystals and Liquid Crystals*, 258(1):107–122, 1995.

[9] Vladimir G Chigrinov, Vladimir M Kozenkov, and Hoi-Sing Kwok. *Photoalignment of liquid crystalline materials: physics and applications*, volume 17. John Wiley & Sons, 2008.

[10] Tuck C. Choy. *Effective Medium Theory Princioples and Applications, Second Edition*. Oxford Univeristy Press, 2016.

[11] Michel L. Dumont, Sophie Hosotte, G. Froc, and Zouheir Sekkat. Orientational manipulation of chromophores through photoisomerization, 1994. URL http://dx.doi.org/10.1117/12.166297.

[12] R. J. Elliott, J. A. Krumhansl, and P. L. Leath. The theory and properties of randomly disordered crystals and related physical systems. *Rev. Mod. Phys.*, 46: 465–543, Jul 1974. . URL https://link.aps.org/doi/10.1103/RevModPhys.46.465.

[13] R Feynman, R Leighton, and M Sands. Feynman lectures on physics, vol ii, part 2: Mainly electromagnetism and matter, 1974.

[14] Jian Gao, Yaning He, Fang Liu, Xi Zhang, Zhiqiang Wang, and Xiaogong Wang. Azobenzene-containing supramolecular side-chain polymer films for laser-induced surface relief gratings. *Chemistry of materials*, 19(16):3877–3881, 2007.

[15] Jack D Gaskill. *Linear systems, Fourier transforms, and optics*, volume 576. Wiley New York, 1978.

[16] James E Harvey, Cynthia L Vernold, Andrey Krywonos, and Patrick L Thompson. Diffracted radiance: a fundamental quantity in nonparaxial scalar diffraction theory. *Applied Optics*, 38(31):6469–6481, 1999.

[17] Johan Hoogboom, Paula M. L. Garcia, Matthijs B. J. Otten, Johannes A. A. W. Elemans, Joseph Sly, Sergiy V. Lazarenko, Theo Rasing, Alan E. Rowan, and Roeland J. M. Nolte. Tunable command layers for liquid crystal alignment. *Journal of the American Chemical Society*, 127(31):11047–11052, 2005. . URL http://dx.doi.org/10.1021/ja0518651. PMID: 16076212.

[18] Che-Ju Hsu, Bao-Long Chen, and Chi-Yen Huang. Controlling liquid crystal pretilt angle with photocurable prepolymer and vertically aligned substrate. *Opt. Express*, 24(2):1463–1471, Jan 2016. . URL http://www.opticsexpress.org/abstract.cfm?URI=oe-24-2-1463.

[19] Kunihiro Ichimura. Photoalignment of liquid-crystal systems. *Chemical Reviews*, 100(5):1847–1874, 2000. . URL http://dx.doi.org/10.1021/cr980079e. PMID: 11777423.

[20] Kunihiro Ichimura. Photoalignment of liquid-crystal systems. *Chemical reviews*, 100(5):1847–1874, 2000.

[21] Keigo Iizuka. *Elements of Photonics, Volume I: In Free Space and Special Media*, volume 41. John Wiley & Sons, 2002.

[22] Shie-Chang Jeng and Shug-June Hwang. Controlling the alignment of polyimide for liquid crystal devices, high performance polymers - polyimides based - from

chemistry to applications. https://www.intechopen.com/books/high-performance-polymers-polyimides-based-from-chemistry-to-applications/controlling-the-alignment-of-polyimide-for-liquid-crystal-devices, 2012. . URL `https://www.intechopen.com/books/high-performance-polymers-polyimides-based-from-chemistry-to-applications/controlling-the-alignment-of-polyimide-for-liquid-crystal-devices`.

[23] R Clark Jones. A generalization of the dielectric ellipsoid problem. *Physical Review*, 68(3-4):93, 1945.

[24] Joel Kauppo. Photoalignment control and surface-relief gratings in commercial azobenzene-based photoalignment layers. Master's thesis, Aalto University, School of Science, 2014.

[25] Joseph B Keller. Effective behavior of heterogeneous media. In *Statistical mechanics and statistical methods in theory and application*, pages 631–644. Springer, 1977.

[26] Ammar A. Khan, G. Dinesha M. R. Dabera, Haider Butt, Malik M. Qasim, Gehan A. J. Amaratunga, S. Ravi P. Silva, and Timothy D. Wilkinson. Tunable scattering from liquid crystal devices using carbon nanotubes network electrodes. *Nanoscale*, 7:330–336, 2015. . URL `http://dx.doi.org/10.1039/C4NR04466E`.

[27] DY Kim, SK Tripathy, Lian Li, and J Kumar. Laser-induced holographic surface relief gratings on nonlinear optical polymer films. *Applied Physics Letters*, 66(10): 1166–1168, 1995.

[28] David Yuk Kei Ko and JC Inkson. Matrix method for tunneling in heterostructures: Resonant tunneling in multilayer systems. *Physical Review B*, 38 (14):9945, 1988.

[29] Tobias AF Konig, Petr A Ledin, Justin Kerszulis, Mahmoud A Mahmoud, Mostafa A El-Sayed, John R Reynolds, and Vladimir V Tsukruk. Electrically tunable plasmonic behavior of nanocube–polymer nanomaterials induced by a redox-active electrochromic polymer. *ACS nano*, 8(6):6182–6192, 2014.

[30] CH Legge and GR Mitchell. Photo-induced phase transitions in azobenzene-doped liquid crystals. *Journal of Physics D: Applied Physics*, 25(3):492, 1992.

[31] John Lekner. Reflection and refraction by uniaxial crystals. *Journal of Physics: Condensed Matter*, 3(32):6121, 1991.

[32] Ohad Levy and David Stroud. Maxwell garnett theory for mixtures of anisotropic inclusions: Application to conducting polymers. *Physical Review B*, 56 (13):8035, 1997.

[33] Lena M Lopatina and Jonathan V Selinger. Theory of ferroelectric nanoparticles in nematic liquid crystals. *Physical review letters*, 102(19):197802, 2009.

[34] Vadim A Markel. Introduction to the maxwell garnett approximation: tutorial. *JOSA A*, 33(7):1244–1256, 2016.

[35] Eleni Mavrona, Sakellaris Mailis, Nina Podoliak, Giampaolo D'Alessandro, Nelson Tabiryan, Maria Trapatseli, J-F Blach, Malgosia Kaczmarek, and Vasileios Apostolopoulos. Intrinsic and photo-induced properties of high refractive index azobenzene based thin films. *Optical Materials Express*, 8(2):420–430, 2018.

[36] Milan Milosevic. On the nature of the evanescent wave. *Applied spectroscopy*, 67 (2):126–131, 2013.

[37] Robert J Moerland and Jacob P Hoogenboom. Subnanometer-accuracy optical distance ruler based on fluorescence quenching by transparent conductors. *Optica*, 3(2):112–117, 2016.

[38] P.H. Moon and D.E. Spencer. *Field Theory for Engineers*. Van Nostrand series in electronics and communications. Van Nostrand, 1961. URL `https://books.google.co.uk/books?id=qF03AAAAIAAJ`.

[39] Alexander Moroz. Depolarization field of spheroidal particles. *JOSA B*, 26(3): 517–527, 2009.

[40] Fiore Pasquale Nicoletta, Daniela Cupelli, Patrizia Formoso, Giovanni De Filpo, Valentina Colella, and Annarosa Gugliuzza. Light responsive polymer membranes: A review. *Membranes*, 2(1):134–197, 2012.

[41] G. A. Pavliotis and Andrew Stuart. *Multiscale Methods - Averaging and Homogenisation*. Springer New York, 2008. URL `http://www.ebook.de/de/product/24984234/g_a_pavliotis_andrew_stuart_multiscale_methods.html`.

[42] Grzegorz Pawlik, Tomasz Wysoczanski, and Antoni C Mitus. Complex dynamics of photoinduced mass transport and surface relief gratings formation. *Nanomaterials*, 9(3):352, 2019.

[43] Eleni Perivolari, JR Gill, N Podoliak, Vasileios Apostolopoulos, TJ Sluckin, Giampaolo D'Alessandro, and Malgosia Kaczmarek. Optically controlled bistable waveplates. *Journal of Molecular Liquids*, 2017.

[44] Nina Podoliak, Oleksandr Buchnev, Mark Herrington, Elena Mavrona, Malgosia Kaczmarek, Antonios G Kanaras, Emmanuel Stratakis, Jean-François Blach, Jean-François Henninot, and Marc Warenghem. Elastic constants, viscosity and response time in nematic liquid crystals doped with ferroelectric nanoparticles. *Rsc Advances*, 4(86):46068–46074, 2014.

[45] Arri Priimagi and Andriy Shevchenko. Azopolymer-based micro-and nanopatterning for photonic applications. *Journal of Polymer Science Part B: Polymer Physics*, 52(3):163–182, 2014.

[46] Arri Priimagi, Jaana Vapaavuori, Francisco J Rodriguez, Charl FJ Faul, Markku T Heino, Olli Ikkala, Martti Kauranen, and Matti Kaivola. Hydrogen-bonded polymer- azobenzene complexes: Enhanced photoinduced birefringence with high temporal stability through interplay of intermolecular interactions. *Chemistry of Materials*, 20(20):6358–6363, 2008.

[47] C.V. Raman and N.S. Nath. The diffraction of light by sound waves of high frequency: Part i. *Proc. Indian. Acad Sci. A2 406-412*, 1936.

[48] JA Ratcliffe. Some aspects of diffraction theory and their application to the ionosphere. *Reports on progress in physics*, 19(1):188, 1956.

[49] Bendaoud Saad. Linearly and circularly polarized laser photoinduced molecular order in azo dye doped polymer films. In *EPJ Web of Conferences*, volume 139, page 00012. EDP Sciences, 2017.

[50] Sergii M Shelestiuk, Victor Yu Reshetnyak, and Timothy J Sluckin. Frederiks transition in ferroelectric liquid-crystal nanosuspensions. *Physical Review E*, 83(4): 041705, 2011.

[51] A.E. Siegman. *Lasers*. University Science Books, 1986. ISBN 9780935702118. URL `https://books.google.co.uk/books?id=1BZVwUZLTkAC`.

[52] Ari Sihvola. Homogenization of a dielectric mixture with anisotropic spheres in anisotropic background. *Electromagnetics*, 17(3):269–286, 1997.

[53] Ari H Sihvola. *Electromagnetic mixing formulas and applications*. Number 47. Iet, 1999.

[54] Erik S Thomson, Larry A Wilen, and John S Wettlaufer. Light scattering from an isotropic layer between uniaxial crystals. *Journal of Physics: Condensed Matter*, 21 (19):195407, 2009.

[55] Lloyd Trefethen. *Spectral Methods in Matlab*. SIAM, 2000.

[56] Jaana Vapaavuori, C Geraldine Bazuin, and Arri Priimagi. Supramolecular design principles for efficient photoresponsive polymer–azobenzene complexes. *Journal of Materials Chemistry C*, 6(9):2168–2188, 2018.

[57] Franklin F.W. Wang. *Introduction to Solid State Electronics (Second Edition)*. Elsevier, 1989.

[58] Eric W. Weisstein. "elliptic cylindrical coordinates." from mathworld–a wolfram web resource. URL https://mathworld.wolfram.com/EllipticCylindricalCoordinates.html.

[59] Deng-Ke Yang and Shin-Tson Wu. *Fundamentals of Liquid Crystal Devices*. JOHN WILEY & SONS INC, 2014. ISBN 1118752007. URL http://www.ebook.de/de/product/22749941/deng_ke_yang_shin_tson_wu_fundamentals_of_liquid_crystal_devices.html.

[60] Oleg Yaroshchuk and Yuriy Reznikov. Photoalignment of liquid crystals: basics and current trends. *Journal of Materials Chemistry*, 22(2):286–300, 2012.

[61] Fion SY Yeung, YL Jacob Ho, Yuet Wing Li, and Hoi Sing Kwok. Liquid crystal alignment layer with controllable anchoring energies. *Journal of display technology*, 4(1):24–27, 2008.

[62] Jingli Zhou, Jing Shen, Jianjun Yang, Yan Ke, Keyi Wang, and Qijin Zhang. Fabrication of a pure polarization grating in a cross-linked azopolymer by polarization-modulated holography. *Optics letters*, 31(10):1370–1372, 2006.