WAVEGUIDE ELECTRO-OPTIC MODULATORS
USING THE PARALLEL-PLATE CONFIGURATION

Henry Robinson

Department of Electrical and Electronic Engineering,
University College, London

A thesis submitted to University College, University of London
for the degree of Doctor of Philosophy

November 1991
"It is a tale
Told by an idiot, full of sound and fury,
Signifying nothing."

Macbeth
Acknowledgements

This project was funded by the Science and Engineering Research Council, and by GEC research (Gt Baddow, Essex) under a CASE studentship; also by the Department of Trade and Industry under the Devices programme of the Joint Opto-Electronic Research Scheme.

I should like to thank the following people for their contributions to this work:

Dr. J.E.U. Ashton and J. Reed, of E.E.V. Co., Chelmsford, for preparing silicon nitride waveguides;

Dr. P.J. Storey, of GEC Hirst Research Laboratories, for preparing silicon-on-sapphire wafers;

Dr. S. Singh and N. Rice, of BP Research Laboratories, for providing amorphous silicon layers;

Professor P.G. LeComber and Dr. R. Gibson, of Dundee University, for providing amorphous silicon layers;

Dr. J. Milledge, of University College, London, for instruction and assistance in X-ray analysis;

G. Critchlow, of Loughborough Consultants, for Auger spectroscopic analysis.

I am grateful to my colleagues at UCL tolerating me, and for providing all kinds of help from verbal to practical: Alan Gorrod and Chris Bunton who contributed many hours of delicate work on the sputtering substrate table, and who, with Maurice Gillett, Alan Drury, Jim Van Sickle, Eric Wheelhouse, and Richard Lewis, built any apparatus too special to be bought or bodged; Tony Overbury, for his expert help with the infrared optical experiments; Fred Stride, Kevin Lee and Fay Morris for providing help in the clean room; Steve Reynolds, Roger Beswick, Maria Fillieul, Mario Petrucci, Chris Watson, and Andy James for providing discussion both entertaining and scientific. In particular, I should like to thank Mario Petrucci for his patience through many hours of spirited discussions.
I am particularly grateful to my supervisor, Professor Chris Pitt, for taking me on in the first place, and then providing scientific discussion, criticism and encouragement when it was needed.

I am deeply grateful to Stewart and Catherine Boyd, for accommodating me for the last three years.

I should like to thank Vaughan Kitchin, who goaded me into starting in the first place.
Abstract

Conventional waveguide electro-optic modulators use the fringing fields below a gap between coplanar electrodes. These fields are applied to a waveguide which is fabricated in an electro-optic substrate. The aim of the present work was to develop a device with an inherently more efficient geometry, viz. the parallel-plate configuration.

A computer model was developed and tested, to calculate optical propagation coefficients and losses in multilayer waveguides incorporating optically absorbing materials such as metals.

Both electro-optic and electrical characteristics were modelled, for a variety of parallel-plate structures. It was found that metallic electrodes engendered unfeasibly high losses via surface plasmons. A novel structure was devised, using silicon electrodes above and below a thin film of lithium niobate. High losses were avoided by operating the structure above the band gap wavelength. Remaining losses were due to free carrier absorption. Based on this scheme, a family of efficient modulator structures was designed.

RF sputtering was used to grow thin films of lithium niobate for the electro-optic layer. The growth characteristics were studied, and a regime was established for growing lithium niobate with a preferred z-orientation as indicated by X-ray crystallography.

A fabrication sequence and a mask set were designed, with a view to fabricating test modulators. As well as the logistical problems, structural difficulties arose from the combination of silicon and lithium niobate: these were identified and overcome, and test devices were fabricated.

These test structures were electrically and optically assessed, and the electro-optic effect was measured.

In the final chapter, the project is reviewed. There is a discussion of the feasibility and implications of combining these structures with silicon integrated circuits. An assessment is made of the contribution to integrated optics.
# Table of contents

<table>
<thead>
<tr>
<th>Acknowledgements</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>5</td>
</tr>
<tr>
<td>Table of contents</td>
<td>6</td>
</tr>
<tr>
<td>List of tables</td>
<td>9</td>
</tr>
<tr>
<td>List of illustrations</td>
<td>10</td>
</tr>
</tbody>
</table>

## Chapter 1: Introduction

1. Integrated optics  
   1.1. Integrated optics vs. electronics  
   1.2. Interconnects in data processing systems  
   1.3. Integrated Optical interconnects  
     1.3.1. Advantages and disadvantages  
     1.3.2. Implementations of integrated optical interconnects  
     1.3.2.1. Free space propagation  
     1.3.2.2. Guided wave propagation  
   1.4. Optical waveguide modulation  
     1.4.1. Directly modulated lasers  
     1.4.2. Waveguide modulators  
   1.5. Optical modulation: interactions and materials  
   1.6. Lithium niobate device structures  
   1.7. The present project  
   1.8. Summary  

References  

## Chapter 2: Computer Modelling

2. The multilayer waveguide model, MOD2  
   2.1. MOD2: mathematical basis  
   2.1.2. MOD2: strategy for finding solutions  
   2.1.3. MOD2: birefringence  
   2.1.4. MOD2: implementation  
   2.1.5. MOD2 results: comparison to literature  
   2.1.6. MOD2 results: comparison to experiment  
   2.1.7. MOD2: scope and limitations  
   2.2. Electrical modelling: VMEFF  
     2.2.1. VMEFF: mathematical basis  
     2.2.2. VMEFF: implementation
Chapter 3: Device modelling

3.1. Figures of merit
   3.1.1. The active length, \(L_a\), and optical efficiency, \(\gamma\)
   3.1.2. Optical attenuation
   3.1.3. Bandwidth

3.2. Optical investigations
   3.2.1. Metallic electrodes
   3.2.2. Low-index buffer layers
   3.2.3. Structures employing silicon electrodes
   3.2.4. Effects of silicon on optical modulation

3.3. Electrical modelling
   3.3.1. Optical absorption vs. electrode resistivity
   3.3.2. Electrode depletion and accumulation

3.4. Test structure design

3.5. Summary

Chapter 4: Lithium niobate thin film deposition

4.1. Programme philosophy

4.2. Lithium niobate
   4.2.1. Lithium niobate thin films
   4.2.2. Poling

4.3. RF Sputter deposition
   4.3.1. Substrate DC Bias
   4.3.2. Substrate heating

4.4. Growth studies
   4.4.1. Film growth
   4.4.2. Optical measurements
   4.4.3. Profilometry and thickness measurements
   4.4.4. Electrical properties
   4.4.5. Pinholes: etching experiment
   4.4.6. X-ray crystallography
   4.4.7. Compositional analysis

References
4.4.8. Scanning Electron Microscopy (SEM) 124

4.5. Summary 125
References 126

Chapter 5: Materials & process development 131
5.1. Materials & processes 131
  5.1.1. The silicon layers 131
    5.1.1.1 PECVD hydrogenated amorphous silicon 132
    5.1.1.2. Silicon-On-Sapphire (SOS) 136
  5.1.2. The lithium niobate layer 137
  5.1.3. Oxide buffer layers 139
5.2. Device fabrication 139
  5.2.1. The process sequences 139
  5.2.2. Final preparations for optical testing 145
5.3. Device design: Photomasks 146
5.4. Summary 151
References 151

Chapter 6: Optical experiments 154
6.1. Inclined crystallographic axes 154
  6.1.1. Electro-optic coefficients 155
  6.1.2. Mode conversion 161
  6.1.3. Inclined crystallographic axes: summary 163
6.2. Measurements: experimental arrangement 164
  6.1.2. Measurement of phase modulation 167
6.3. Measurements: experimental results 173
  6.3.1. Polarisation 173
  6.3.2. Launch positioner 175
  6.3.3. DC response 177
  6.3.4. Frequency response 179
6.4. Conclusions 181
References 182

Chapter 7: Conclusions 184
7.1. Achievements of this work 184
7.2. Parallel-plate modulators and VLSI: feasibility 186
7.3. Further work 188
7.3.1. Device improvements

7.3.1.1. Materials

7.3.1.2. Device geometry

7.3.2. Novel applications and implementations

7.3.3. Other devices

7.4. Conclusions

References

Appendices

2A. MOD2 program organisation

3A. Refractive indices and electro-optic effect in lithium niobate

3B. Optical effects of free carriers

4A. RF sputtering

4B. Geometry of X-ray diffraction

4C. X-ray diffraction photographs of rhombohedral lithium niobate

5A. Ion beam milling

6A. Calculation of the electro-optic effect in <012> oriented lithium niobate

List of tables

2.1. VMEFF: results for typical modulator structure

3.1. Characteristics of surface plasmons

3.2. Electro-optic contributions of various layers of the modulator

5.1. The parallel-plate modulator fabrication sequences

6.1. Direction cosines for conversion between waveguide and crystal axes

6.2. Modulation per volt, of each mode of the waveguide, due to the electro-optic effect in the lithium niobate layer

6.3. Electro-optic modulation contributions of various layers of the modulator for each guided mode
# List of illustrations

<table>
<thead>
<tr>
<th>Chapter 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1. HOEs for interconnects</td>
</tr>
<tr>
<td>1.2. WSI interconnects</td>
</tr>
<tr>
<td>1.3. Electro-optic modulation - schematic</td>
</tr>
<tr>
<td>1.4. Δβ waveguide coupler</td>
</tr>
<tr>
<td>1.5. TIR waveguide coupler</td>
</tr>
<tr>
<td>1.6. Mach-Zehnder interferometer</td>
</tr>
<tr>
<td>1.7. Cutoff modulator</td>
</tr>
<tr>
<td>1.8. Thinned-substrate parallel plate waveguide modulator</td>
</tr>
<tr>
<td>1.9. Proposed thin film modulator structure</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1. Optical waveguide - schematic diagram</td>
</tr>
<tr>
<td>2.2. MOD2: axes</td>
</tr>
<tr>
<td>2.3. Plot of R vs. real &amp; imaginary components of β/κo</td>
</tr>
<tr>
<td>2.4. MOD2: Flow diagram</td>
</tr>
<tr>
<td>2.5. Lossy waveguide characteristics</td>
</tr>
<tr>
<td>2.6. Silicon nitride waveguides: β/κo graph</td>
</tr>
<tr>
<td>2.7. Leaky waveguide: spiralling modal distribution</td>
</tr>
<tr>
<td>2.8. Silicon nitride waveguides: loss graph</td>
</tr>
<tr>
<td>2.9. VMEFF: modelled structure</td>
</tr>
<tr>
<td>2.10. VMEFF: electric fields at various frequencies</td>
</tr>
<tr>
<td>2.11. Predicted characteristics vs frequency</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1. Surface plasmons: modal plots</td>
</tr>
<tr>
<td>3.2. Ag/LiNbO3/Ag: TM characteristics</td>
</tr>
<tr>
<td>3.3. Ag/LiNbO3/Ag: TE characteristics</td>
</tr>
<tr>
<td>3.4. Ag/LiNbO3/Ag: modal profiles</td>
</tr>
<tr>
<td>3.5. Ag/SiO2/LiNbO3/SiO2/Ag: TM characteristics</td>
</tr>
<tr>
<td>3.6. Ag/SiO2/LiNbO3/SiO2/Ag: TE characteristics</td>
</tr>
<tr>
<td>3.7. Ag/SiO2/LiNbO3/SiO2/Ag: modal profiles</td>
</tr>
<tr>
<td>3.8. Characteristics of thick silicon-electrode modulators</td>
</tr>
<tr>
<td>3.9. Characteristics of silicon-electrode modulators vs thickness of upper electrode</td>
</tr>
</tbody>
</table>
3.10. Characteristics of silicon-electrode modulators vs thickness of lithium niobate layer 81
3.11. Lπ vs thickness of lithium niobate layer at 5GHz 84
3.12. Free carrier absorption & resistivity (bulk silicon) vs. doping density 86
3.13. Depletion effect: schematic diagram 88
3.14. Predicted effect of depletion on modulation 89

Chapter 4
4.1. Phase diagram of lithium niobate 95
4.2. Crystalline structure of rhombohedral lithium niobate 97
4.3. RF sputtering apparatus: schematic diagram 99
4.4. RF matching unit: circuit diagram 101
4.5. Substrate table 103
4.6. Method of film thickness and refractive index measurement by optical waveguiding 106
4.7. Effects of index and thickness variations on optical waveguide measurements 107
4.8. Birefringence vs. temperature 108
4.9. Method of Talystep measurement using ion beam milled edges 109
4.10. Talystep: film profiles & roughnesses 110
4.11. Deposition rate vs RF power 111
4.12. Length of step coverage - test capacitors & modulators 112
4.13. IV characteristics of lithium niobate thin film capacitor 113
4.14. Pinholes experiment: (micrograph) 114
4.15. X-ray diffraction photographs of deposited films 116-117
4.16. X-ray diffraction photograph of lithium niobate film deposited at high RF power 118
4.17. X-ray oscillation and Weissenberg photographs of lithium niobate film on <012> sapphire 119-120
4.18. X-ray diffraction photograph of lithium niobate film deposited on SOS 121
4.19. Haloes on lithium niobate films, in the neighbourhood of underlying silicon and silica features 122
4.20. Auger spectrum 123
4.21. Auger profile of lithium niobate sputtered film 124
4.22. Scanning electron micrograph of thick lithium niobate film 125
### Chapter 5

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1. Optical absorption processes in silicon</td>
<td>133</td>
</tr>
<tr>
<td>5.2. Thermal delamination (micrograph)</td>
<td>135</td>
</tr>
<tr>
<td>5.3. Ion beam milling: etch rates</td>
<td>138</td>
</tr>
<tr>
<td>5.4. Fabrication process sequence (I - simple)</td>
<td>143</td>
</tr>
<tr>
<td>5.5. Fabrication process sequence (II - with silica buffers)</td>
<td>144</td>
</tr>
<tr>
<td>5.6. Scanning electron micrograph of cleaved modulator</td>
<td>145</td>
</tr>
<tr>
<td>5.7. Finished device (micrograph)</td>
<td>146</td>
</tr>
<tr>
<td>5.8. Merlin: modulator layout</td>
<td>147</td>
</tr>
<tr>
<td>5.9. Vernier alignment boxes</td>
<td>148</td>
</tr>
<tr>
<td>5.10. Merlin: chip layout</td>
<td>149</td>
</tr>
<tr>
<td>5.11. Test capacitors</td>
<td>150</td>
</tr>
<tr>
<td>5.12. Step coverage problems</td>
<td>150</td>
</tr>
</tbody>
</table>

### Chapter 6

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1. Crystal axes and waveguide orientation</td>
<td>157</td>
</tr>
<tr>
<td>6.2. Effect of electric field on index ellipsoid</td>
<td>158</td>
</tr>
<tr>
<td>6.3. Crystal axes: transformation</td>
<td>159</td>
</tr>
<tr>
<td>6.4. Inclined ellipsoids &amp; TE-TM mode conversion</td>
<td>162</td>
</tr>
<tr>
<td>6.5. Optical bench layout</td>
<td>165</td>
</tr>
<tr>
<td>6.6. Optical arrangement for end-firing and signal collection</td>
<td>166</td>
</tr>
<tr>
<td>6.7. End-firing apparatus (photographs)</td>
<td>167</td>
</tr>
<tr>
<td>6.8. TE-TM interference: schematic diagram</td>
<td>169</td>
</tr>
<tr>
<td>6.9. TM0-TM1 interference: schematic diagram</td>
<td>171</td>
</tr>
<tr>
<td>6.10. Radiation patterns, TM0 and TM1</td>
<td>172</td>
</tr>
<tr>
<td>6.11. Radiation patterns for various phase shifts</td>
<td>173</td>
</tr>
<tr>
<td>6.12. Measured signal vs. half wave plate position</td>
<td>174</td>
</tr>
<tr>
<td>6.13. Measured signal vs. detection polariser position</td>
<td>175</td>
</tr>
<tr>
<td>6.14. Measured signal vs. launch objective position</td>
<td>176</td>
</tr>
<tr>
<td>6.15. Measured signal vs. DC bias</td>
<td>178</td>
</tr>
<tr>
<td>6.16. Theoretical device output vs. DC bias</td>
<td>178</td>
</tr>
<tr>
<td>6.17. Leakage current of device</td>
<td>179</td>
</tr>
<tr>
<td>6.18. Measured signal vs. frequency</td>
<td>180</td>
</tr>
<tr>
<td>6.19. Effect of visible illumination on frequency response</td>
<td>180</td>
</tr>
</tbody>
</table>

### Chapter 7

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1. Combined modulator &amp; passive waveguide</td>
<td>187</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>7.2. Top-contacted device</td>
<td>190</td>
</tr>
<tr>
<td>7.3. Buffered lower silicon edges</td>
<td>191</td>
</tr>
<tr>
<td>7.4. Device with extended lower electrode</td>
<td>191</td>
</tr>
<tr>
<td>7.5. Characteristics of TE modulator using epitaxial lithium niobate on sapphire</td>
<td>192</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

Abstract

This thesis concerns the design and fabrication of an integrated optical device; the primary aim of this chapter is to place the work in context. The chapter outlines the ethos of integrated optics, focusing in particular on its role in the field of interconnects for electronic systems. At high frequencies, the performance of electronic circuits becomes limited by the rate at which signals can be transmitted from one part of the circuit to another. Integrated optics offers a competitive method of tackling this problem not only over long distances, where optical fibre communications are already firmly established, but also over shorter distances (down to a few centimetres). The devices reported in this thesis are intended primarily for this purpose: they are waveguide modulators, which may be used for imparting a signal to a guided optical carrier. It is argued that there is a potential demand for a compact and efficient device for performing this task, capable of being integrated with electronic circuits both in terms of fabrication and performance.

The research reported here arose from a recent programme of materials research at UCL[^1-^2]: it is explained how the motivation for this work was to capitalise on advances made during that research, by designing and fabricating a device which exploited the advantages of the new materials.

In the latter part of the chapter, attention is focused on the project itself, and a preview of the research is given. It is explained how the project developed, how some early theoretical results led to an opportunity to steer it towards the field of semiconductor devices, and how the research was carried through to a successful conclusion.

1.1. Integrated optics

The term "integrated optics" refers to the combination of several optical functions within a single device. Historically, the field emerged in the 1960s[^3] and was largely based on the concept of optical waveguiding[^4], in which a beam of light may be guided, e.g. along a thin dielectric film deposited on a substrate of lower refractive index. By suitably patterning the film, e.g. by etching it into a stripe, light may be
confined to a narrow channel\textsuperscript{[4]}. This allowed light to be routed between sources and detectors; it also implied concentration of the optical beam within a small cross-sectional area, so that compact electro-optic modulators\textsuperscript{[5,6]} could be realised, with a concomitantly low drive power.

The concept of an all-optical integrated circuit suffers from a severe problem in the difficulty of fabricating wholly optical active components, because of the tendency of light beams to pass through one another without interacting. This is not to say that interaction between several light beams is impossible: however, it requires an intermediate process such as photorefraction. The most promising solution to this problem is a class of structures known as self-electro-optic effect devices (SEEDs), in which a primary light beam generates electron-hole pairs in a semiconductor, and the resulting change in conductivity is used to alter the voltage applied to an electro-optic device which in turn modulates a secondary optical beam\textsuperscript{[7]}. There are two main problems with these devices. The first is that the intermediate process usually involves the movement of charge carriers so that the bandwidth becomes limited by electronic considerations. The notable exception to this is a class of quantum well devices which rely on the quantum-confined Stark effect (QCSE)\textsuperscript{[8]}, where an exciton absorption peak is altered by the application of an electric field. The ultimate limit on the speed of this effect is believed to be better than 1ps\textsuperscript{[7]}. However, it should be pointed out that this refers to the rapidity with which the absorption can be switched on: the switch-off, or relaxation time, involves charge transport to remove the photogenerated carriers, and is typically no better than 1ns or so. The second problem with SEEDs is that the packing density is very low, by a factor of over 100, compared to electronic devices\textsuperscript{[7,9]}.

1.1.1. Integrated optics and electronics

Although integrated optics on its own is unlikely to compete with electronic ICs as an alternative technology for data processing, it is clear that it offers advantages in certain respects (speed, bandwidth, free space propagation). There has been a great deal of interest in exploiting these advantages by using integrated optics to provide high performance interconnects for electronic ICs.

The primary motivation behind this concept lies in the inherently high speed and bandwidth of light as compared to electronic signals. Light may be viewed as a carrier wave with a frequency of the order of $10^{14}$Hz. Modulation at $10^{10}$Hz, which is near the limit of electronic capabilities\textsuperscript{[10]}, represents only a small fraction of this carrier frequency and may therefore be treated as a narrow-bandwidth signal. The alternative
way of viewing this is to observe that the modulation frequency may be raised by at least a couple of orders of magnitude before it begins to broaden significantly the spectral linewidth of the light.

1.2. Interconnects in data processing systems

The speed of a data processing system is a figure of merit of obvious importance. This speed is determined by the operating frequency and by the number of information channels being processed in parallel[11].

The maximum operating frequency is limited not only by the speed of individual circuit elements, but also by the limitations of the interconnect network which links different parts of the circuit[10,11,12,13,14]. By far the simplest and cheapest interconnect system uses metal tracks and wires. At high frequencies (beyond $10^9$ Hz or so), these suffer from problems such as signal attenuation, propagation delays, and crosstalk. The first two of these problems may be alleviated to some extent by cramming more circuit elements into denser layouts, but this exacerbates the crosstalk problem because of the increased mutual reactances as the interconnects are packed closer together. An additional problem is that the fabrication yield drops as the circuit size and density are increased[9].

As a result of these problems, the scale of integration of individual ICs is approaching a limit. In order to make further advances, the industry is examining wafer-scale integration (WSI)[9], multichips[15] and hybrid[9] circuits in an attempt to continue increasing the circuit density - essentially by bringing the chips closer together. Whilst these approaches alleviate the problem of fabrication yield, they require very high interconnect and pin-out densities[9].

It is evident, then, that the problem of interconnects is becoming a critical issue in the design and performance of integrated systems. The requirement is emerging for alternative technologies such as 3D integration[9,16] or integrated optics[12]. 3D integration involves stacking ICs or wafers vertically, drilling through them so as to shorten the interconnect lengths required by the conventional edge-connection methods. The difficulties with this approach lie in forming the connections in the centre of the chips or wafers, and the consequent impact on circuit yield of such a risky process near the end of the fabrication sequence where each failure carries a high cost.
1.3. Integrated Optical interconnects

1.3.1. Advantages and disadvantages

The speed and bandwidth advantages of integrated optics were discussed earlier. Before discussing integrated optical interconnects in great detail, it is important to be aware of the difficulties of this approach.

A fundamental drawback of integrated optical interconnects lies in the use of energy, as opposed to potential, for carrying information. Optical circuitry is subject to losses from more or less fundamental mechanisms (diffraction) as well as limitations in the engineering (coupling and bending losses). Channel waveguides exhibit bending losses\(^{[17,18]}\) related to the guide structure and the radius of curvature. In practical systems, the minimum radius of curvature is of the order of centimetres, so that it is hard to imagine the circuit density of electronic ICs being reproduced in integrated optics. As an alternative, waveguide mirrors have been demonstrated\(^{[19,20]}\), with reasonably high efficiency, but this poses constraints on the guide fabrication techniques and the close tolerances imply a significant loss of yield. Optical signals can couple parasitically, by energy transfer between neighbouring or intersecting guides\(^{[21,22]}\) or by pickup of scattered light. These effects all tend to restrict the maximum packing density of integrated optical components and waveguides.

Another important disadvantage of the integrated optical interconnects lies in the relatively high complexity, size and cost of the interconnect - light source, modulation, propagation, detection, and associated electronics - when compared to those of a simple metal track.

While these considerations remain overwhelming disadvantages for very short-range applications, e.g. within individual ICs where propagation distances are no more than a few millimetres, they become outweighed by the speed and bandwidth advantages over longer distances. One may estimate that there is a critical distance, currently of the order of centimetres\(^{[14]}\), above which the optical solution becomes the best option for applications where very high frequency operation is important. This critical distance is apt to decrease as the frequency of operation increases, and as advances in integrated optics reduce the size and cost of optical interconnects.
1.3.2. Implementations of integrated optical interconnects

The function of an optical interconnect may be split into four parts: carrier generation, typically by means of a semiconductor laser\[^{23}\]; modulation, whereby a signal is imparted to the optical carrier; transmission, either guided or in free space; finally, reception, e.g. by means of a photodiode\[^{24}\], where the signal is converted back to electrical form.

The following sections outline the two approaches to the problem, i.e. free space and guided wave propagation.

1.3.2.1. Free space propagation

Free space interconnects make use of holographic optical elements (HOEs)\[^{25}\], suspended above the plane of the substrate, to reflect and focus the optical energy from light sources onto detectors as shown in figure 1.1(a). An alternative is to use transmission HOEs\[^{26}\], illustrated in figure 1.1(b). Although transmission HOEs are less compact than reflection HOEs, the optical efficiency tends to be higher. One special case which is particularly amenable to the use of transmission HOEs is clock distribution, where a single clock signal is to be focused onto a number of detectors distributed over a large area\[^{26}\].

The HOE approach offers a neat way of achieving large densities of high frequency interconnects to the entire surface of each chip, with little expense of real-estate, in contrast to guided wave and metallic interconnects which occupy a significant area and need to be routed through the edge of each chip.

Reflection HOE interconnects suffer from several engineering difficulties. The design of HOEs involves a compromise between flexibility and optical efficiency\[^{27}\]. One approach uses multi-faceted elements, with one facet dedicated to each laser. This enables each laser beam to be redirected at will, but the limited size of the facets leads to beam spreading by diffraction so that the detector has to be correspondingly large if it is to collect a substantial portion of the signal power. The alternative method is to use a single HOE which simultaneously redirects all the laser beams to their respective destinations. The problem with this is that a fraction of the optical power from each source is misdirected in unwanted diffraction orders which contribute to noise and crosstalk. In either case, the HOE needs to be aligned with very great precision with respect to both sources and detectors. This problem is not confined to alignment at the
packaging stage: the device must be capable of withstanding subsequent drift due to thermal expansion.

![Diagram of holographic optical element](image)

**Figure 1.1. Various holographic optical interconnect schemes**

The design of reflection HOE systems is further complicated by a compromise between the maximum allowable source-detector separation and the distance between the HOE and the substrate, which also limits the resolution obtainable with a given facet size and the consequent signal-to-noise ratio. Kostuk et al.[27] calculated that source-detector separations of about 2 to 3 cm might be achieved with a HOE 1cm distant from the source-detector plane, with resolutions of about 100μm at the detectors, assuming a facet diameter of 1mm. Such large spot sizes imply correspondingly large detectors,
with consequent expense of real estate, high capacitance and low bandwidth. The
detectors may be made smaller at the expense of optical power efficiency.
Alternatively, a multi-layer HOE may be used to extend the range and improve
resolution\cite{28}, but this method suffers from additional complexity and alignment
difficulties.

1.3.2.2. Guided wave propagation

Guided wave transmission involves confining the optical energy by means of an
optical fibre or a waveguide\cite{3}, wherein it may be modulated or switched as will be
discussed in the next section. An advantage of guided wave transmission over HOEs,
in the present context, lies in the similarity between the technologies of waveguide
devices and electronic ICs: both are based on planar structures, and they are fabricated
using similar techniques - thin film deposition, photolithography, etching, diffusion
and so forth. Waveguides on silicon substrates have been demonstrated using materials
such as silicon nitride and oxynitride\cite{29,30}, and in SIMOX (separation by implanted
oxygen) structures\cite{31} which are gaining acceptance as high performance silicon IC
substrates\cite{32}.

Lastly, the confinement of the optical signal to a small cross-sectional area,
typically a few microns square, implies comparatively low modulator drive power and
high bandwidth as well as a low real-estate cost.

Waveguide-based optical interconnects suffer from a number of difficulties.
Although the fabrication technology is broadly similar to that of semiconductors, the
materials needed for active components such as modulators are not all entirely
compatible. Lithium niobate, which is one of the most important materials for electro-
optic modulation, is a complex material and is not easy to deposit or etch. Moreover, it
contains lithium which is a fast diffuser and an n-type dopant in silicon\cite{33,34} so that
contamination is a potential problem if these are processed together at high
temperatures.

The other major problem with the waveguide approach lies in the weakness of
electro-optic interactions, so that typical modulators need to be quite long (of the order
of a centimetre), offsetting to some degree the advantage gained through cross-sectional
confinement. The modulator length may be reduced only at the expense of drive
voltage or optical transmission. Figure 1.2 illustrates a possible optical interconnect
scheme for WSI using waveguide modulators; similar arrangements may be envisaged.
for hybrid or multichip devices. It is clear from this that the device length is an important design parameter. One might estimate an upper limit of a couple of millimetres or so for a viable modulator in this type of application if the wafer surface utilisation is not to be dominated by the interconnect network.

![Expanded diagram showing waveguide crossovers with modulators and detectors to enable transfer of signals from one guide to another as well as to and from the electronic ICs.](image)

Figure 1.2. Schematic diagram of a possible optical interconnect scheme for wafer scale integration. Practical systems might require more waveguides and modulators. The scale of the devices, compared to electronic IC dimensions, is clearly an important consideration.
Another important aspect in the design of a waveguide modulator for this type of application is material and fabrication compatibility with VLSI technology. If a hybrid approach is adopted, this issue becomes less of a constraint as the devices are fabricated separately from the electronics: on the other hand, there is then the added complication of aligning, mounting and bonding the modulator to the VLSI substrate, and consequent problems in optical coupling to the waveguide[35,36]. The monolithic integration approach avoids this problem but the constraints of compatibility with VLSI fabrication are severe. VLSI device processing involves high temperatures, and diffusion of dopants and foreign species (such as lithium) needs to be strictly controlled. Active optical components cannot therefore be fabricated until after most of the electronic device fabrication is complete. Temperatures must then be kept as low as possible, i.e. below 700°C or so depending on factors such as duration and ambient atmosphere and the presence of potentially harmful impurities[34].

1.4. Optical waveguide modulation

1.4.1. Directly modulated lasers

There are many methods of electronically modulating an optical carrier in a waveguide. The most obvious method is direct modulation of the laser via the injected current[37]. In principle, this has the advantage of simplicity, since no additional optical components are needed, but it suffers from a number of engineering problems. The density of interconnect transmitters on a single substrate is limited by the laser power dissipation, which is related to threshold current. This brings about a requirement for low threshold current, which is observed to conflict with the needs for high signal-to-noise ratio and high frequency modulation[38]. These problems restrict direct modulation to relatively long range applications, where the size, complexity and expense of discrete components is justified by an increase in speed and bandwidth which cannot be achieved by other means.

1.4.2. Waveguide modulators

The alternative to direct laser modulation is to use a waveguide electro-optic modulator[6,39]. In this approach, the optical waveguide comprises a volume of some electro-optic material, with a set of electrodes, as illustrated in figure 1.3.
Figure 1.3. Examples of waveguide and electrode arrangements for electro-optic waveguide modulation. An electric field applied across the electrodes leads to a change in refractive index, altering the propagation characteristics of the waveguide.

The propagation characteristics of the waveguide may be altered, via the electro-optic effect, by the application of an electrical signal to the electrodes. Various schemes for utilising this type of effect are briefly described below.

Modulation of the phase propagation velocity of light within the waveguide may be used for switching optical energy between two adjacent waveguides[40] as shown in figure 1.4. A light wave travelling down one guide sets up a weak excitation field in the other guide. Energy is gradually transferred from one guide to the other until all the energy is in the second guide (whereupon the process begins again in reverse). Essentially, the two guides behave as weakly coupled oscillators, the coupling strength depending on the distance between the guides and the phase propagation velocities in each. The light may be switched between the two output guides by suitably altering the relative phase velocities.
Figure 1.4. Directional $\Delta\beta$ coupler (schematic). $L$ is an odd-integer number of coupling lengths so that the light entering one guide is transferred to the other. When $V$ is suitably altered, the coupling length changes so that $L$ becomes an even-integer number of coupling lengths, and light entering one guide exits in the same guide.

Figure 1.5. TIR waveguide switch. Light entering one guide is reflected into the other guide by means of the electro-optically induced change in refractive index in the region between the electrodes.

Switching may be achieved by total internal reflection between two waveguides intersecting at a shallow angle\cite{41}, as illustrated in figure 1.5.

The change in phase propagation velocity may be converted to amplitude modulation, by interference with a reference signal. Usually, two guides are modulated in opposite senses, doubling the relative phase change per unit active length, as in the Mach-Zehnder interferometer\cite{42} illustrated in figure 1.6.
Figure 1.6. Mach-Zehnder interferometer (schematic). Light is split between two waveguide branches, either using a Y-branch as shown or a 3dB directional coupler. At the output, the two guides are recombined: if the applied voltage is such as to develop a phase shift of $\pi$ radians between the two signals, the output is effectively switched off.

Figure 1.7. Cutoff modulator (cross section schematic): when a field is applied via the electrode structure, the refractive index in the underlying waveguide is altered in such a way that the waveguiding conditions are no longer satisfied, so that the optical energy is dumped into the substrate.

A waveguide cut-off modulator$^{[43]}$, illustrated in figure 1.7, makes use of the conditions which a structure must fulfil, with regard to dimensions and refractive indices, in order to support guided propagation$^{[4]}$. A guide is fabricated precariously close to the cut-off condition, and the transmitted light may be switched off by using an electro-optically induced change in refractive index to tip the balance between guiding and cut-off. This scheme suffers from the disadvantage that, close to the cutoff
condition, much of the optical energy travels deep down in the substrate where the electric field is weak so that the electro-optic interaction is comparatively inefficient: in addition to this, some of the dumped optical energy can be coupled back into the guide beyond the modulator thus leading to a degradation of the signal-to-noise ratio.

1.5. Optical modulation: interactions and materials

Thermo-optic modulation has been demonstrated using a strip of resistive material above a waveguide. When this is heated by an electrical current, the waveguide refractive index is altered via a thermo-optic effect. This scheme benefits from simplicity as it does not require exotic materials. Most materials exhibit the required thermo-optic effect, and devices using this effect have been demonstrated using silicon and silica waveguides. Nevertheless, it is severely limited in speed by its dependence on thermal time constants.

Electro-optic modulation may be achieved by using the electric field to alter the depth of a semiconductor depletion region, where the refractive index is lowered in the presence of mobile carriers. This type of modulator has a clear advantage in terms of material compatibility with semiconductor ICs, and has been demonstrated using III-V semiconductors as well as silicon. The main drawbacks are that this electro-optic mechanism is comparatively weak, and optical losses are introduced by the presence of the charge carriers. (This phenomenon is relevant to the operation of the device structures reported in this thesis, and will be re-examined in some detail in Chapter 3).

Another possibility is to use a material with a high electro-optic coefficient. One such material is lithium niobate, in which the electro-optic effect arises from the asymmetry of the crystal structure. Electron wavefunctions, associated with certain ionic bonds within the material, become distorted under the influence of an electric field. This distortion gives rise to a change in polarisability (and dielectric constant) and hence in refractive index. Although this is associated with a relatively high dielectric constant (c. 30 along the c-axis), so that the stored energy and electrical driving power per unit volume are high, the effect is maintained up to very high frequencies so that, in principle, very high bandwidths can be achieved. The chief disadvantage of this material lies in its structural complexity and the difficulty of processing it or growing it in thin layers. This will be discussed in greater detail in Chapter 4.
Lithium niobate is not the only material to exhibit this electro-optic mechanism. Others include gallium arsenide, gallium phosphide, quartz and lithium tantalate, but lithium niobate is a material in which the effect is particularly strong\textsuperscript{51}. It also exhibits a piezo-electric effect which has been put to use since the early days of integrated optics in SAW devices\textsuperscript{52}. These factors combine to make lithium niobate arguably the most important electro-optic material today, supported by a considerable body of knowledge and experience in growing and processing\textsuperscript{50,53}.

1.6. Lithium niobate device structures

It is clear from the literature that the design of high performance electro-optic modulators is dominated by the weakness of the electro-optic interaction. This may be illustrated by an order-of-magnitude calculation of the change \(\Delta n\) in refractive index\textsuperscript{39} in a typical device:

\[
\Delta n = \frac{1}{2} n^3 r_{33} V G,
\]

where \(n\) is the unperturbed index of lithium niobate, \(r_{33}\) is the strongest electro-optic coefficient of lithium niobate\textsuperscript{54}, and \(V\) is the potential difference applied across a gap \(G\); \(\Gamma\) is a measure of the electrostatic efficiency of the structure\textsuperscript{39}. \(L_\pi\), the device length required to accumulate an optical phase change of \(\pi\), is then given by

\[
L_\pi = \frac{\lambda G}{n^3 r_{33} \Gamma V}
\]

Evaluating this expression, with \(\lambda = 10^{-6}\)m, \(G = 2\times10^{-6}\)m, \(n = 2.2\), \(r_{33} = 30 \times 10^{-12}\)m/V, \(\Gamma = 0.5\), and \(V = 5V\), \(L_\pi\) comes to \(2.5\times10^{-3}\)m or 0.25cm. The voltage-length product, \(VL_\pi\), is a useful and often-quoted figure of merit, and is of the order of 1Vcm. It should be added that the length of a practical Mach-Zehnder structure is often greater than this because of the additional lengths required for y-branches at each end of the modulator.

The active length of a device has a direct impact on bandwidth, because of the difference between optical and microwave propagation velocities\textsuperscript{55}. The microwave velocity is low because of the high dielectric constant of lithium niobate. The optical wave, travelling faster than the microwave, experiences a gradually phase-shifting electric field as it travels down the modulator. The net modulation is thus reduced and may be annulled if this phase mismatch accrues to an integer number of cycles.
Various phase-reversal techniques have been used to mitigate this effect\textsuperscript{[55,56]} but it remains an important limitation on the bandwidths of these devices.

A measure of performance which is often used to compare devices, from the point of view of bandwidth, is the ratio between drive voltage and the bandwidth\textsuperscript{[55]}. Considering the electro-optic interaction described above, and the relationship between device length and bandwidth as implied by the velocity mismatch problem, it becomes apparent that although one might attempt to increase the bandwidth by reducing the device length, the drive voltage would increase proportionally, so that this ratio is unchanged. The voltage/bandwidth ratio is thus a figure of merit which depends on the efficiency of the device structure independently of this compromise. In the present context, however, it is more convenient to consider devices assuming a fixed operating voltage of 5\text{V}, this being compatible with TTL (transistor-transistor logic) circuits, and consider the bandwidth on its own as a useful figure of merit.

It is interesting to note that the bandwidths of almost all the devices reported to date tend to be limited by electronic considerations. An obvious trade-off would therefore be to sacrifice optical bandwidth for electronic, e.g. by means of a multiple-pass device such as a Fabry-Perot cavity. The problem with this lies in the difficulty of fabricating efficient waveguide mirrors\textsuperscript{[19]}. Nevertheless, it is perhaps surprising that the literature reveals little work along these lines other than studies on waveguide losses \textsuperscript{[57,58]}, where the bandwidth was traded for sensitivity and the devices were modulated by gentle temperature changes, the cavity characteristics being used to yield accurate information about the waveguide losses, a temperature sensor\textsuperscript{[59]} and two theoretical studies\textsuperscript{[60,61]}.

Most of the devices to date make use of a coplanar-electrode arrangement, in which the electrodes are fabricated by etching a suitable pattern in a thin metal film deposited over the waveguide\textsuperscript{[39]}. This is the arrangement illustrated in figure 1.3. The electric field between such electrodes extends, to some degree, into the waveguide below. This scheme benefits from great ease of fabrication, but it is geometrically inefficient. The minimum size of the optical field confinement region is of the order of the optical wavelength. An optical buffer layer is usually required between the electro-optic waveguide and the electrodes, to distance the optical energy from the metal in order to avoid severe absorption losses. The presence of this buffer layer greatly reduces the electric field in the electro-optic region, as will be discussed in Chapter 3: this is a severe effect because of the high dielectric constant of lithium niobate (about 30 along the crystal z-axis, compared to 3.8 for silicon dioxide which is often used as a buffer layer). The thickness of this buffer also needs to be of the order of the
wavelength. The inter-electrode gap must be of about this size or greater. Although a smaller gap yields a more intense electric field, it also reduces field penetration into the electro-optic waveguide. Another restriction is imposed on the electrode spacing by the difficulty of fabricating a narrow gap with no short circuits over the several millimetres needed for electro-optic interaction. As a result of all this, the inter-electrode gaps are of the order of several microns, and it is difficult to achieve a high electro-optic overlap, i.e. to match the electric and optical field distributions and confine them to the electro-optic region\[39\].

The large dielectric constant of lithium niobate has been mentioned as a disadvantage in several respects (high electrical stored energy density and device driving power; reduction of the microwave velocity in the electrode structure; and diminished electric field appearing across the lithium niobate in the presence of buffer layers). It is worth noting that this high dielectric constant of lithium niobate is connected with the strong electro-optic behaviour. The high dielectric constant implies a high polarisability, and hence a significant distortion of the electronic wavefunctions within the unit cell under the influence of an electric field: it is precisely this distortion, coupled with a strong crystal asymmetry, which leads to changes in the refractive index. The implication is that this is a fundamental problem, and it is unlikely that an alternative material can be found combining a low dielectric constant with a high electro-optic coefficient.

![Diagram of Waveguide Modulator](image)

**Figure 1.8.** Waveguide modulator using the parallel plate geometry\[62\].

An alternative scheme has also been demonstrated, using the parallel plate structure shown in figure 1.8\[62\]. A waveguide was formed in a lithium niobate substrate, and metal electrodes were applied to front and back faces. The substrate was thinned by grinding and polishing before forming the back electrode, so as to increase
the electric field for a given voltage. Although this arrangement is more geometrically efficient than the coplanar configuration, the inter-electrode gap was comparatively large (about 15µm), being determined by the thickness to which the lithium niobate could be thinned without damage. The electric field was correspondingly low so that the product $VL_{\pi}$ was about 3Vcm.

1.7. The present project

This research project came into being when UCL's thin film lithium niobate programme\cite{1,2} began to yield promising results. It seemed that the difficulties of growing thin films of lithium niobate were being overcome so that some effort should be directed towards exploiting the properties of the new material by fabricating thin-film electro-optic modulators.

These devices would combine the geometrical efficiency of a parallel-plate configuration with a narrow inter-electrode gap (determined by the lithium niobate film thickness, i.e. less than a micron). This is illustrated in figure 1.9. The narrow gap leads to an intense electric field for a given voltage; the optical energy is tightly confined in the region of maximum electric field strength; the buried electrode makes best use of the $r_{33}$ electro-optic coefficient of lithium niobate by aligning the electric field along the crystal c-axis.

![Proposed waveguide electro-optic modulator using the parallel plate configuration.](image)

Figure 1.9. Proposed waveguide electro-optic modulator using the parallel plate configuration.
The first phase of this programme, discussed in Chapters 2 and 3, was aimed at developing a thorough theoretical understanding of the parallel-plate structure. The close proximity of electrodes and optical energy was a potential cause of large optical absorption losses. Chapter 2 describes the development and experimental testing of a waveguide modelling algorithm capable of handling optically absorbing materials such as metals. Chapter 3 describes how these models were used to investigate a variety of possible structures. The original concept, using metal electrodes, was found to suffer from severe optical losses. This was alleviated by using silicon electrodes and operating the devices in the infra-red, beyond the silicon band-gap wavelength where optical absorption was weaker. This departure from the original brief had the benefit of steering the project towards established VLSI technology.

The electrical characteristics of the structures were modelled. The development of the electrical simulation is described in Chapter 2. In Chapter 3, this is combined with results from the optical modelling to assess the electro-optic performance of the devices up to high frequencies. The electrical and optical characteristics of the structures were found to be related via free charges in the electrodes, which governed both electrical conductivity and free carrier optical absorption. This led to a compromise between optical attenuation and device bandwidth.

Finally, as a result of the theoretical work, a family of parallel-plate modulators was designed. These were predicted to be no more than a few hundred microns in length, and to exhibit high bandwidths (of the order of 10s of GHz).

The second phase of the work is discussed in Chapters 4 and 5. Chapter 4 is concerned with the growth of the electro-optic layer by RF sputtering. Chapter 5 is concerned with devising a suitable fabrication sequence and mask set for the structures designed in the theoretical phase of the programme. The compatibility of the new structures with VLSI technology was demonstrated by fabricating them on silicon-on-sapphire, a commercially available VLSI substrate material. Although many techniques could be drawn directly from established silicon device technology, a number of logistical and material compatibility problems had to be overcome.

The third and final phase of the programme, discussed in Chapter 6, is concerned with optical assessment of the devices. Electro-optical measurements were taken and compared to theoretical predictions. Agreement was found to be within a factor of about 2; the remaining discrepancy could be accounted for by material imperfection.
Conclusions of the project are presented in Chapter 7. It is argued that the project was successful in so far as the ultimate aim was achieved, viz. the demonstration of a parallel plate electro-optic waveguide modulator. A great deal of further work is prompted by this success. The performance of the prototype devices was severely limited by the choice of top electrode material: there remains the task of optimising fabrication parameters in order to exploit the potential capabilities predicted by the theoretical work. As mentioned above, these devices are expected to be compatible, in terms of materials and size as well as performance, with silicon IC technology; indeed, devices were fabricated on silicon IC-grade substrates. However, this does not constitute an exhaustive test, and it remains to demonstrate the integration of modulators with functioning CMOS circuitry before compatibility is proven.

1.8. Summary

The subject of integrated optics, and its role in the field of interconnects, has been outlined. Whilst optical fibres are well established in long-range applications, the implementation of integrated optical devices for very short range interconnects remains complicated and expensive. The need for a compact waveguide modulator structure has been identified. The present research project was set up to satisfy that need by exploiting the recent advances in UCL's MBE programme.

As the following chapters will recount, this project was successful in demonstrating the operation of the first parallel-plate electro-optic waveguide modulator using thin film lithium niobate.

References


Chapter 2: Computer Modelling

Abstract

The first phase of this research programme was aimed at developing a thorough understanding of the performance characteristics of parallel-plate modulator structures. To this end, this chapter describes the development of a set of computer models.

A program, named "MOD2", was developed to calculate optical propagation constants and losses in multilayer waveguides; it was supplemented by a graphics program, "MPLOT", for examining the behaviour of field and energy distributions in the modelled structures.

Another model, "VMEFF", was written to calculate the electrical characteristics of the structures, in order to simulate the behaviour of the parallel plate arrangement when driven with a time-varying electrical signal. The purely electrostatic fields in such an arrangement are straightforward: indeed, the situation becomes trivial if the aspect ratio (electrode width to separation) is large enough to justify ignoring fringing fields at the edges of the structure. The situation becomes slightly more complicated if the electrodes are resistive and are driven at high frequency, so that the charging of the capacitor becomes limited by the distributed RC time constant. This effect places a limit on the electrical bandwidth of the structure.

Clearly, this behaviour depends on a number of geometrical and structural factors, so that the results are influenced by the choice of device layout. VMEFF was written to model the behaviour of the devices which were fabricated later in the programme. VMEFF combined optical data calculated by MOD2 with material and structural parameters to simulate device structures and calculate figures of merit.

2.1. The multilayer waveguide model, MOD2

Optical waveguiding\textsuperscript{[1]} may very loosely be described as a phenomenon whereby light is guided using the principle of an optical cavity, i.e. several surfaces between which the light is reflected to and fro. In what is probably the simplest configuration, illustrated in figure 2.1, the reflecting planes are formed by the upper and lower surfaces of a thin film deposited on a substrate: the reflection coefficients are determined by the refractive indices of the three regions thus defined, and the angle of
incidence of the light on the reflecting surfaces. If the refractive index of the film is higher than those of the surrounding media, total internal reflection may occur so that light is propagated along the structure without radiative loss. The situation can be analysed mathematically to show that guiding can only take place for certain discrete reflection angles, which fulfil a resonance condition. Each solution to this set of conditions is usually referred to as a "mode" of the waveguide, and is conventionally indexed according to the integer $m$ which describes the phase shift $m\pi$ corresponding to a round trip of the cavity. Considering the Fresnel coefficients, which relate both the phase and amplitude of transmitted and reflected waves, it can be seen that the modal angles generally depend on the plane of polarisation. Birefringence can also contribute to the polarisation-dependence of the modal properties.

For each guided mode, one can calculate or measure an "effective refractive index" which is defined by the ratio of the wavenumber in the guide (conventionally written as $\beta$) to the free space wavenumber $k_0$. In general, $\beta$ is complex, the imaginary term representing gain or loss in the guide. Gain may be due to the presence of an amplifying medium, as in a laser: loss may be due to leakage (imperfect total internal reflection), absorptive media or to scattering from imperfections in the guide structure.

In a waveguide electro-optic phase modulator, guided light is phase-modulated by means of a change in the effective refractive index $\beta/k_0$. Accurate calculation of $\beta/k_0$ was necessary in the present project, to model both the electro-optic effect and the guide losses. Calculations of loss are of particular importance in this work, since the mobile charges which are present in electrical conductors (such as needed to form the electrodes) cause optically absorption and must, by the very nature of the parallel plate configuration, form integral parts of the waveguides.
2.1.1. MOD2: mathematical basis

The mathematical basis of MOD2 is the same as that used by Walpita\cite{3} in a similar study (although his method of solution was not divulged), and stems from the complex formulation of the Maxwell equations\cite{4}. Axes, illustrated in figure 2.2, are chosen such that the x-axis is perpendicular to the plane-stratified structure. The z direction is taken along the direction of propagation of the guided waves. The y-axis then lies across the guide to complete a right-handed set of axes.

Using the $e^{-ikx}$ dependence of a wave within each layer\cite{4}, the fields at the top of a given layer are expressed in terms of those at the bottom; field continuity requirements then yield an expression relating the fields on either side of each interface; applying these equations to all the layers in turn, the field in the semi-infinite region above the guide (the "cover") can be expressed in terms of the field in the substrate region below the guide. The guided modes are obtained by looking for solutions to these equations for which the fields far from the guide tend to zero.
A plane wave may be expressed in terms of its amplitude $A$:

$$A = A_0 e^{i(\omega t - k \cdot r)} \quad \text{where} \quad k = n k_0,$$

$k_0$ being the free-space wave-vector.

(In the first equation, the sign of the exponent is arbitrary: the usual convention is as shown here).

In general, these become tensor equations, the refractive index $n$ being replaced by appropriate components which describe the index ellipsoidal$^2$: throughout this text, this simplified notation is used for the sake of clarity.

The vector $k$ is in general complex, as is the amplitude $A_0$. The imaginary part of $k$ gives rise to an exponential variation in the wave amplitude: this may be interpreted as a loss or gain term. (A thumbnail calculation shows that this imaginary component must be negative in the case of loss, using the conventional sign notation as above.)

Henceforth, for the sake of clarity, the $e^{i\omega t}$ term is absorbed into $A_0$.

Remembering that the $y$ direction is at right angles to the direction of propagation, the vector $k$ may be split into two components, $\beta$ and $k_x$, along the $z$ and $x$ directions respectively. The relation

$$\beta^2 + k_x^2 = k^2,$$

where $k$ represents the magnitude of the vector $k$, implies

$$k_x = \pm \sqrt{k^2 - \beta^2}. \quad \text{(2.1)}$$

Thus for a given value of $\beta$, there are two waves: these may be denoted by the superscripts $+$ and $-$ according to the sign of the square root in equation (2.1).

Equations may now be set up to relate the fields on either side of each interface, and at the top and bottom of each stratum in the structure. To do this, it is convenient to express the amplitude of a modal field in terms of a $2 \times 1$ vector, with one component for the upward-travelling wave and one for the downward-travelling wave. This vector $A$ is a generalised field vector which may be used to represent either electric or magnetic fields, and it may be manipulated using straightforward matrix equations. For example, within stratum number $m$,
\[
\mathbf{A}_m = \begin{pmatrix} A_{+0m} \exp(-i k_{x_m} x) \\ A_{-0m} \exp(+i k_{x_m} x) \end{pmatrix}
\]

where an obvious subscript notation has been adopted. The relation between the fields at the top and bottom of stratum \(m\) can then be expressed:

\[
\mathbf{A}_m(\text{top}) = \mathbf{T}_m \cdot \mathbf{A}_m(\text{bottom}), \quad \text{where } \mathbf{T}_m \text{ is the transfer matrix given by}
\]

\[
\mathbf{T}_m = \begin{pmatrix} \exp(-i k_{x_m} h_m) & 0 \\ 0 & \exp(+i k_{x_m} h_m) \end{pmatrix}
\]

The field components parallel and perpendicular to the interface are used to establish the relationships of fields in neighbouring strata. For a TM\(^{[4]}\) mode,

\[
D_x k = 0 \implies \frac{D_x}{D_z} = -\frac{\beta}{k_x}
\]

Then the continuity of the electric potential implies

\[
\frac{D_{z m+1}}{\varepsilon_{z m+1}} = \frac{D_{z m}}{\varepsilon_{z m}}
\]

while Gauss' Law yields

\[
D_{x m+1} = D_{x m}
\]

Similarly, for TE\(^{[4]}\) modes

\[
\frac{H_x}{H_z} = -\frac{\beta}{k_x}, \quad \text{with } H_x \text{ and } H_z \text{ both continuous across the boundary.}
\]

These lead to the following transfer equations, relating the wave components on either side of each interface:

\[
\mathbf{M}_{m+1} \mathbf{A}_{m+1}(\text{bottom}) = \mathbf{M}_m \mathbf{A}_m(\text{top}), \quad \text{where}
\]

\[
\mathbf{M}_m = \begin{pmatrix} 1 & 1 \\ k_{x_m}/\varepsilon_{x_m} & -k_{x_m}/\varepsilon_{x_m} \end{pmatrix}
\]

for TM modes, and

\[
\mathbf{M}_m = \begin{pmatrix} 1 & 1 \\ k_{x_m} & -k_{x_m} \end{pmatrix}
\]

for TE modes.

Having established these, a relationship between the fields in the superstrate and the substrate may be derived by multiplying together all the transfer matrices: denoting the top stratum by the label \(M\),
In addition to this equation, the conditions for waveguiding are that the field amplitudes at \( \pm \infty \) tend to zero: clearly, this implies

\[ A_{-oc} = A_{+os} = 0. \]

Finally, these may be combined to give a single equation, whose solutions correspond to waveguiding modes of the structure:

\[
\begin{bmatrix} A_{-oc} \\ A_{+os} \end{bmatrix} = G \begin{bmatrix} 0 \\ A_{-os} \end{bmatrix}
\]

(2.3)

Expanding, and dividing one component of this vector equation by the other,

\[
\frac{M_{c11}}{M_{c21}} = \frac{(G_{11}M_{s12} + G_{12}M_{s22})}{(G_{21}M_{s12} + G_{22}M_{s22})}
\]

This may be re-arranged to yield

\[
R = \frac{M_{c21}(G_{11}M_{s12} + G_{12}M_{s22})}{M_{c11}(G_{21}M_{s12} + G_{22}M_{s22})} - 1 = 0.
\]

(2.5)

The problem is thus reduced to the task of searching for solutions to this equation, by varying \( \beta/k_0 \).
This formulation for the waveguiding condition may be re-arranged in many ways: although mathematically equivalent, some of these are less suitable than others for computer solution because they require higher precision. This may be illustrated by considering the general form of these equations,

\[ A - B = 0, \quad (2.6) \]

where \( A \) and \( B \) are functions of a variable \( x \).

If this equation is to be solved for \( x \) to within an accuracy of \( \Delta_x \), i.e. to \( \log(x/\Delta_x) \) significant figures, the absolute precision required of \((A - B)\) is given by

\[ \Delta(A - B) = \Delta_x (A' - B'), \]

where the dashes signify derivatives with respect to \( x \). The corresponding requirement of significant figures for \( A \) is then

\[ \Delta_A = \log(A/\Delta(A-B)), \]

with a similar expression for \( \Delta_B \).

If \( A \) and \( B \) are large, a correspondingly high precision may be required to avoid severe problems from rounding errors. By re-formulating the original equation (2.6) as

\[ (A/B) - 1 = 0, \quad (2.7) \]

the required accuracy is reduced to

\[ \Delta(A/B) = \Delta_x (A'B - AB')/B^2, \]

which is of the order of \( \Delta_x (A' - B')/B \), with correspondingly lower requirements of significant figures.

This choice of formulation was found to have a critical effect on the ability of the program to solve the equations. The argument is tantamount to the observation that the difference between two large but similar numbers cannot be so accurately computed as their ratio. A similar argument applies when \( A \) and \( B \) are small: inspection reveals that the safest route, for computer solution in the general regime, is to take the ratio so that \( R \) is dimensionless - although, as is discussed below, even this precaution did not guarantee success.
Figure 2.3. Plot of the logarithm of the squared absolute magnitude of $R$ as a function of the trial value of $\beta/k_0$. Solutions to the waveguiding conditions correspond to zeros of $R$. Two discontinuities are visible, terminating at nodes at $(1.0,0.0)$ and $(0.5,-0.5)$ corresponding to the indices of the cover and substrate regions respectively.
Some example plots, of the logarithm of squared absolute magnitude of \( R \) as a function of the real and imaginary parts of \( \beta/k_0 \), are shown in figure 2.3. Two lines of discontinuity are visible: these arise from the quadratic equations (2.1) for \( k_x \). In the intermediate layers of the structure, both positive and negative values of \( k_x \) are used since upward and downward waves are present: in the cover and substrate, only one root is chosen, in order that the field should tend to zero at \( \pm \infty \). This implies that from a mathematical standpoint, \( R \) is a four-valued function (one value corresponding to each combination of choices for \( k_{xc} \) and \( k_{xs} \)), of which only one value represents waveguiding according to the definition: however, these four values lie on a single continuous surface, with four nodes at \( \beta/k_0 \) values corresponding to the roots of

\[
\beta^2 = n_c^2 k_0 \quad \text{and} \quad \beta^2 = n_s^2 k_0^2
\]

Another salient feature of the plots is that each guided mode seems to be accompanied by a peak in \( R^2 \): some of these peaks go to infinity. These peaks have little physical significance, arising from maximal values of \( A/B \) in equation (2.7). The poles, where \( R^2 \to \infty \) occur at values of \( \beta/k_0 \) for which \( B = 0 \); these, and similar (though less remarkably visible) points where \( A = 0 \) (and \( R^2 = 1 \)) can be shown to correspond to situations where the upward (or downward) waves vanish at certain interfaces in the structure.

Note that these equations may be re-arranged to give the reflectivity of the structure, as a function of angle of incidence and wavelength. This idea has been put into practice elsewhere, though confined to lossless media (real indices)\(^5\). This could be used as a powerful diagnostic tool as it does not require waveguiding per se: only the reflectance need be monitored. This is the principle of operation of interferometric thickness monitors.

### 2.1.2. MOD2: strategy for finding solutions

The search strategy is illustrated by the flow chart in figure 2.4. The computer program calculates \( R \) for a series of trial values of \( \beta/k_0 \) (subroutine RSTEP). These trial values are taken on the real axis, between user-defined upper and lower limits. When a minimum is found, a "homing" algorithm is called (subroutine RSOLVE). This algorithm uses a complex formulation of the Newton-Raphson iteration. The solution is refined until the calculated value of \( R \) falls below a pre-determined threshold; it is then stored, and the program reverts to the stepping routine where it left off.
Note 1. \( \beta/k \) steps along the real axis. A Newton-Raphson algorithm is used to home in towards \( R = 0 \).

Note 2. \( \beta/k \) is now allowed off the real axis. A Newton-Raphson algorithm is used to home in towards \( R = 0 \).

Note 3. \( \beta/k \) steps along the real axis, from where stage (1) left off at point (B). A Newton-Raphson algorithm is used to home in towards \( 1/R = 0 \).

Figure 2.4. MOD2 flow chart
The first snag with this approach was that although R was found to vary more or less cyclically with \( \beta/k_o \) as seen in figure 2.3, it was found that the cycle period could be very uneven. In particular, R tended to change very rapidly when \( \beta/k_o \) was near the bulk refractive index of a substantial portion of the structure. This was because the x-components of the wave-vectors (i.e. components normal to the guide surface) were changing rapidly, and R was more directly related to these components than to \( \beta \) itself. This variation in the period of R made it awkward to choose a step size for \( \beta/k_o \). If the step was too small, time was wasted scrutinising regions where R varied only gradually, and the program took a long time to run: if on the other hand the step was too large, some of the modes could be stepped over and missed.

In addition to this, sharp peaks with very steep gradients were found to occur close to solutions, as can be seen in figure 2.3. Calculations at discrete intervals could yield a rather confusing picture of the situation unless very small steps were used. A method was developed to take this into account, by adjusting the step size at each iteration according to the local gradients of R. This was done using the Newton-Raphson method described above, this time searching alternately for R and 1/R as the program waded through the peaks and troughs in its quest. Limits were imposed on the maximum step size in order to avoid stepping over minima as mentioned above. These limits were calculated according to a "search resolution" specified by the user.

Rapid fluctuations in R could also be generated by thick, low-index regions such as buffer layers. This was because the \( k_x \) term for such a layer had a large imaginary component, giving a large exponential term in the transfer matrix for the layer. This, in turn, led to large values of \( G_{11} \) and \( G_{12} \) (or \( G_{21} \) and \( G_{22} \)), introducing rounding error problems as discussed above. This problem was largely overcome by splitting G into two component matrices, \( G_s \) and \( G_c \), such that (cf. equation (3))

\[
G_c = \prod_{MR} (M_m T_m^{-1} M_m^{-1}) \quad \text{and} \quad G_s = \prod_{MR-1} (M_m T_m M_m^{-1})
\]

and so equation (2.4) became \( G_c M_c A_c = G_s M_s A_s \), with a suitably re-arranged expression (2.5) for R. (Note that the quantity R itself was altered, but the conditions for waveguiding remained \( R = 0 \)).

In this re-arrangement, MR refers to a "reference" layer within the structure; the computer program was written to let this be chosen by the user.
This enabled the two large exponential factors to be split between the numerator and denominator in the equation for $R$, reducing the gradients to manageable levels for most situations. The problem was thus alleviated, although not entirely eliminated. If, for example, a guide is modelled with two high index regions separated by thick low-index layers, the structure may begin to resemble a pair of weakly coupled guides, with symmetric and anti-symmetric modes which tend towards degeneracy as the separation increases. An example of this is shown in figures 3.2 and 3.4 of Chapter 3, where the fundamental and first harmonic modes of a silver/dielectric/silver sandwich both converge on the characteristics of the isolated surface plasmon. This is an unavoidable problem, but one which does not occur often in useful, practical situations: where necessary, a particular situation may usually be examined by scrutinising the range of interest using a very high search resolution.

A potential problem with the program was that of "false" minima in $R$. These could occur if a minimum in $R$, as found by RSTEP, did not lead to a valid solution. This was checked by putting a limit on the number of homing iterations in RSOLVE; if the computed value of $R$ was then still above the threshold, the solution was rejected. This limit on the number of homing iterations was taken from the user-specified "search resolution".

A problem was that two or more minima in $R$, as found by RSTEP, could lead to a single solution, for example when modelling a near-degenerate situation (see above). This was avoided quite simply, by comparing each solution to the previous one.

One question which arose from this choice of search strategy, was that of the search domain: was it justified simply to step $\beta/k_o$ along the real axis, and to hope that solutions would all be sufficiently close to the axis to give rise to appropriate minima in $R$? Might there not be solutions either too far away, or protected by a "ridge" of high values of $R$ so that RSOLVE, following always the steepest gradient downwards, could not reach them? Although these questions were not rigorously resolved, it was argued that although such solutions might conceivably exist, they would be of little practical use. This was because an imaginary part of $\beta/k_o$ equal to -i corresponds to an absorption loss of $1/e$ per wavelength, or about $10^7$ dB/cm at optical wavelengths! Practical losses are much lower than this, so that the imaginary part of $\beta/k_o$ would be much smaller than the typical step size used by RSTEP. The same argument, coupled with the assumption (supported by plots such as that depicted in figure 2.3), that $R$ does not exhibit such wild excursions as the "ridge" case would require for solutions
close to the real axis, suggested that the model was likely to be capable of providing complete results for most practical purposes.

2.1.3. MOD2: birefringence

In order to cater for birefringent materials, the refractive index ellipsoids were used in full: thus the three indices for each layer could be specified independently. This was subject to the constraint that the crystal and waveguide axes were aligned. It would have been possible to modify the program so as to relax this constraint, but this would have rendered the present calculations unnecessarily cumbersome and slow for most practical situations.

As will be discussed in the following chapters, devices were fabricated using severely mis-oriented lithium niobate where the extraordinary axis was inclined both to the guide normal and to the direction of propagation. This situation was analysed using an approximate method, which will be discussed in Chapter 3.

2.1.4. MOD2: implementation

Ryan-McFarland Fortran was chosen for this work, since it combines high speed, precision, and a highly modular structure. This is a version of Fortran 77 which provides for double-precision-complex arithmetic such as is required for accurate calculations of this nature. The work was carried out on IBM PC compatible computers, fitted with 8087 math co-processors in order to speed up the calculations. The program structure is laid out in Appendix 2.A, along with example input data files and corresponding results.

The program was also translated into Pascal, which made it directly accessible by other programs: e.g. for extracting thickness and index figures from $\beta/k_0$ measurements, as will be described in Chapter 4. However, the Pascal version was slower by a factor of about five, so the Fortran version remained in great demand.

A logical extension to this work was to write a routine to display modal field plots. This extension, "MPLOT", was written in Pascal, and used the values of the $x$-components of wave-vectors and the interface relations described above to construct wave functions in each stratum. This proved to be an invaluable tool to gain insight
into the physical situations under scrutiny, as well as to verify the labelling of the predicted modes.

2.1.5. MOD2 results: comparison to literature

Figure 2.5. Data calculated using MOD2, for comparison to Reisingers results.
The program was initially tested by comparison with data from the literature. Propagation constants and losses were predicted for some structures investigated by Reisinger[6]. In this paper, Reisinger describes theoretical calculations of propagation constants, losses, and modal distributions in multilayer waveguides structures with metallic overlays. He went on to fabricate and characterise a number of waveguides, and found good agreement with the predictions.

Results from MOD2 were found to agree precisely with Reisinger's own predictions. The most significant of these results are shown in graphical form in figure 2.5: these are plots of effective index and log(loss coefficient), displayed in a similar manner to the graphs in Reisinger's paper so as to allow direct comparison.

2.1.6. MOD2 results: comparison to experiment

The model was also experimentally tested. Propagation coefficients and losses were measured, of waveguides formed by depositing silicon nitride on thermally oxidised silicon. The measurements were in reasonable agreement with the model predictions.

Silicon wafers were oxidised and partially etched back, to provide oxide layers of a range of thicknesses up to 1μm. These were then re-oxidised slightly, to eliminate any pinholes generated in the etch. Finally, LPCVD silicon nitride was deposited on the oxide.

Film thicknesses were measured using an automated interferometric microscope. Refractive indices were measured using an ellipsometer. The effective indices of guided modes were measured by the prism coupling method[7]. The measurements were found to be repeatable to within about 0.0003. Optical losses were measured by the video camera method[8], coupling into the guide via a prism. The range of measurable losses was limited to approximately 15dB/cm at the top end by the difficulty in obtaining a sufficiently long bright streak. In low loss measurements, uncertainty arose from the non-uniform distribution of scattering centres. Since only the intrinsic losses were of interest in this study, the Rayleigh scattering contribution added uncertainty to the measurement.

The effective refractive indices was found to be governed chiefly by the thickness of the nitride layer. In all cases where the losses were sufficiently low to measure anything, the influence of the silicon was minimal and so the guide could be
approximated by a thin film of silicon nitride on an oxide substrate. Effective indices are plotted versus nitride thickness in figure 2.6.

Figure 2.6. Effective indices of silicon nitride waveguides vs. silicon nitride thickness. The curves were calculated using MOD2: the upper curve of each pair illustrates the effect of an uncertainty of 0.01 in the refractive index of the silicon nitride. The points represent measurements.

The intrinsic losses in these structures arose from two effects: absorption in the silicon, and leakage to the substrate. Leakage arose from the high refractive index of the silicon (3.85) compared to the guiding medium (silicon nitride, 1.99). These contributions were investigated separately, using the computer model, by plotting losses versus real and imaginary parts of the substrate refractive index. The results indicated that, within the range of this experiment, the dominant effect was leakage, whereas the absorption contribution was much smaller than the experimental uncertainty. The leakage effect can be seen clearly from the modal field distribution in a typical guide which is plotted in figure 2.7.

The loss was influenced by a combination of the effective index and the thickness of the low-index buffer layer, since the former determined the extent of the evanescent field in the oxide while the latter governed the width of the isolation barrier between the guide and the underlying silicon. The losses are plotted in figure 2.8, versus predicted values.
Figure 2.7. Modal distribution in a leaky silicon nitride waveguide. The $e^{i\omega t}$ term gives rise to rotation about the vertical axis as shown. The field in the silicon region propagates downwards.

All the measurements were in reasonable agreement with the predicted values. This lent credence to the model predictions within the leakage regime, i.e. where absorption losses were not significant. The comparison to literature (see § 2.1.5), on the other hand, provided validation in the presence of layers with very high loss. While
not constituting proof positive of the model's validity, this provided substantial quantitative evidence in support of it.

2.1.7. MOD2: scope and limitations

If the verification described above provided grounds to believe that the model predictions were reliable, there remains the caveat that the model does not take into account scattering losses from structural imperfections (Rayleigh scattering). Having said this, the visualisation program, MPLOT, provides an indication of the power density as a function of depth in the structure, from which one can estimate the relative influences of scattering centres in various layers. An important example of this is scattering due to particles of dust on the guide surface: if the power density at the surface of the guide is very low, these particles tend not to cause much scattering.

MOD2 only simulated slab guides, i.e. structures in which there was no lateral confinement. This was not felt to be a severe limitation, since the effects of lateral confinement can be calculated (in the lossless régime) using finite element techniques. If the effects of loss and lateral confinement are small perturbations on the guided mode (i.e. if the associated \(k\)-vector components are small compared to the real part of \(\beta\)), then one may reasonably assume the second-order effects to be small. A complete (though approximate) solution is then obtained by superimposing the losses, calculated using MOD2, on the real \(\beta/k_0\) values and modal distributions from finite element calculations.

2.2. Electrical modelling: VMEFF

Electrical modelling was needed for assessing the performance of potential device structures. In a parallel-plate capacitor configuration, the inter-electrode capacitance is greater than in coplanar geometries. Constraints on the electrode material led to comparatively large resistivities, as will be seen in the next chapters. Hence the distributed RC time constant of the electrode structure was an important consideration, resulting in phase and amplitude variations of the electric field across the width of the waveguide channel.

A computer model, VMEFF, was developed to calculate these effects for the test structures reported in this study.
2.2.1. VMEFF: mathematical basis

Figure 2.9 shows a schematic diagram of a section of a modulator structure, comprising a dielectric slab of thickness $h$ and width $W$, sandwiched between two electrodes. The lower electrode is grounded on one side: a sinusoidal signal is applied to the opposite side of the upper electrode. By symmetry arguments, it is clear that no current flows along the modulator.

Let the following terms be defined:

- subscripts $u$, $l$ denoting upper and lower electrodes respectively;
- potential $V_u(x)$, $V_l(x)$;
- charge density $\sigma_u(x)$, $\sigma_l(x)$ per unit area;
- current density $j_u(x)$, $j_l(x)$ per unit length of electrode;
- dielectric constant $\varepsilon$ of the dielectric slab;
- electrode sheet resistances $\rho_u$, $\rho_l$.

Let the signal applied to the upper electrode be $V_0 e^{i\omega t}$.

Considering a Gauss' cylinder through one electrode,

$$\sigma_u = \varepsilon \varepsilon_0 (V_u - V_l)/h$$

(2.8)
Ohm's Law implies

\[ j_u = \frac{1}{\rho_u} \frac{\partial V_u}{\partial x} \]  \hspace{1cm} (2.9)

Charge conservation implies

\[ \frac{\partial j_h}{\partial x} = - \frac{\partial \sigma_u}{\partial t} = - \text{i} \omega \sigma_u \] \hspace{1cm} (2.10)

Differentiating (2.9) with respect to x, and combining with (2.10) and (2.8),

\[ \frac{\partial^2 V_u}{\partial x^2} = \text{i} \omega \varepsilon_0 \rho_u (V_u - V_l)/h \] \hspace{1cm} (2.11)

By symmetry, subscripts may be swapped: subtracting (2.11) from its counterpart, and writing \( \Delta = V_u - V_l \), this leads to a wave equation in \( \Delta \):

\[ \frac{\partial^2 \Delta}{\partial x^2} = \text{i} \omega \varepsilon_0 (\rho_u + \rho_l) \Delta/h \]

This has a general solution

\[ \Delta = \Delta_+ e^{-(1+i)kx} + \Delta_- e^{+(1+i)kx} \]

where \( k = \pm \sqrt{\omega \varepsilon_0 (\rho_u + \rho_l)/2h} \).

Now substituting into (2.8), equation (2.10) may be integrated to give

\[ j_u = \frac{(1+i)k}{\rho_u + \rho_l} \left\{ \Delta_+ e^{-(1+i)kx} - \Delta_- e^{+(1+i)kx} \right\} + c_u \]

where \( c_u \) is a constant of integration. Substituting this into (2.9) and integrating gives

\[ V_u = \frac{\rho_u}{\rho_u + \rho_l} \left\{ \Delta_+ e^{-(1+i)kx} + \Delta_- e^{+(1+i)kx} \right\} - c_u x + d_u \] \hspace{1cm} (2.12)

with a similar equation for \( V_l \). Equation (2.12) may be subtracted from its counterpart to give an equation for \( \Delta \), from which it appears that \( \rho_u c_u = \rho_l c_l \) and \( d_u = d_l \). The four unknowns \( \Delta_+ , \Delta_- , c_u , d_u \) may then be evaluated using the boundary conditions \( V_l(0) = 0 , V_u(W) = V_0 , j_u(0) = 0 , j_l(W) = 0 \). Integrating \( \Delta/h \) from 0 to \( W \), leads to an expression for the average electric field,

\[ \overline{\Delta/h} = \frac{(1-i)(\Delta_+ + \alpha \Delta_-)(1-1/\alpha)}{2Wkh} , \] \hspace{1cm} where \( \alpha = e^{(1+i)kW} \)
Note that the apparent asymmetry in $\Delta_+$ and $\Delta_-$ is due to the choice of origin. These terms represent counter-propagating waves launched from opposite sides of the structure, attenuating with distance. This is visible in figure 2.10, which shows plots of potential distributions in the structure at various frequencies above and below the bandwidth. Bearing this in mind, the quantities $\Delta_+$ and $\alpha \Delta_-$ are symmetrical. This nuance was important when writing the computer model, where a judicious choice of variables was used to avoid numerical overflow problems.

Figure 2.10. Applied electric field (complex) as a function of position (x) across the waveguide, at frequencies near the bandwidth ($f_{bw}$), showing attenuation and phase lag. Time-dependence ($\exp(i\omega t)$) is represented by rotation about the x axis as shown.
Finally, the electrical impedance of the structure was obtained from $V_o/j_0(W)$:
As well as providing useful information for electrical driving requirements, this was an important quantity since it could easily be measured and could therefore be used for testing the accuracy of the model as well as for measuring the dielectric constant $\varepsilon$.

2.2.2. VMEFF: implementation

These equations were used to write a computer program. This was written in Pascal, since the computations were fairly short and running speed was not crucial. The high compilation speed of Pascal, on the other hand, made it easy to explore a range of characteristics simply by modifying the program.

VMEFF used sets of structural parameters (thicknesses, etc) including optical data provided by MOD2 to calculate device characteristics such as the electro-optic modulation as a function of driving frequency, and the bandwidth and optical absorption loss as a function of electrode doping levels: these results will be presented in Chapter 3.

2.2.3. VMEFF: validation

No experimental validation was performed at this point since the situations covered by this model were relatively simple. Qualitative arguments were used to justify the general shapes of the curves which were obtained from the calculations and which are presented in figure 2.11. Some degree of quantitative confirmation of the model was obtained in special cases (very low or very high frequency), where simplifying assumptions could be made. At low frequencies, when the RC time constant was small compared to the cycle period, the electrodes became fully charged, and the field was simply $V/h$. At very high frequencies, the capacitative coupling was strong so that the electrodes were practically short circuited, and the impedance was dominated by resistance, which was then determined by the upper and lower electrode resistances in parallel.

The bandwidth was of the order of $1/RC$, where $R$ was given by this high-frequency value, and $C$ was the total capacitance. Although the situation was somewhat complicated by the distributed nature of the resistance and capacitance, this
yielded an order-of-magnitude estimate of bandwidth. The model calculations were in good accord with these estimates, as indicated in table 2.1.

(a) modulation vs. frequency: this is a plot of the electro-optic modulation amplitude, averaged over the width of the device, and normalised with respect to DC values.

(b) electrical impedance vs. frequency, for the same structure as above.

Figure 2.11. Typical device characteristics versus frequency, predicted by the computer model VMEFF.
Table 2.1. Comparison between VMEFF calculations and order-of-magnitude calculations for a typical vertical modulator structure.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resistance: (high frequency)</td>
<td>2.5 Ω</td>
<td>2.5 Ω</td>
</tr>
<tr>
<td>Capacitance: (low frequency)</td>
<td>1857 pF</td>
<td>1857 pF</td>
</tr>
<tr>
<td>Bandwidth:</td>
<td>34 GHz</td>
<td>28 GHz</td>
</tr>
</tbody>
</table>

2.3. Summary

The computer model MOD2 was written for optical modelling of multilayer waveguides employing optically absorbing media. MOD2 was verified by comparing calculations to experimental data covering a wide range of materials. The application of MOD2 to electro-optic device structures will be discussed in Chapter 3. As will be described in Chapter 4, the solver algorithm of MOD2 was also applicable to the task of analysing waveguides, to obtain thickness bulk refractive index values from waveguiding measurements.

The electrical behaviour of the parallel-plate electrode geometry was modelled using VMEFF. This program modelled potential distributions across the width of the modulator structure, as functions of material properties, geometry and frequency, so as to calculate device bandwidth and impedance.
References


Chapter 3: Device modelling

Abstract

The mathematical tools developed in the last chapter were put to use in exploring a variety of structures. In this chapter, figures of merit are introduced and defined, enabling a direct comparison between different arrangements. The electro-optic efficiency is expressed as a combination of electrostatic and optical efficiency terms. It is shown that the geometrical efficiency of the parallel plate configuration is offset by the large optical losses due to the proximity of metal electrodes: these losses may be reduced by means of buffer layers, but this severely reduces the electrostatic component of the electro-optic efficiency.

In order to reduce these optical losses, the use of semiconductor electrodes was explored. As well as forming efficient device structures, this offered the advantage of being largely compatible with the idea of monolithic integration with semiconductor ICs. A family of modulators was designed. The electrical model, VMEFF, was used to predict the electrical characteristics of these structures.

3.1. Figures of merit

A number of indices of performance were defined, in order to quantify the assessment of various arrangements.

3.1.1. The active length, \( L_\pi \), and optical efficiency, \( \gamma \)

Clearly, a figure was needed to measure the electro-optic efficiency of a modulator structure. The \( \pi \) modulation length was selected: this being the active length, in a single-sided modulator (i.e. one in which the reference beam is unmodulated, as opposed to a push-pull configuration) required in order to achieve 100% intensity modulation which corresponds to a phase change of \( \pi \).

In order to calculate this, a given structure was first optically modelled, by means of MOD2, using the standard refractive indices. The calculation was then
repeated, using indices perturbed by the electric field applied to the electro-optic region. From these two calculations, the derivative $d/dn \{\beta/k_o\}$ was obtained:

$$\frac{d\{\beta/k_o\}}{dn} = \frac{\{\beta/k_o\}(\text{perturbed}) - \{\beta/k_o\}(\text{unperturbed})}{\Delta n}.$$ 

Finally, the active length $L_\pi$ could be calculated by considering the phase change over a distance $L$:

$$\Delta \phi = L_\pi \Delta \beta,$$

so that

$$\pi = L_\pi \Delta \beta = L_\pi k_o \frac{d\{\beta/k_o\}}{dn} \Delta n.$$ 

The index change, $\Delta n$, was obtained from the electro-optic relationship which relates changes in the refractive index ellipsoid to the applied electric field (see Appendix 3.A):

$$\Delta n = -\frac{1}{2} n^3 r_{33} E;$$

writing $k_o = \frac{2\pi}{\lambda}$, and $\gamma = \frac{d\{\beta/k_o\}}{dn}$,

$$L_\pi = \frac{\lambda}{n^3 r_{33} E \gamma} \quad (3.1)$$

The electric field $E$ was calculated using a standard driving voltage of 5V, applied between the two electrodes.

From these equations, it is apparent that $L_\pi$ is inversely proportional to the electric field (for a given voltage), and is therefore directly proportional to the thickness of the electro-optic region in the absence of passive dielectric buffer layers. The effect of buffer layers is to weaken the electric field and increase $L_\pi$; this will be discussed in more detail below. This initial result, however, is important as it indicates that the inter-electrode gap should be small if the structure is to be efficient.

Although the active length, $L_\pi$, was useful for comparing device structures from the point of view of operating devices, it was sometimes useful to examine the electrical and optical aspects separately, in order to gain insight into the reasons underlying the behaviour of these structures. $L_\pi$ also depends on $\gamma$, which will henceforth be referred to as the "optical efficiency" of the structure, as it relates the change in the effective refractive index of the guide to the change in bulk index in the electro-optic layer. $\gamma$
depends not only on the fraction of appropriately polarised light travelling in the electro-optic region, but also on the nature of the optical field in that region (i.e. whether it is evanescent or not).

\( \gamma \) may be derived from Maxwell's equations, as described by Kogelnik\(^1\); the change in \( \beta \) is given by

\[
2P \Delta \beta = \varepsilon_0 \omega \int \Delta \varepsilon \mathbf{E} \mathbf{E}^* \, dx
\]

(3.2)

where \( P \) is the power flowing along the waveguide, given by the integral of the time-averaged \( z \)-component of the Poynting vector \( S_z = \text{Re}(\mathbf{E} \times \mathbf{H}^*) \) over the waveguide cross-section, and \( \Delta \varepsilon \) is the change in dielectric constant at the optical frequency \( \omega \). The modal profiles obtained from the computer model, MOD2, were normalised such that the normalised power flow was equal to 1W/cm. It was therefore convenient to rearrange this in terms of \( S_z \). Using the same notation as in Chapter 2, the electric field components of a TM-polarised guided wave may be expressed in terms of the superposition of upward and downward travelling waves,

\[
E_x = \frac{\beta}{k} \{ E_+^0 \exp(-i k_x x) + E_-^0 \exp(+i k_x x) \}
\]

\[
E_z = \frac{k}{k} \{ E_+^0 \exp(-i k_x x) - E_-^0 \exp(+i k_x x) \}
\]

where \( k \) is the wavenumber in the guide medium, as opposed to the wavenumber \( k_0 \) in vacuo; similarly, making use of the well-known relationship between the electric and magnetic fields of an electromagnetic wave, i.e.

\[
\varepsilon \varepsilon_0 E^2 = \mu_0 H^2
\]

the transverse magnetic field is then given by

\[
H_y = n \sqrt{\frac{\varepsilon_0}{\mu_0}} \{ E_+^0 \exp(-i k_x x) + E_-^0 \exp(+i k_x x) \}
\]

\[
= \frac{n^2 k_0}{\beta} \sqrt{\frac{\varepsilon_0}{\mu_0}} E_x
\]

In most cases of practical interest, \( \beta/k \) is close to 1, (i.e. \( \beta/k_0 \) is not far different from the refractive index of the guide medium), so that \( k_x/k \) is small. The \( z \) components of electric field may therefore be neglected, so that the right hand side of equation (3.2) may be re-written
\[ \varepsilon_0 \omega \int \Delta \varepsilon \cdot E \cdot E^* \, dx = \varepsilon_0 \omega \int \Delta \varepsilon \cdot E_x^2 \, dx \]

\[ = \frac{\beta/k_0}{n^2} k_0 \int \Delta \varepsilon \cdot E_x H_y \, dx \]

where the relationship \( \sqrt{\varepsilon_0 \mu_0} = k_0/\omega \) has been used.

Substituting into equation (3.2), replacing \( \Delta \varepsilon \) by \( 2n \Delta n \), and dividing by \( 2k_0 \), the change in \( \beta/k_0 \) may be obtained:

\[ \Delta \left\{ \beta/k_0 \right\} = \int \Delta n \frac{\beta/k_0}{n} E \wedge H^* \, dx, \text{ so that} \]

\[ \gamma = \int \frac{\beta/k_0}{n} E \wedge H^* \, dx, \]

where the integration is taken over the stratum in which the index is altered. The \( E \wedge H^* \) term represents the fraction of energy travelling in the modulation region, and the fraction \( \{\beta/k_0\}/n \) may enhance or diminish the efficiency depending on whether the field in that region is evanescent or propagating. In the case of TE modulation, a similar equation applies, but the expression for \( H \) in terms of \( E \) is altered, giving

\[ \gamma = \int \frac{n}{\beta/k_0} E \wedge H^* \, dx \]

In most cases, \( \gamma \) was calculated by comparing perturbed and unperturbed values of \( \beta/k_0 \), as described above. This alternative analysis, however, afforded a useful insight into the situation, as well as providing a method of calculating \( \gamma \) from an approximate knowledge of the field distribution. The latter was useful for examining situations where the refractive index ellipsoid was not aligned with the waveguide axes, as will be seen in Chapter 6.

### 3.1.2. Optical attenuation

Another important figure was judged to be the optical attenuation of a device. This is clearly a matter of importance in the design of any optical system. The absorption loss, incurred over the active length \( L_a \), was selected as a useful figure of merit in this respect.
The insertion loss of a device depends on several other factors such as coupling losses and scattering. These were not calculated here, as they are not intrinsic to the modulator structure. Coupling losses depend on the structural discontinuities at the ends of the device, and therefore on the neighbouring optical elements. Scattering depends on inhomogeneities in the device structure. This was neglected here, because scattering was not expected to play a dominant role in determining the device loss. Although the figures vary widely with different deposition techniques, scattering losses of 1dB/cm or less are usually quite achievable with a suitable choice of deposition conditions for most optical materials. As will be seen below, typical device lengths were less than a millimetre, suggesting scattering losses of less than 0.1dB.

3.1.3. Bandwidth

Thirdly, the device bandwidth was calculated. One of the advantages of an efficient electro-optic interaction is that modulation can be achieved in a short active length, so that the electrical/optical velocity mismatch no longer limits the modulation bandwidth as tends to be the case in coplanar devices (see Chapter 1). In these devices, where the electrode spacing may be smaller and the capacitance correspondingly higher, the bandwidth may become limited instead by the electrical bandwidth of the electrode structure. Looking at the graphs presented in figure 2.11 of Chapter 2, which illustrate typical device characteristics as functions of frequency, the bandwidth is a measure of the frequency at which the device performance begins to degrade. In keeping with common practice, the bandwidth was taken to be the frequency at which the modulation amplitude was attenuated by 3dB with respect to DC.

3.2. Optical investigations

The optical waveguiding model, MOD2, was used to investigate the behaviour of a variety of structures, calculating the figures of merit where possible in each case.

3.2.1. Metallic electrodes

The straightforward approach to a parallel-plate structure, using metal electrodes with low-index buffer layers above and below an electro-optic layer, suffered from
high losses (of the order of several hundred dB/cm or more). These losses stemmed from surface plasmons\cite{2} excited on the metal planes.

Surface plasmons are guided modes in the TM polarisation at a metal-dielectric interface. They arise from the largely negative dielectric constant of the metal: this, in turn, is due to free electrons in the metal, which oscillate in antiphase to the optical electric field. The finite resistivity of the metal provides a mechanism for dissipating the oscillation energy, so that these modes exhibit very high losses. This mechanism is the same as that responsible for free carrier absorption in semiconductors, and which is discussed below in § 3.4.1. and in Appendix 3.B. Table 3.1 shows the effective indices and losses, calculated using MOD2, associated with niobium, aluminium, gold and silver: these are given for interfaces with air and with z-cut lithium niobate. In the latter case, optical efficiencies were calculated and are presented in the last column, labelled "γ". The bulk refractive index data in column 1 were obtained from reference [3]. This illustrates the very high surface plasmon losses, and indicates that the lowest losses are obtained using silver.

Table 3.1. Characteristics of surface plasmons at 0.6328 μm

<table>
<thead>
<tr>
<th>Metal (index)</th>
<th>Air:</th>
<th>LiNbO₃:</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\beta/k_0)</td>
<td>loss (dB/cm)</td>
<td>(\beta/k_0)</td>
</tr>
<tr>
<td>Niobium 2.5-2.6i</td>
<td>0.9993</td>
<td>33,100</td>
</tr>
<tr>
<td>Aluminium 1.2-7.0i</td>
<td>1.0095</td>
<td>2,930</td>
</tr>
<tr>
<td>Gold 0.155-3.21i</td>
<td>1.0519</td>
<td>4,680</td>
</tr>
<tr>
<td>Silver 0.0663-4.045i</td>
<td>1.0320</td>
<td>949</td>
</tr>
</tbody>
</table>
Figure 3.1. Modal profile for silver/z cut lithium niobate surface plasmon.
The high optical efficiency figures, over 100% for all but the niobium, may be interpreted partly as a consequence of the reversal of the z-component of the Poynting vector in the metal, which is shown in figure 3.1 along with the electric and magnetic field distributions. The negative Poynting vector in the metal is compensated for by more than 100% of the total flow in the lithium niobate. In addition to this, $\beta/k_0$ is considerably higher than the index of lithium niobate, further enhancing $\gamma$ as discussed above (§ 3.1.1). The optical efficiency, depending to first order upon the fraction of total energy flowing in the lithium niobate, is thus over 100%.

A sandwich of lithium niobate between silver electrodes was modelled: the characteristics, as a function of lithium niobate thickness, are presented in figure 3.2. The fundamental and the first harmonic modes converge on the performance figures of the surface plasmon. As the layer becomes thicker, these modes begin to look like weakly coupled plasmons on the two interfaces. The higher order modes, meanwhile, exhibit the standing waves one might normally expect to see in a resonant cavity. This is illustrated by the modal profiles presented in figure 3.4. The more uniform energy distribution across the structure, in the higher order modes, gives rise to a gradual decline of the loss, as the metal represents a dwindling fraction of the structure and the effective index and optical efficiency approach the bulk figures for lithium niobate. Even for these modes, however, the losses do not become low enough to operate a practical device, unless the layer of lithium niobate is made so thick that the geometrical advantage of the parallel-plate structure is lost. The TM2 mode, for example, reaches an optical efficiency of nearly 100% for a lithium niobate thickness of 6000Å, implying an active length of only 230μm, but the loss of 800dB/cm yields a device loss of 18dB. Although it might be argued that this could be reduced by using even thicker lithium niobate layers, the lithium niobate layer quality would have to be very high to avoid losses via mode conversion. There is evidence, as will be discussed in Chapter 4, to suggest that the surface roughness (and hence also scattering, both to radiation modes and between guided modes) increases with the thickness of the deposited film.

Surface plasmon-type losses might be avoided by using TE modes rather than TM: indeed, the TE mode losses were found to be slightly lower. Although TM modes were required in order to utilise the vertical applied electric field with the large $r_{33}$ electro-optic coefficient of lithium niobate for efficient modulation, TE modes could be modulated either using the $r_{13}$ electro-optic coefficient, or using an off-axis orientation of lithium niobate so that the $r_{51}$ coefficient might be accessible (as will be discussed in some detail in § 6.1 of Chapter 6). Figure 3.3 shows the characteristics of the TE modes of these structures, for direct comparison to figure 3.2. In spite of the much
lower electro-optic coefficient, the structures are generally more efficient in this polarisation. This is largely due to the $n/(\beta/k_0)$ factor in the optical efficiency, which is about 3.5 near the cutoff for the TE0 mode (1000Å lithium niobate thickness). Even so, the device losses are high: the optimum structure for TM uses 1000Å lithium niobate, yielding $L_\pi = 26\mu m$, with a loss of 47dB, whereas TE with 2000Å lithium niobate yields $L_\pi = 100\mu m$, with a loss of 20dB.

Figure 3.2. Optical waveguiding characteristics of silver/lithium niobate/silver structure, versus lithium niobate thickness (TM modes, $\lambda = 0.6328\mu m$).
Figure 3.3. Optical waveguiding characteristics of silver/lithium niobate/silver structure, versus lithium niobate thickness (TE modes, $\lambda = 0.6328 \mu m$).
Figure 3.4. Modal distributions (transverse field) of the silver/lithium niobate/silver structure (\( \lambda = 0.6328 \mu m \)).
3.2.2. Low-index buffer layers

One way to reduce the optical absorption losses was to use low-index buffer layers, such as SiO₂ (n = 1.46). The main drawback with this was that the electric field across the LiNbO₃ was then greatly reduced, particularly since LiNbO₃ has a large dielectric coefficient (about 30, compared to about 3.8 for SiO₂). This may be seen by considering the potential drop, $V_D$, applied to a multilayer stack. If layer $i$ has thickness $h_i$, dielectric constant $\varepsilon_i$, and experiences a field $E_i$ and a potential drop $V_i = h_i E_i$, Gauss' Law implies

$$\varepsilon_i E_i = D,$$

which is a constant for all $i$.

This leads to an expression for the electric field across layer $j$:

$$E_j = \frac{V_0}{\varepsilon_j \sum \{h_i / \varepsilon_i\}}$$

From this it can be seen how the potential drop is divided across the layers in proportion to their "electrostatic thickness" $h_i / \varepsilon_i$. The potential drop across the lithium niobate is thus reduced, out of proportion to the simple geometrical factor. For example, if a simple metal-lithium niobate-metal sandwich is modified by inserting a thickness of silicon dioxide of thickness equal to that of the lithium niobate, the field across the lithium niobate drops by a factor of 8.9; one such buffer on each side reduces the field by a factor 16.8.

Another argument against the use of these layers was that they would carry optical energy, so that the modulation of the lithium niobate would only act on a reduced fraction of the light.

The characteristics of symmetrical buffered-electrode structures were studied. A typical set of results is plotted in figures 3.5 (TM) and 3.6 (TE), versus the thickness of the lithium niobate layer, using silver electrodes and 0.2μm SiO₂ buffer layers. Typical modal profiles are shown in figure 3.7. The losses were greatly reduced using buffers at least 0.2μm thick, but these structures were multi-moded. $L_\pi$ generally decreased with the lithium niobate thickness, as the proportion of optical energy in the electro-optic region increased, reaching a minimum value of rather more than 1000μm (TM0 mode) for a lithium niobate thickness of 0.5μm or so. Beyond this, $L_\pi$ began to increase once more, as the electric field became weakened by the thickening of the lithium niobate layer.
Figure 3.3. Characteristics of the symmetrical buffered-electrode structure, using silver electrodes and 0.2μm silica buffer layers, as functions of lithium niobate layer thickness ($\lambda = 0.6328\mu m$).
Figure 3.6. Characteristics of the TE modes of the symmetrical buffered-electrode structure, using silver electrodes and 0.2 \( \mu \text{m} \) silica buffer layers, as functions of lithium niobate layer thickness (\( \lambda = 0.6328 \mu \text{m} \)).
3.2.3. Structures employing silicon electrodes

Alternative structures were examined, including the use of silicon as an electrode material. The optical absorption in such devices could be kept low by operating above the silicon band-gap wavelength (c. 1.2μm). These structures seemed particularly interesting because of the possibility of fabricating on a silicon or silicon-on-insulator substrate, eventually integrating modulators with silicon ICs.

Figure 3.7. Modal profiles (transverse field) of the buffered metal electrode structure (λ = 0.6328μm).
Figure 3.8. Characteristics of silicon/lithium niobate/silicon structure, versus upper silicon layer thickness. The lithium niobate thickness is 0.5μm; λ = 1.53μm.
It was found that a simple silicon-lithium niobate-silicon sandwich yielded promising results. This structure has the peculiarity that it is a composite waveguide. The refractive index of silicon is very high (c. 3.5 at 1.5μm wavelength), so that each silicon layer tends to behave as a separate optical waveguide. If the structure is sufficiently thin, the overlap between the silicon-waveguide modes becomes strong enough to phase-lock them together to form composite modes. For very thin structures, only one (symmetric) mode is supported: as the thicknesses of the layers are increased, first an anti-symmetric and then higher order modes begin to appear. The properties of the guided modes of some comparatively thick structures (layer thicknesses of the order of 0.5μm) are plotted in figure 3.8 as functions of the thickness of the upper silicon layer. Alongside these graphs are the characteristics of the upper layer alone, i.e. the same structure except that the lower silicon layer thickness is zero. This shows how the properties of the composite structure are essentially a superposition of the characteristics of the individual silicon layers.

Although the study of thick waveguides was useful for understanding the behaviour of this type of structures, it was felt unlikely that such thick multi-moded structures would provide practical devices for the present purpose. This was because of the difficulty in ensuring good isolation between the modes: unless layer quality was extremely good, inter-mode scattering would make it very difficult to isolate and study one optical mode independently of the others. The great thickness of the lithium niobate layer would exacerbate this situation, both by contributing a great deal of scattering and via the long Lπ.

Thinner structures were modelled, in which the layer thicknesses were of the order of 0.2μm. These were found to be single- or double-moded, and the optical efficiency yielded Lπ’s of the order of a few hundred microns.

For fixed upper electrode and lithium niobate thicknesses, a well-defined optimum (in terms of electro-optic efficiency) was found for the thickness of the lower electrode (see figure 3.9). Similarly, the upper electrode thickness could be optimised. Modal plots, shown inset in figure 3.9, indicate that these performance maxima occur when the optical energy is fairly equally distributed between the two silicon layers, maximising the energy in the active layer of lithium niobate between them.
Figure 3.9. Characteristics of the silicon/lithium niobate/silicon structure, as functions of the thickness of the lower silicon layer thickness. ($\lambda = 1.53\mu m$).
Figure 3.10. Characteristics of the silicon/lithium niobate/silicon structure, as functions of lithium niobate thickness. ($\lambda = 1.53\mu m$).

It was found that $L_\pi$ increased approximately linearly with the thickness of the electro-optic region, as shown in figure 3.10. Although $\gamma$ increased slightly as the lithium niobate layer was made thicker, the behaviour of $L_\pi$ was dominated by the $E$ term in equation (3.1), and hence by the inter-electrode gap. $\beta/k_\nu$ was high, because of the preponderance of high-index material (silicon, $n = 3.47$) in the structure, so that the ($\beta/k_\nu/n$) ratio was usually greater than 1, enhancing the optical efficiency in the TM0 mode.

Figure 3.10 also illustrates how $L_\pi$ tends to a finite value as the lithium niobate layer thickness tends to zero. This is because for small lithium niobate thickness $h$, $\gamma$ is proportional (to first order) to the fraction of optical energy carried in the lithium niobate layer, which is in turn proportional to $h$. 

Thicknesses:
- Upper electrode: 0.2µm
- Lithium niobate: 0.0-0.3µm
- Lower electrode: 0.15µm

Width: 3µm
writing $\alpha = \frac{\beta/k_0}{n} E^\dagger \mathbf{H}^\dagger$, and substituting in equation (3.1), with $E = V/h$,

$$L_{\pi} = \frac{\lambda h}{n^3 r_{33} V \alpha h} = \frac{\lambda}{n^3 r_{33} V \alpha},$$

which is finite independently of $h$.

Although a real effect, this could not be experimentally verified (or exploited!) because of the finite breakdown strength of the dielectric layer; besides which, the device bandwidth, related to the lithium niobate thickness via the inter-electrode capacitance, dwindles to zero so that such a device would be practically useless.

An interesting detail in these results concerns the behaviour of the TM1 mode. Modal plots also showed that the energy density in the lithium niobate was low in the case of TM1. This, and the lower effective index, reduced the optical efficiency of the TM1 modes far below that of TM0. The absorption losses, on the other hand, could be made similar by a suitable choice of layer thicknesses. This suggested the possibility of interfering these two modes in a single waveguide for amplitude modulation. In a Mach-Zehnder arrangement (see Chapter 1), two signals are spatially separated by means of a split waveguide: in this alternative scheme, they may be separated in the spatial frequency domain instead, as two modes of the same guide. So long as the two modes remain distinct and one is modulated relative to the other, the output signals may be interfered to produce amplitude modulation. This will be discussed further in Chapter 6.

### 3.2.4. Effects of silicon on optical modulation

A second electro-optic effect was predicted to take place in these structures, because of the optical perturbation due to mobile charge carriers in the silicon electrodes\(^4\). These carriers, whether electrons or holes, are set in motion by the oscillatory optical electric field, contributing to the electric polarisation in the material. The effects on the real and imaginary parts of the refractive index can be calculated, using a classical oscillator model (see Appendix 3B).
These effects were expected to lead to optical modulation because of the capacitative charge induced on the electrodes, which was dependent to first order on the applied voltage. Calculations showed that the magnitude of this effect was small compared to the lithium niobate electro-optic effect (see Table 3.2). Not only were the individual contributions to the phase and amplitude modulation small, but also the effects were in opposite directions for similarly doped electrodes and would tend to cancel each other out. (This was because a field which removed carriers from one electrode would supply them to the other, and vice versa.) However, the effects might be measurable, particularly if the effect of the lithium niobate was greatly diminished for some reason (e.g. poor effective poling).

Table 3.2. Electro-optic effects of lithium niobate and silicon layers in the proposed structure: 0.2\(\mu\)m, 0.22\(\mu\)m, 0.165\(\mu\)m upper, lithium niobate and lower layer thicknesses respectively. The figures give changes in index, in parts per 10\(^6\) per volt of applied signal.

<table>
<thead>
<tr>
<th>Layer</th>
<th>(\Delta n)</th>
<th>mode</th>
<th>(\Delta (\beta/k_0))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>real</td>
<td>imag.</td>
<td>real</td>
</tr>
<tr>
<td>Upper silicon:</td>
<td>-40</td>
<td>-0.58</td>
<td>TM0: -10</td>
</tr>
<tr>
<td>(Induced charge 3.8x10(^{16})/cm(^3))</td>
<td></td>
<td></td>
<td>TM1: -11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TE0: -26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TE1: -13</td>
</tr>
<tr>
<td>Lithium niobate:</td>
<td>670</td>
<td>0</td>
<td>TM0: 420</td>
</tr>
<tr>
<td></td>
<td>187</td>
<td>0</td>
<td>TM1: 54</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TE0: 32</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TE1: 14</td>
</tr>
<tr>
<td>Lower silicon:</td>
<td>49</td>
<td>0.73</td>
<td>TM0: 7</td>
</tr>
<tr>
<td>(Induced charge 3.8x10(^{16})/cm(^3))</td>
<td></td>
<td></td>
<td>TM1: 10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TE0: 10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TE1: 32</td>
</tr>
</tbody>
</table>

These figures were calculated using p-type silicon for both upper and lower electrodes. The effects in the silicon layers are opposite to each other, although their direction relative to the lithium niobate effect depends on the poling state (i.e. the direction of the positive c-axis) of the latter.
3.3. Electrical modelling

The general form of the electrical behaviour (bandwidth, impedance) of these parallel-plate modulator structures was discussed in Chapter 2. The graphs at the top of figures 3.9 and 3.10 indicate the variations of bandwidth as a function of layer thicknesses for the structures discussed above, assuming an overlap width of 3\(\mu\text{m}\) and a fixed electrode resistivity of 8.3 m\(\Omega\)-cm. If instead the sheet resistance is assumed to be fixed, the bandwidth becomes independent of electrode thickness. The relationship between bandwidth and sheet resistance may be obtained by using the secondary label on the abscissa of the graph in figure 3.9.

These graphs suggest that bandwidths of the order several GHz should be achievable using suitably conductive electrode materials. The effects of velocity mismatch between the electrical signal and the optical wave (see Chapter 1) were estimated by calculating the effective signal wavelength seen by the optical wave. The optical signal velocity is of the order of \(c/n\), where \(c\) is the speed of light in vacuo and \(n\) is the effective refractive index: similarly, the electrical signal velocity is of the order of \(c/\sqrt{\varepsilon}\). Substituting \(n = 2.5\) and \(\varepsilon = 30\), the optical wave travels approximately twice as fast as the electrical wave. The electrical signal wavelength is given by \(c/f\sqrt{\varepsilon}\), where \(f\) is the frequency. At \(f = 10\text{GHz}\), this 1.2cm, so that the optical wave experiences an effective modulation wavelength of 2.4cm, which is almost two orders of magnitude longer than the active length of these devices. This was therefore not expected to play a significant part in limiting the bandwidth.

![Figure 3.11. \(L_n\) versus lithium niobate thickness, for a silicon/lithium niobate/silicon structure operating at 5GHz.](image)

The bandwidth plot in figure 3.10 shows how the bandwidth is directly proportional to the lithium niobate thickness, via the \(1/h\) dependence of the capacitance term in the RC time constant of the structure. This led to a direct compromise between
electro-optic performance \( (L_n) \) and bandwidth. For a given frequency, there was a clearly defined optimum lithium niobate thickness as illustrated in figure 3.11.

The width of the overlap region directly affected both the resistance and the capacitance of the structure, so that the bandwidth varied as \( 1/W^2 \). The results quoted above were based on widths of 3\( \mu \)m: the bandwidths could be increased by a factor of 9 by reducing this to 1\( \mu \)m, with no serious penalty in electro-optic performance except for the increased proportion of optical energy in the (unmodulated) lateral evanescent regions. The problems with this scheme were expected to lie in alignment and patterning difficulties (c. 1\( \mu \)m minimum overlap), and with the finite width required to support a guided mode, with a sufficient preponderance of optical energy within the overlap region (c. \( \lambda/2\pi \) or 0.2\( \mu \)m minimum guide width).

Electrical drive power was estimated by calculating the electrical energy stored in the device, assuming a 5V signal, multiplied by the operating frequency. Although a sinusoidal modulation could be achieved with lower power by means of a resonant circuit, such an arrangement would be unsuitable for signal transmission applications where the modulator would have to be terminated and impedance-matched to the drive circuitry. The electrical energy stored on a capacitor is given by \( CV^2/2 \), so that the total power at a frequency \( f \) is \( CV^2f/2 \). \( C \) is equal to \( \varepsilon_0WL_\pi/h \), with notation as used in Chapter 2. Substituting, the drive power is given by

\[
\frac{\varepsilon_0WL_\pi V^2f}{2h} = 160 \text{ mW},
\]

at \( f = 10 \text{ GHz} \), for a device with \( \varepsilon = 30, \ W = 3 \mu \text{m}, \ L_\pi = 320 \mu \text{m}, \ V = 5 \text{ Volts}, \ h = 0.2 \mu \text{m} \). Note that this power is not dissipated within the device itself but in the terminating load resistor.

### 3.3.1. Optical absorption vs. electrode resistivity

The electrical and optical properties of the structures have so far been discussed separately. There is, however, an important interaction, via the mechanism of free carrier absorption which was mentioned above (§ 3.2.4). The mobile carriers in the electrodes are necessary for electrical conduction, but they also lead to optical absorption. The carriers are set in motion by the optical electric field, and the finite resistance of the silicon provides a mechanism for dissipating this oscillatory energy. Fundamentally, optical power is coupled to the phonon structure of the material. The magnitude of the effect may be calculated, using a classical oscillator model, as
described in Appendix 3.B. The calculations are in good qualitative agreement with experiment[5], and are shown in figure 3.12, although the exact quantitative relationship depends on the values used for mobility and effective mass of the carriers. This interaction led to a trade-off between bandwidth and optical absorption loss, which is illustrated in figure 3.12 for a typical device structure.

The electrical properties of the electrodes are likely to depend on the deposition techniques chosen for each electrode. In the present study, PECVD doped hydrogenated amorphous silicon and epitaxial silicon-on-sapphire were used; these behave very differently from each other, as will be discussed in Chapter 5, and the properties of each depend on deposition conditions and subsequent processing.

\[ Q - U \]

\[ \begin{cases} 18 & 19 \\ 20 & \end{cases} \]

\[ \log \{ \text{dopant density, } /\text{cm}^3 \} \]

Figure 3.12. Absorption (\(\alpha\)) and resistivity (\(\rho\)) of p-type single crystal silicon, as functions of dopant density. The resistivity was calculated using data from Sze[7]; the absorption data points were obtained from Schmidts experiment[6]; the absorption curve was calculated using data from Sze in the relationship described in Appendix 3.A, using an effective mass of 0.28m_e.

3.3.2. Electrode depletion and accumulation

In the discussion so far, modelling of device behaviour has relied on the assumption that the charge induced on the electrodes resides at the interface with the lithium niobate. In a practical device, this assumption may not be valid: in particular, a weakly doped electrode may be driven into depletion[6] so that the field across the lithium niobate no longer varies linearly with the applied voltage. The situation is illustrated in figure 3.13. Essentially, the space charge density in the depleted electrode
material is limited to the dopant density, so that there is a simple relationship between the induced charge and the depletion depth. In the case of a uniform dopant density $N$, the depletion depth $D$ is proportional to the total induced charge $\sigma$:

$$\sigma = ND. \quad (3.3)$$

Gauss' Law\[7\] then implies a linear variation of electric field across the depletion region: one boundary condition is that the electric field is zero at the conducting boundary, so that

$$E = \frac{\rho x}{\varepsilon_s \varepsilon_0} \quad \text{and} \quad V = \frac{\rho x^2}{2 \varepsilon_s \varepsilon_0} \quad (3.4)$$

where $x$ is measured from the extremity of the depletion region, and $\varepsilon_s$ is the dielectric constant of silicon. Gauss's Law also provides a relationship between the electric fields $E_s$ and $E_l$ on either side of the silicon/lithium niobate boundary:

$$E_l = \frac{\varepsilon_s E_s}{\varepsilon_l} = \frac{\rho D}{\varepsilon_l \varepsilon_0}. \quad (3.5)$$

The total applied voltage $V_o$ may then be expressed as the sum of the potential drops across the depleted silicon and the lithium niobate:

$$V_o = \frac{\rho D^2}{2 \varepsilon_s \varepsilon_0} + \frac{\rho Dh}{\varepsilon_l \varepsilon_0} \quad (3.6)$$

where $h$ is the lithium niobate thickness. Finally, this quadratic equation may be solved for $D$, and the electric field across the lithium niobate may be calculated using equation (3.3).

The devices fabricated in this study employed hydrogenated amorphous silicon for the upper electrode (see Chapter 5). The active doping density of this material was low, so the effect of depletion was significant.

This calculation was incorporated into VMEFF. The calculated relationship between phase modulation and applied voltage is shown in figure 3.14. Note that the depletion effect only takes place for one polarity of applied voltage, corresponding to the removal of majority carriers from the weakly doped electrode.
In the opposite polarity, the electrode is driven into accumulation\cite{6}, and the space charge is limited only by the density of states which tends to be greater than $10^{19}/\text{cm}^3/\text{V}$ (even in amorphous material)\cite{8}, so that the accumulation depth and the concomitant voltage drop across it are small in comparison to the thickness of the lithium niobate and the total applied voltage respectively. This may be verified by calculating the field corresponding to a small voltage, say 0.1V, across the accumulation region. The space charge is then of the order of $10^{18}q_e/\text{cm}^3$ where $q_e$ is the electronic charge; substituting this into equation (3.3), the corresponding accumulation depth is given by

$$D = \sqrt{\frac{2\varepsilon_1\epsilon_0 V}{\rho}} \approx 110\text{Å}.$$
The electric field across the lithium niobate may then be obtained using equation (3.5), and comes to $17\text{V/}\mu\text{m}$, or $3.4\text{V}$ across a thickness of $0.2\mu\text{m}$. Thus almost all the voltage drop is across the lithium niobate, and this accumulation effect may be neglected for the present purposes.

Before leaving this topic, it should be added that a practical device is likely to suffer from flat-band shift[6], which is due to fixed charges in the dielectric (in this case, the lithium niobate layer). This leads to a finite electric field even for zero applied voltage, so that the transition between accumulation and depletion will be shifted from zero volts. The voltage shift is proportional to $\sigma_f/C$, where $\sigma_f$ is the fixed charge density, and $C$ is the capacitance per unit area, proportional to the dielectric constant divided by the dielectric thickness; assuming that the dielectric thickness and $\sigma_f$ are of the same order as that found in typical MOS transistors, where the flat-band shift is of the order of a volt, the corresponding shift in these devices may be expected to be reduced by a factor of the ratio between the dielectric constants, to about $0.2\text{V}$.

![Device structure diagram](image)

Device structure:

Upper electrode:
- thickness: $0.2\ \mu\text{m}$
- doping: $1.6\times10^{13}/\text{cm}^3$ (p)
- dielectric constant: 11.7

Lithium niobate:
- thickness: $0.2\mu\text{m}$
- dielectric constant: 65.5

Lower electrode:
- thickness: $0.15\ \mu\text{m}$
- doping: $5\times10^{18}/\text{cm}^3$ (p)
- dielectric constant: 11.7

Figure 3.14. Depletion depth and voltage across the lithium niobate layer, as functions of the total applied signal.
3.4. **Test structure design**

Having identified a promising family of parallel-plate modulator structures, some suitable test devices were designed.

The comparatively minor variations of performance with the layer thicknesses implied that fabrication tolerances would be quite wide, particularly since these devices were intended merely for demonstration rather than optimum performance. Another encouraging result to emerge from this analysis was that the lithium niobate layer would carry only a small fraction of the optical energy. This, coupled with the short active lengths, suggested that the constraints on the quality of the lithium niobate layer might be relaxed - in particular the absorption and scattering losses, and possibly even the requirement that the material be single crystal. Thus emerged the possibility of using oriented polycrystalline material, relaxing the requirement for a lithium niobate substrate from which to seed the layer.

Consequently, other substrate materials were considered, such as fused silica and sapphire. The choice of sapphire was particularly interesting, as it offered the possibility of combining modulators and silicon devices on a single substrate, since sapphire is an established substrate material for silicon IC fabrication, as will be discussed in Chapter 5. Optical modelling results for the silica and SOS-based structures were found to be similar to those using lithium niobate substrates. This was not unexpected, given the comparatively low fraction of energy carried in the substrate.

A standard device structure was selected for fabrication, using 0.15\(\mu\)m and 0.20\(\mu\)m lower and upper electrode thicknesses respectively, with 0.2\(\mu\)m lithium niobate. The choice of lithium niobate thickness was influenced by the need to ensure adequate electrical breakdown strength. The figure of 0.2\(\mu\)m was selected after experimental measurements of breakdown strength, as will be discussed in Chapter 4.

3.5. **Summary**

Parallel-plate modulator structures were optically modelled. The electro-optic efficiency of these structures was seen to depend on an electrostatic term, governed by the inter-electrode gap, and an optical efficiency term which depended on the fraction of optical energy travelling in the electro-optic layer, and on the ratio of \(\beta/\kappa_0\) to the refractive index of the layer in which modulation was taking place. This analysis showed that the inter-electrode gap, and the presence of passive buffer layers, exerted a
crucial influence on the device efficiency. This led to a conflict of requirements: a small gap with no buffer layers, as required for optical efficiency, meant that a substantial portion of optical energy would travel in the electrode material - which, being conductive, led to unacceptably large optical losses via free carrier absorption. These losses were enhanced, when using metallic electrodes, by surface plasmon effects whereby a concentration of optical energy was confined by the metal-dielectric interface. TE modes were found to offer remarkably high efficiencies; but the absorption losses, although lower than for TM modes, remained too high for practical devices. Low-index buffers alleviated this problem, but only at the expense of electrostatic efficiency.

The problem of optical loss was alleviated by using silicon electrodes with a thin layer of lithium niobate between them. By operating beyond the silicon bandgap, the optical losses were kept down to an acceptable level. This enabled a substantial overlap between the optical field and the electrode material. These structures were studied, and a practical device was proposed. The modulation effects in this structure were identified and evaluated. The devices were electrically modelled: although the prototype devices fabricated in this work were limited to the kHz regime by the high resistivity of available material, ultimate bandwidths in excess of 10GHz were predicted for devices employing material of lower resistivity, with corresponding optical absorption losses of the order of 3dB, an active length of 350μm, and $V_\pi$ of 5V.

Finally, it should be added that although this modelling has been based on accessing the $r_{33}$ and $r_{13}$ coefficients of lithium niobate, with the assumption that the material is aligned with its $c$ axis normal to the plane of the waveguide, the results may be applied to other situations by substituting an appropriately altered electro-optic coefficient. This was put into practice, as will be discussed in Chapter 6, for the analysis of the prototype devices, in which the lithium niobate grew with an $<012>$ orientation, whereupon the $r_{51}$ coefficient became accessible.

References


Chapter 4: Lithium niobate thin film deposition

Abstract

This chapter describes the development of a process for growing the thin films of lithium niobate which were required to fabricate the electro-optic device structures discussed in Chapter 3.

The chapter begins with an introduction, setting forth the aims and the programme philosophy. There follows a short discussion about lithium niobate, identifying the issues of crystal structure and stoichiometry seen to be of crucial importance in the device. The published literature on thin film growth of lithium niobate is reviewed. RF sputter deposition is briefly introduced. Apparatus was adapted to meet the requirements of lithium niobate deposition: this is discussed in some detail.

The growth studies are discussed. A range of analytical methods was used to measure film properties. Thickness, surface roughness and refractive index were studied using Talystep™, and optical waveguiding[1]. X-ray diffraction[2] was a particularly important technique, since it provided a rapid measure of the crystalline quality of the films. Dielectric behaviour was measured by depositing on conductive layers and making thin-film capacitors. Other film properties, such as pinhole densities and etching characteristics were also investigated.

4.1. Programme philosophy

Given the device-oriented purpose of the work, the emphasis of this materials programme was on the development of device quality thin films in the shortest reasonable time rather than a thorough characterisation of growth kinetics. The first step in such an exercise was to establish a set of criteria which the films had to fulfil. These were determined by the design and modelling work discussed in Chapter 3: essentially, the films had to be optically transparent, electrically insulating, and electro-optic. The required optical transparency was determined by the need for a reasonably strong output signal from the device. A contribution of 1dB to the device loss was felt to be an acceptable figure, corresponding to a guide attenuation of 25dB/cm. The electrical resistance through the lithium niobate had to be greater than that of the silicon,
to ensure that the electric field was applied across the electro-optic region rather than the electrodes. The waveguide structure was shown in figure 2.9 of Chapter 2. The resistance of the electrode structure is dominated by the upper electrode, with a resistivity of $10^4 \Omega \cdot \text{cm}$ in the devices fabricated here; assuming a lateral separation of 10 $\mu$m between the guide and the metal contact, this yields a conductivity of $2 \times 10^{-6} \Omega^{-1} \text{cm}$ per cm of active length. With a lithium niobate thickness of 0.2 $\mu$m and a waveguide width of 2 $\mu$m, this implies a minimum lithium niobate resistivity of $5 \times 10^6 \Omega \cdot \text{cm}$.

The analysis of Chapter 3 indicated that, theoretically, the electro-optic behaviour of the structure was dominated by the $r_{33}$ coefficient of lithium niobate, by about a factor of ten, over effects in the silicon (see Table 3.2 in Chapter 3). The electro-optic effect in these films therefore had to be within a factor of ten of the bulk figure of $31 \times 10^{-12} \text{m/V}$ in order to be unambiguously attributable to the lithium niobate.

### 4.2. Lithium niobate

The growth, structure and properties of lithium niobate are well documented\(^{[3-14]}\). The electro-optic effect in lithium niobate results from the lack of centrosymmetry in its crystal structure\(^{[3]}\). The growth of thin films of lithium niobate, suitable for the present electro-optic device application, therefore depends on obtaining the correct crystal structure and poling state. As was explained in Chapter 3, the strongest electro-optic coefficient, $r_{33}$ ($30.8 \times 10^{-12} \text{m/V}$)\(^{[14]}\), is put to most efficient use by aligning the applied electric field and the optical polarisation with the c-axis of the crystal: in the present context, this implies growing a film where the c axis is oriented normal to the surface. The $r_{51}$ coefficient is also strong ($28 \times 10^{-12} \text{m/V}$)\(^{[14]}\), and may be accessed when the c-axis is inclined to the surface normal as will be discussed in Chapter 6.

Lithium niobate may be viewed in terms of a combination of niobium oxide ($\text{Nb}_2\text{O}_5$) and lithium oxide ($\text{Li}_2\text{O}$) as illustrated in the phase diagram\(^{[5,6]}\) which is reproduced in figure 4.1. From this phase diagram, it appears that the formation of the electro-optic phase of lithium niobate (rhombohedral LiNbO$_3$) can occur over a narrow range of lithium:niobium ratio, about 2% or so at 500°C. On either side of this region, mixed phases are obtained. Although the rhombohedral phase does not necessarily vanish outside this range, it becomes diluted by other phases and the effective electro-optic coefficients may be expected to be reduced accordingly.
The phase diagram cited above is concerned with the existence regime of LiNbO$_3$ in terms of the lithium:niobium ratio, but does not give any indication of the influence of oxygen composition on the crystal structure. Little data is available on this subject, although it is reasonable to suppose that small variations in oxygen composition will result in vacancies or interstitial defects, while large oxygen deficiencies may be regarded as an excess of niobium and lithium, which are then likely to form interstitial defects. Such defects are likely to contribute to electrical conduction and optical absorption. Studies involving the processing of lithium niobate indicated that the tendency was for Li$_2$O to be released under thermal treatment\cite{15}: in the present work, the expectation was that there would be a tendency towards lithium deficiency as the deposition of crystalline layers involved high substrate temperatures.

![Phase diagram of the system Li$_2$O - Nb$_2$O$_5$](image)

Figure 4. 1. Phase diagram of the system Li$_2$O - Nb$_2$O$_5$\cite{4,5}

4.2.1. Lithium niobate thin films

The growth of lithium niobate in thin films is not new: the first reports date as long ago as 1969, using DC triode sputtering\cite{16} and the technique known as epitaxial growth by melting (EGM)\cite{17-20}, in which a powdered crystal is scattered on the substrate and then melted. Films were also grown by dipping into a melt\cite{20}. These
techniques produced electro-optic films of several microns in thickness, whereas the devices reported here called for films of no more than about 0.2μm in thickness. Thin films have been reported also by molecular beam epitaxy in this laboratory[21-23], but problems of oxygen deficiency and sheer process complexity made this an unattractive approach for this project.

Evidence from the literature[24-36] suggested that suitable films could be grown by RF sputtering. The transition temperature between amorphous and crystalline growth regimes, was reported by most authors as being between 450 and 600°C, depending on substrate material as pointed out by Hewig et al[28]; Griffel et al[30] reported the transition to be as low as 380°C. It is possible that this wide range reflects differences in measurement techniques: it was confirmed experimentally in this work that temperature measurements could vary by over 100°C depending on the position of the measurement thermocouple.

Most authors used sputter targets enriched with lithium oxide, so as to counteract the effects of lithium out-diffusion. All the authors cited[24-36] employed oxygen-argon mixtures for sputter ambient: a broad optimum was observed by Hewig at 30 ±10% oxygen. Sputter pressures varied between 2 and 20mTorr, although most of the work was at the high end of this range.

The choice of substrate material seems to have exerted a crucial influence on film properties. Growth on carefully lattice-matched material such as <111> MgO[26] and Gd₃Ga₅O₁₂[27] showed good c-oriented epitaxy, while attempts to grow on <111> silicon and c-sapphire resulted in strained layers. Several authors[28-30,35] reported oriented growth on glass substrates, mostly with preferred c-orientation although Hewig found <012> oriented growth. Only two authors[24,35] reported electro-optic coefficients in these films, and both were reduced by more than an order of magnitude with respect to bulk material. One study[32] showed from measurements of pyroelectrically induced charge that the c-oriented film (on c-sapphire substrate) had grown in the positive c-direction.

It should be added that in most cases, these studies were aimed at growing rather thicker films, c. 0.5μm, than warranted in the present work (0.1 to 0.2μm). As was remarked by several authors[28,34], the film quality tended to be better on the thinner films. This augured well for the fabrication of the proposed devices.
4.2.2. Poling

The crystal structure of lithium niobate\textsuperscript{9,10} is illustrated in figure 4.2. Essentially, this may be viewed as a succession of layers of oxygen ions arranged on hexagonal grid, stacked up in the c direction. The oxygen ions thus form an array of octahedra, each containing a niobium or lithium ion or a vacancy in a regular succession. The poling state of the material is governed by the ordering of lithium, niobium and vacancies in successive octahedra along the c direction. From this, it can be seen that the poling state of the material may be reversed by displacing each lithium ion along the c direction, effectively swapping it for the vacancy in the next octahedron\textsuperscript{37}. This process requires a supply of energy, to overcome the potential barrier which prevents spontaneous re-poling, and the application of an electric field in order to shift the ions in the desired direction.

![Figure 4.2. Schematic representation of the structure of lithium niobate, showing how the metal ions fit into the oxygen lattice.](image)

It may be argued that the mechanism responsible for preferred orientation on amorphous substrates should also give rise to poling of the growing layer. The preferred orientation is due to a strong orientation dependence of the average bonding energy of atoms, or groups of atoms (nascent crystallites) on the growing surface. This is in good accordance with the observed orientation dependence of chemical
etching of lithium niobate\textsuperscript{7,38}. Chemical etching has been shown to distinguish sharply between the positive and negative c faces\textsuperscript{38}. It is therefore quite possible that this orientation dependence should give rise to selectivity between the growth of microcrystallites, favouring those with the greatest bond energy and correspondingly the slowest etch rates, i.e. the positive c-face. This argument is supported by observations of electro-optic effects and of poled growth, which were cited above\textsuperscript{32}. As a result, it was expected that, if this work succeeded in growing oriented layers, no ex-situ poling would be necessary to ensure that the films exhibited an electro-optic effect.

A technique has been reported\textsuperscript{37,39} for re-poling at relatively low temperatures, i.e. well below the Curie point. This technique relied on electron bombardment to provide the energy necessary for the lithium atoms to overcome the potential barrier. This provided a possible method of poling the thin films grown for the present devices, should they turn out to grow without a predominant poling state.

4.3. RF Sputter deposition

RF sputter deposition is a well-known technique, which has been extensively covered in standard textbooks\textsuperscript{40} and other literature\textsuperscript{1,41-43}. A brief introduction is provided in Appendix 4.A.

The equipment used in this work was a Materials Research Corporation Sputter/Etch Module, model 8620; this was extensively modified before and during the present project. The system is shown schematically in figure 4.3. The vacuum process chamber contained a target electrode and substrate table. The target was a niobium tray, covered with close-fitting fragments of lithium niobate. Particulate contamination of the substrates was minimised by fixing them face downwards.

The process chamber was evacuated by a rotary-backed diffusion pump, to a base pressure of about $10^{-6}\text{Torr}$. A liquid nitrogen cooled cold trap was used to reduce the partial pressure of water and back-streaming from the diffusion pump. The substrates could meanwhile be heated to $500^\circ\text{C}$. Process gas (typically argon, or an argon-oxygen mixture) was introduced, and the pressure was controlled between 5 and 30mTorr by throttling the gate valve. An RF signal, of typically 100W at 13.56MHz, was then applied to the target electrode (of 10cm diameter), establishing a glow discharge in the region above the target. A shutter protected the substrates for the first few minutes of the discharge, both to allow the system to settle into a stable operating
regime and to sputter-clean the target. The shutter was then removed for a period (typically two hours or so) to allow deposition to take place on the substrates.

4.3.1. Substrate DC Bias

The composition and structure of the growing film are influenced by electric fields in the vicinity of the substrate. These fields govern the bombardment of the film by ions and electrons. There is a tendency for oxide films to grow oxygen-deficient,
because of the volatility of oxygen\textsuperscript{[1]}. This may be counteracted by introducing a partial pressure of oxygen into the discharge: an electric field across the substrate sheath can then be used to attract oxygen ions from the plasma into the growing film\textsuperscript{[43]}. 

The conventional method of establishing these electric fields is to apply a small RF signal to the substrate table: the rectifying behaviour of the sheath then gives rise to a DC field\textsuperscript{[43,44]} (see Appendix 4.A). This method suffered from being difficult to control. The RF signal for the substrate table was derived by power splitting in the impedance matching circuit; the split ratio was not easy to adjust without upsetting the impedance matching condition, and vice versa. This was of particular importance during the first few minutes of layer growth. Precise control was essential at this stage, since the first few monolayers could have a seminal influence on the structure of the rest of the film. When the sputter shield was removed to begin deposition, the RF power matching circuit was usually disturbed and required re-adjustment. 

An alternative method was developed: this consisted in applying a negative DC bias directly between the substrate table and ground, via an RF choke. This technique was tested, and found to be effective: control of the matching condition and bias became completely independent, and growth studies indicated that the bias was exerting an influence over film properties. 

A schematic diagram of the circuit is shown in figure 4.4. An accurate analysis of the situation is a profoundly complex problem involving the details of the behaviour of the RF plasma, and is not within the scope of this thesis. A plausible explanation, consistent with observations, is as follows. The substrate table was capacitatively coupled via the sheath to the plasma, and by small stray capacitance to ground. The negative DC bias repelled electrons: as the bias increased, this led to a thickening of the sheath and a consequent weakening in the RF coupling between the plasma and the substrate table. The RF voltage across the sheath therefore increased. The arrangement could thus be regarded as a convenient way of controlling the power split ratio. It should be noted that the DC bias induced on the surface of an insulating substrate is not, in general, equal to the directly applied DC bias, but is given approximately by half of the peak-to-peak RF voltage across the substrate sheath (see Appendix 4.A).
4.3.2. Substrate heating

An important reason for modifying the matching network was to incorporate substrate heating. Although the unit was fitted with a substrate heater, it was originally connected in such a way that it was impossible to heat while applying RF power; possibly this design was intended to protect the heater power supply from RF power.
The heater and the spike thermocouple were re-wired, via RF chokes, so as to overcome this (see figure 4.4).

A computerised controller was built in order to enable the temperature to be slowly ramped and monitored. This controlled the power input to the substrate heater by means of a belt drive to the Variac transformer knob: feedback was provided by measuring the Variac output voltage. Meanwhile, the temperature was monitored by means of a spike thermocouple set in the back of the substrate table. The system was calibrated by placing thermocouples on the substrate table in lieu of the substrates, then slowly increasing the heater power while monitoring the temperature. In this way, a plot of substrate table temperature vs. heater power was obtained and stored as a lookup table for the control software. The thermal mass of the system was such that a lag of up to 30 minutes was observed between setting and reaching a given temperature: this was taken into account in the thermal cycles for deposition runs. The long-term repeatability of the system was found to be better than 5°C, which was similar to the temperature variations across the deposition region.

Substrate heating provided structural problems as well as electrical. The substrate table was essentially a small chamber, containing a heater element and water cooling channels. The substrates were clipped to a plate which was bolted to the bottom of the table. Water cooling was necessary to avoid overheating nearby parts of the apparatus.

The table had to be modified in order to gain access to the heater element for servicing. The new design, shown in figure 4.5, used a bolted bottom plate with a copper gasket seal. The heated baseplate incorporated a radial bellows structure: this was to overcome a tendency for the gasket seal to spring vacuum leaks when the baseplate expanded and distorted at high temperatures or during thermal fluctuations. The radial bellows not only reduced heat losses to the periphery of the plate, so that the edge could be cooler for a given temperature at the centre, but it also enabled the hot central region to expand without unduly stressing the periphery, and the vacuum seal was retained even up to the highest temperatures.

A beneficial side-effect of this design was that, by a judicious arrangement of the bellows grooves, the compressive stress on the bellows was transferred to a downwards force, pressing the hot central region of the baseplate against the substrate retaining plate and ensuring a good thermal contact. This was demonstrated by the deposition pattern on the back of the retaining plate after a hot deposition run: material had diffused into the small but inevitable gap between the plate and the table, but there remained a sharply defined region in the centre where no deposition had taken place.
Figure 4.5. Substrate table assembly.

The improved thermal contact to the central region of the substrate plate was accompanied by a reduction in radial heat flow, because of the thermal impedance of the bellows structure as well as the slightly poorer thermal contact between the plate and the table near the edges. The reduced heat flow led to a corresponding reduction in the thermal gradient across the central region, so that the temperature uniformity was improved. These variations were measured to be about 5°C across the deposition area at a typical deposition temperature (400°C).
4.4. Growth studies

4.4.1. Film growth

A succession of lithium niobate growth runs were performed with the modified sputtering apparatus. A typical run schedule was as follows.

On the evening preceding the run, substrates were cleaned using trichloroethylene (with ultrasonic agitation), acetone and isopropyl alcohol (IPA); they were left in clean IPA overnight.

In the morning, the substrates were rinsed in de-ionised water and placed in a bath of hydrogen peroxide, to which was added concentrated sulphuric acid. Here they remained for 20 minutes. They were then rinsed in running de-ionised water for a minimum of 10 minutes, blown dry and clipped to the substrate holder. The substrate holder was immediately bolted into position, and pumpdown was initiated within ten minutes of blowing the substrates dry.

During pumpdown, the liquid nitrogen cold trap on the diffusion pump was charged, and an automatic refilling device was put in place to keep the trap charged for the duration of the process. The temperature was set to ramp up at 2°C per minute, to a setpoint typically 30°C above the process temperature (for the purpose of outgassing).

Half an hour after reaching this setpoint, the system was set to ramp down to the process temperature. When this was reached, sputter gas was admitted and the glow discharge was set up with the deposition shield in place. After a further half hour, when the temperature and process conditions had stabilised, the sputter shutter was opened to allow deposition.

At the end of the deposition, the RF system was switched off, the process gas valve was shut and the temperature was set to ramp down at 2°C per minute to room temperature. When the temperature dropped below 100°C, the liquid nitrogen cold trap autofiller was switched off. Specimens were not removed from the chamber until they reached room temperature.

Films were characterised to assess suitability for the purpose of devices. Thickness, roughness and optical properties were of obvious importance, and were examined using optical waveguiding and surface profilometry. Electrical characteristics were important, to ensure dielectric integrity (i.e. lack of short circuits) of the devices: this was measured using parallel plate capacitor structures. One of the most important
4.4.2. Optical measurements

A technique was devised to obtain refractive index measurements of very thin films (down to less than 100Å). This is illustrated in figure 4.6. Films were grown on characterised slab waveguides of silicon nitride on silicon dioxide. These were fairly thin (mostly monomode) guides, so that an appreciable fraction of the optical signal would travel in the lithium niobate overlayers. The effective refractive indices, $\beta/k_0$, were then measured using the prism coupling method\(^1\) in both coated and uncoated regions: the difference between these was ascribed to the presence of the deposited layer. There remained an ambiguity in the result, as there were two unknowns: the possible solutions to the problem could thus be plotted as a curve on a graph of layer index versus thickness. The ambiguity was resolved by repeating the measurements, this time with a high-index oil over the guide. This altered the optical energy distribution, giving a slightly different index-thickness curve: the intersection between the two graphs therefore yielded a unique solution. The experiment was performed both in TE and in TM polarisations, to obtain the film birefringence and to check for agreement between the TE and TM thickness values. The uncertainties in thickness and refractive index obtained by this technique are correlated, giving an "error parallelogram" as shown. A computer program was written to implement this, using the solver algorithm of MOD2 (see Chapter 2). The raw $\beta/k_0$ data was placed in an input file, including estimated error bars on each $\beta/k_0$ measurement: from this, the program would calculate thickness and bulk index values for the underlying nitride and the deposited film.
(a) perturbation $\Delta$ leads to a curve representing possible film parameters.

From measurements using oil cover

Intersection yields unique solution

(b) repeating the measurement using oil under the prism gives a second curve.

(c) repeating the exercise in the other polarisation, both film indices and the film thickness are obtained.

Figure 4.6. Optical measurement of film thickness and refractive indices.

Disagreements were observed between thickness measurements obtained in TE and TM. This was interpreted in terms of thickness or index variations across each sample. The analysis technique described above relied on the assumption that the
measurements with air and oil covers were of the same guide, i.e. that the film thickness and index in both measurements were identical: if the prism was displaced between measurements, and the film thicknesses or indices differ slightly between the two measurement sites, the assumption was invalidated and errors arose. This hypothesis was tested by computer simulation. The computer model was used to calculate an idealised set of $\beta/k_0$ measurements, together with the effects of small thickness and index variations (5% and 0.01 respectively) on the values of $\beta/k_0$; perturbed values were then substituted for the oil-covered measurements, and these were resubmitted to the analysis program. The result, illustrated in figure 4.7, shows that the calculated thickness and index values were distorted by about 10% and 0.04 respectively, which is similar to the discrepancies in experimental data. From this, it was inferred that the observed discrepancies could be accounted for by a surface roughness of the order of 5%, or index variations of the order of 0.01.

![Figure 4.7](image)

Figure 4.7. Effects of 50Å thickness variations on the film parameters calculated from optical waveguiding measurements. The crosses indicate the measured structures: the boxes and parallelograms show the computed film parameters and error margins assuming uncertainties of 0.0050 in $\beta/k_0$. The circles illustrate the effects of index variations of 0.01.

Note that the distortions in thickness and index both give rise to roughly parallel deviations, along the major axes of the error parallelograms on the index-thickness graph. As a result of this, the calculated figures can be used to estimate the film birefringence with an accuracy of typically about 0.01, i.e. more accurately than the absolute index values.
Films grown at low temperatures exhibited refractive indices lower than those of lithium niobate. This could be the result of a lower packing density, due to occlusion of argon and excess oxygen from the plasma as well as the amorphous structure of the material. Films grown at higher temperatures showed higher indices and marked birefringence was observed as illustrated in figure 4.8, in reasonable agreement with values expected for c-oriented rhombohedral lithium niobate.

![Graph showing birefringence vs. deposition temperature.](image)

**Figure 4.8.** Birefringence vs. deposition temperature. The transition temperature between amorphous and crystalline growth regimes was found to lie within the range indicated by $T_x$ (from X-ray data: see Section 4.4.5).

It was also observed that a significant amount of TE-TM conversion was taking place in the films. This was inferred from the extraneous mode lines which appeared when waveguiding at He-Ne wavelengths: when coupling into the TM0 mode, for example, a mode line would appear in a position corresponding to the TE0 mode. A rough estimate was that up to 20% of the power was converted, over a distance of two millimetres or so (i.e. the size of the coupling spot), in the most severe cases. This may be extrapolated to predict the severity of the effect in devices: the inverse-fourth dependence of scattering on wavelength\[^{45}\] reduces the effect by a factor of 34, so that the power conversion over a distance of 1mm at 1.53μm (compared to the measurement wavelength of 0.6328μm) may be expected to be much less than 1%.
4.4.3. Profilometry and thickness measurements

Attempts were made to measure the layer thickness by other methods, in order to corroborate the optical measurements. Talystep™ measurements were taken on the edges of shielded areas, but these were inconclusive as the edge profiles were distorted. An alternative scheme was devised, using lithium niobate on silicon-on-sapphire (SOS)
substrates, illustrated in figure 4.9. An edge was ion beam milled through the lithium niobate and part way into the underlying silicon; the remainder of the silicon was removed using an SF₆ plasma. This plasma did not etch the lithium niobate or the sapphire, so that the lithium niobate thickness could be measured by taking Talystep™ measurements of the total step and subtracting the (previously measured) silicon thickness. (SOS and ion beam milling will be discussed in Chapter 5). Profile measurements are shown in figure 4.10, showing surface roughness to be of the order of 5%, in agreement with the optical data.

![Graph showing Talystep measurements](image)

**Figure 4.10.** Talystep measurements showing surface roughness.

Optical and Talystep™ thickness measurements were used to calculate deposition rates. The deposition rate was influenced primarily by RF power. The effects of other parameters, such as deposition chamber pressure, were largely masked by variations which were chiefly ascribed to deposition pressure instabilities: these led to changes in RF matching and hence to the RF power density applied to the plasma. Deposition rates are plotted as a function of RF power density in figure 4.11.
Figure 4.11. Deposition rate versus RF power, as measured by Talystep (+) and optical waveguiding (Δ).

4.4.4. Electrical properties

The film resistivity was measured using parallel-plate capacitors. High resistivity and breakdown voltage were essential in order to be able to apply an electric field across the layer for optical modulation, as explained in § 4.1.

Lithium niobate was deposited on thin films of gold and of silicon, and aluminium capacitor dots were subsequently deposited on top. Many of these devices were short-circuit, with very low resistance: a few were open circuit. This behaviour suggested localised defects such as pinholes rather than bulk conductivity. SEM examination revealed a number of features which may indeed have been pinholes.

There was some correlation between the film thickness and the incidence of short circuits, with the thicker films showing fewer failures. It was therefore decided to fabricate devices with rather thicker layers, (c. 0.2μm, rather than 0.6μm). It was hoped that this, possibly coupled with the use of silicon nitride buffer layers, would give a reasonable chance of fabricating several devices free of short circuits.

In a continuation of this experiment, capacitors were fabricated alongside test modulators on device substrates. In one type of capacitor, the upper plate was provided by the top silicon layer; in another type, the top silicon layer was removed and the upper plate was provided by the metal layer. It was found that the metal-electrode
type were electrically either short-circuit or, at best, rather fragile, with breakdown fields usually about 9V/µm. The silicon-electrode type were mostly free of short circuits, and were tested up to 90V/µm. The high resistivity of the silicon upper electrode no doubt conferred some degree of protection, by limiting avalanche effects. The large numbers of initially short-circuit capacitors, as well as the lower breakdown fields, implied that metallisation directly onto the lithium niobate layer was detrimental to its dielectric integrity.

Test capacitors

![Diagram of test capacitors](image)

**Figure 4.12. Comparison of step coverage lengths and areas of test capacitors and modulators**

Leakage currents measurements were taken on capacitors using silicon electrodes, i.e. structures similar to the devices, yielding resistivities of typically $10^{10} \Omega$–cm, which provided a margin of three orders of magnitude over the minimum value specified in § 4.1. On the other hand, measurements on completed modulators exhibited higher leakage currents than this figure suggests. It was inferred that the leakage current on the modulators was associated with a step coverage problem. The design of the modulators included a great length of step coverage, where the lithium niobate and upper silicon climbed up onto the lower silicon electrode. The design of the test capacitors was such as to minimise the length of this step, as shown in figure 4.12. The figure presents a comparison between typical dimensions and leakage currents for the two structures, showing that leakage was correlated with step length rather than area. This implied that the resistivity calculated from the capacitor measurements represented only a lower limit on the film resistivity, but that the
performance of devices might be limited instead by step coverage effects unless an alternative layout was devised to minimise the step length.

In most cases, leakage currents were observed to behave asymmetrically, with the positive current greater than the negative (the positive polarity being defined as flowing from the upper contact to the lower). The low-frequency positive leakage currents were observed to drift, taking up to twenty seconds to respond to a change in the DC bias voltage. This behaviour is illustrated by the photographed oscilloscope trace in figure 4.13, which is the IV characteristic of a 1mm² dot of gold-on-chrome, deposited on a 1500Å lithium niobate film over a gold-on-chrome ground-plane. The photograph was exposed for 1 second, during which the trace drifted upwards, appearing as a broadened trace: there followed a ten second pause, and then another 1 second exposure. The second trace is not broadened, indicating that the leakage characteristic had reached an equilibrium.

![Photograph:](Image)

**Figure 4.13.** I-V characteristic of a 1mm² thin film capacitor (gold/chrome electrodes, 1500Å lithium niobate film).

Device leakage currents were compared to measurements of test resistors. In most cases, currents were substantially smaller than would have been expected if the lithium niobate was consistently breaking down along the steps. A typical result, on layer P43 (2300Å thick) showed device leakage currents of 150 to 380nA at 10V, (over a 3000µm device length) compared to 900nA calculated for fully short-circuited structures (i.e. with the entire voltage drop being applied across the silicon on either side of the overlap region). The main conclusion from these experiments was that thick (0.2µm) films might be expected to yield several operable devices, although the leakage...
current would lead to a potential difference across the silicon and a consequent reduction in the field applied to the lithium niobate.

### 4.4.5. Pinholes: etching experiment

An experiment was performed to ascertain the incidence of pinholes in the deposited lithium niobate layers. In this experiment, lithium niobate layers were first deposited on device substrates (patterned SOS). The specimens were then subjected to an SF$_6$ plasma for five minutes under conditions which would have given a silicon etch rate of 0.6μm/minute. Microscopic examination revealed that the underlying silicon had not been etched, implying that the lithium niobate layer offered good protection and was therefore free of pinholes.

As an adjunct to this experiment, a partial ion beam milling step was performed between lithium niobate deposition and plasma etching. Half of the surface of each specimen was covered with resist, and the specimens were ion beam milled for a duration calculated to remove approximately half the thickness of the lithium niobate. The resist was then stripped and the experiment continued with the SF$_6$ plasma process. Microscopic examination showed that, in regions which had been partially ion beam milled, the underlying silicon had been eroded, as can be seen in the micrographs in figure 4.14. The erosion was chiefly observed at the edges in the silicon pattern, implying a lithium niobate step coverage fragility. This may have been enhanced, or even caused, by electrostatic damage in the SF$_6$ plasma.

![Figure 4.14. Optical micrographs of SF$_6$ plasma-etched specimens of silicon with a covering of sputtered lithium niobate. The specimen in the left hand photograph was partially ion beam milled prior to plasma etching: erosion of the silicon is visible as a blue discolouration, predominantly at the edges of the pattern.](image-url)
The important implication for device processing was that the sputtering process yielded essentially pinhole-free layers.

4.4.6. X-ray crystallography

X-ray diffraction was used, to probe the crystalline structure and orientation of the films. The geometry of X-ray reflections is well documented[2]; Appendix 4.B gives a brief outline of important results.

Rhombohedral lithium niobate belongs to space group R3c, with six formula units per unit cell and with lattice parameters a and c of 5.148 and 13.864Å respectively[9]. The X-ray pattern of lithium niobate is determined almost entirely by the niobium atoms, since niobium has a much higher atomic number (Z = 41), and hence a higher X-ray scattering factor, than oxygen (Z = 8) or lithium (Z = 3). It follows that the poling state of the material cannot easily be determined using X-rays.

A set of Laue, oscillation, rotation and powder photographs were taken using commercially grown single crystal material, for reference purposes (see Appendix 4.C). X-ray patterns from grown films were then compared directly to these to determine whether the patterns were consistent with the growth of rhombohedral lithium niobate. The X-ray structure of the grown films depended critically on the nature of the substrate as well as the processing conditions.

It was found that layers grown at high RF power on amorphous substrates and on z-cut lithium niobate tended to exhibit a preferred c-orientation normal to the plane. This was examined by taking 5° oscillation photographs, with the oscillation axis in the plane of the specimen, centred on the position expected for the <006> reflection. Figures 4.15(a) to (d) show typical results, of films grown on fused silica, on LPCVD silicon nitride (on top of thermally oxidised silicon), on patterned silicon-on-sapphire (SOS) with a 200Å PECVD silicon nitride buffer layer, and on a z-cut lithium niobate substrate respectively. The white radiation streaks in figures 4.15(b) to (d) are caused by the underlying substrates. The patterns due to the deposited films appear as arcs of constant θB on these photographs: these may be viewed as sections of powder lines, the intensity profile along the arc indicating the distribution of crystallites about the average orientation. In figures 4.15(a) to (c), the misorientation is of the order of 10°; the films grown on z-cut lithium niobate, on the other hand, showed much more definite orientation, as seen in figure 4.15(d).
Key to photographs: identification of some of the major features of rhombohedral lithium niobate visible in the photographs (a) to (c). (See reference photographs and note on indexing, Appendix 4C).

(a) on fused silica

(b) on (amorphous) LPCVD silicon nitride on oxidised silicon.

Note white radiation streaks due to substrate silicon

(c) on (amorphous) PEVD silicon nitride on SOS

Figure 4.15. X-ray diffraction photographs of sputtered lithium niobate films (continued overleaf).
Figure 4.15 (continued). X-ray diffraction photographs of sputtered lithium niobate films.

Note that the photograph in figure 4.15(d) was taken with the orthorhombic y-axis of the substrate set along the oscillation axis, so that the pattern due to the substrate does not exhibit mirror symmetry: the spots due to the thin film, on the other hand, are symmetrically disposed about the centre-line, indicating that some degree of twinning has taken place within the layer.

The transition temperature, between amorphous and crystalline growth, was determined to lie between 315°C and 350°C for layers grown on silicon nitride. Above 370°C, the intensity of the central region of the X-ray arcs were fairly constant.

RF power density was important in determining the structure of films on fused silica substrates. Below 1.3W/cm², films appeared to be amorphous; above this, the films were found to be crystalline. Films on silicon nitride, however, were crystalline with RF powers even as low as 0.5W/cm².

Deposition pressure was varied between 5 and 25mTorr, but exerted only slight influence on the film structure as determined by X-rays. 20mTorr was adopted for device depositions, as this made the glow discharge easier to stabilise and yielded marginally stronger <006> preferred orientation.

DC Bias was varied from 10V to 50V, for RF power densities of 0.5 and 1.3W/cm² while maintaining the pressure at 20mTorr. Below 20V bias, crystallinity was weaker or absent, whereas no change was discerned above about 30V.

Films deposited at high RF power densities (2.6W/cm²) exhibited some additional X-ray features, illustrated in figure 4.16. The clearest of these was an arc, similar to the <006> reflection, displaced from it by about 1° (Bragg). This did not correspond with any line in the rhombohedral lithium niobate pattern, and may
therefore indicate the presence of a separate phase. For device applications, it was decided to deposit films at lower RF power densities (1.3W/cm²) where the X-ray data indicated that a single homogeneous phase could be obtained.

Substrate: amorphous LPCVD silicon nitride on oxidised silicon.

Figure 4.16. X-ray diffraction photograph of a lithium niobate film sputtered at high RF power. The <006> feature has acquired a companion; other additional features are visible. (cf. figure 4.15(b)).

Films grown on patterned SOS with PECVD silicon nitride buffer layers have been discussed above. Films were also grown on patterned SOS without such buffer layers. SOS consists of a layer of epitaxial <100> silicon on single crystal sapphire. The sapphire orientation is close to <012>, but is a few degrees off so that the surface has atomic steps every hundred atoms or so which are believed to help initiate epitaxial growth of the silicon[47]. Films on these substrates exhibited two growth structures.

On exposed <012> sapphire regions, growth was epitaxial: the X-ray patterns were consistent with y-oriented lithium niobate, with <100> twinning planes[48]. These were examined and identified by taking inclined beam 180° oscillation and equi-inclination Weissenburg photographs[2], such as those shown in figure 4.17.
(a) film grown on \( <012> \) sapphire substrate, oscillation about the normal to the plane.

(b) reference \( \text{LiNbO}_3 \) specimen, oscillation about \( <010> \).

Figure 4.17 (a,b). Inclined beam oscillation photographs showing spots characteristic of \( <010> \) oriented lithium niobate. Inclination: \( 5^\circ \); oscillation: \( 204^\circ \).
Figure 4.17 (c,d). Equi-inclination Weissenberg X-ray diffraction photographs showing spots characteristic of $<010>$ oriented lithium niobate. Both were set with $\mu = 4^\circ40'$, to obtain the first layer visible in the photographs 4.17 (a) & (b).
The other growth structure observed on these substrates was a preferred <012> orientation. This is discernible in figure 4.18, which is a long exposure 15° oscillation photograph where the X-ray beam was directed onto a region where there was a substantial coverage of silicon (c. 40%): lithium niobate <012> and <024> spots are visible. (The individual silicon islands were below the in-plane resolution of the X-ray beam, so that an archipelago was studied). When aiming the X-rays at a portion of the substrate devoid of silicon, these features were not apparent. These results suggested that lithium niobate had grown with a preferred <012> orientation on patterned silicon on sapphire, i.e. on device regions.

![Figure 4.18](image)

Figure 4.18. 15° oscillation photograph of a sputtered film of lithium niobate on SOS. lithium niobate <012> and <024> spots are clearly visible.

Layers were grown on unpatterned SOS, i.e. the same situation as described above, except that the silicon coverage was continuous: these films showed no significant X-ray patterns at all.

The absence of crystalline growth, at low deposition rates on silica and on large expanses of silicon, may be accounted for by the diffusion of lithium into the substrate. Lithium is highly mobile: the diffusion length in crystalline silicon, over the thermal process outlined here (370°C for over an hour) is of the order of 1mm\(^4\). According to this hypothesis, the disruption of crystalline growth is due to a breakdown of stoichiometry: at high RF power (and growth rate), the supply of lithium is fast enough to ensure that removal of lithium from the growing layer becomes diffusion-limited and only affects a thin layer at the bottom of the growing film. The problem might not be observed when growing on silicon nitride because it is an excellent diffusion barrier\(^5\).
Further evidence in support of this explanation was provided by the observation of crystalline growth in a very thick (0.77\(\mu\)m) film grown on fused silica, using comparatively low power (1W/cm\(^2\)). According to the lithium diffusion hypothesis, the removal of lithium becomes diffusion-limited because of the increased quantity of lithium as well as the greater distance between the growing surface and the unsaturated substrate.

![Photomicrograph of silicon and silica patterns on sapphire, covered by a sputtered film of lithium niobate. The regions surrounding the silicon patterns are discoloured, giving the appearance of haloes. These indicate thickness or refractive index variations in the lithium niobate film. The effect is strongest in the vicinity of silica patterns.](image)

The same hypothesis may also account for the lack of crystalline growth on large areas of silicon. If the silicon provides a sink for lithium, large areas of silicon might be expected to cause significant lithium-depletion and disruption of the crystal structure. Where the silicon only occurs in small islands, the effect would be reduced. Microscopic examination revealed that silicon patterns had a visible effect on the surrounding lithium niobate, up to a distance of tens of microns. This is visible in figure 4.19, where each silicon island is surrounded by a faint but sharply defined halo. The change in colour in the halo may be due to a change in refractive index due to the altered composition.

This effect is a potential drawback for the silicon dioxide buffer process described in Chapter 5. Microscopic examination showed that the halo appeared to be enhanced in the presence of silicon dioxide buffers.
4.4.7. Compositional analysis

To some extent, X-ray data may be taken to give an indication of composition: if the material is far off stoichiometry, it is not likely to form the desired structure. Conversely, observation of the LiNbO$_3$ X-ray pattern indicates that the composition is close to stoichiometric, or that it contained an appreciable fraction of material of the correct structure. Having said this, it should be borne in mind that it is possible for the composition to alter, e.g. by diffusion of mobile species or by implantation, without destroying the crystal structure$^{[51]}$. Strictly speaking, these inferences from X-ray data are therefore only valid in the growing film.

Auger spectroscopy$^{[52]}$ was performed in an attempt to relate changes in crystallinity to compositional variations. The technique is relatively insensitive to lithium, because the lithium signal is partially superimposed on a niobium signal and an instrumental artifact, as may be seen in figure 4.20. In spite of this, measurements were repeatable to within about 1%.

![Auger spectrum of a sputtered lithium niobate film, illustrating the Nb, Li, and O features and the instrumental artifact in the low-energy region. Carbon and silicon peaks are also visible.](image)

Figure 4.20. Auger spectrum of a sputtered lithium niobate film, illustrating the Nb, Li, and O features and the instrumental artifact in the low-energy region. Carbon and silicon peaks are also visible.
Growth runs were performed with periodically decremented bias, (over the range 50 to 20V in 5V steps) and the oxygen concentration was then measured as a function of depth. A profile is shown in figure 4.21, and it is clear that the variations, if present, are not much larger than the random fluctuations which are of the order of 1%. Contrasting this with the measured change in crystallinity by X-rays over the same range (see § 4.4.6), it seems that the Auger stoichiometry does not provide a sensitive indication of a film property of importance to its structure (and hence to the growth of device-quality films).

![Image of Auger depth profile](image)

Figure 4.21. Auger depth profile of atomic species in lithium niobate deposited upon silicon nitride substrate with a stepped DC bias.

Although this does not contradict the hypothesis that the X-ray data is a sensitive indication of deviation from stoichiometry in these films, the important caveat remains that the observed dependence of film structure on bias may not be due to subtle changes in stoichiometry, but to some other mechanism, such as an enhancement of surface temperature and mobility enhancement due to ion bombardment.

**4.4.8. Scanning Electron Microscopy (SEM)**

Layers were examined in a SEM. In most cases, the surface was almost featureless. In the case of layer P14 (on fused silica), which was a thick layer (c. 0.77µm), columnar grains became visible, with a lateral dimension of about 0.1µm: this is reproduced in figure 4.22. The exposed surfaces of these grains were generally rounded, giving the film a rough appearance. This is in agreement with data in the
literature\cite{26,34} as well as the observation of poor orientation in the X-ray pattern as the rough surface effectively presents a wide range of surface orientations in the later stages of growth.

![Scanning electron micrograph of a 0.77μm thick layer of lithium niobate on silica.](image)

**Figure 4.22.** Scanning electron micrograph of a 0.77μm thick layer of lithium niobate on silica.

### 4.5. Summary

Modifications to the sputtering apparatus were successful in achieving the desired range and accuracy of process control for these experiments. A novel scheme was devised for applying DC bias to the substrates: growth studies confirmed that this offered a useful dimension of process control.

Lithium niobate thin films were grown and characterised. Crucial film parameters were measured, viz. crystal structure and orientation, electrical leakage and breakdown strength, and optical waveguiding. The growth process was optimised to grow films which met almost all the device requirements which were outlined at the beginning of this chapter. The exceptions to this were the poling state and electro-optic coefficients of the material; these measurements will be discussed in Chapter 6.

The operating conditions selected for growing device films were as follows:

- **Power:** 1.3 ±0.05W/cm²
- **Temperature:** 370 ±10°C
- **Pressure:** 20 ±5mTorr
- **DC bias:** 40 ±10V
Under these conditions, the growth rate was 200 ±10 Å/hour, and the resulting films on silicon islands appeared under X-ray analysis to exhibit the desired LiNbO$_3$ structure, albeit with a preferred <012> orientation instead of the <006> orientation which was sought. As was mentioned in § 4.2, and will be discussed in Chapter 6, this material exhibited an electro-optic effect via the $r_{51}$ coefficient.

References


51. C.W. Pitt, private communication.

Chapter 5: Materials & process development

Abstract

This chapter is concerned with fabrication of the modulators which were described in Chapter 3. The development of a fabrication sequence was essentially a matter of putting together a series of standard processes, for depositing and patterning the various layers in turn - silicon, lithium niobate, silicon, then metal. In the cases of silicon and metal, the techniques were carried over from semiconductor device technology[1].

Deposition of electro-optic lithium niobate was of crucial importance to the success of the project: this was discussed in Chapter 4. Patterning of the lithium niobate was achieved by ion beam milling[2]. Both deposition and patterning of the lithium niobate introduced problems when combined with the silicon processes. The dependence of the lithium niobate crystal structure on the underlying substrate was discussed in Chapter 4; lithium niobate deposition was found to cause delamination of certain silicon layers; reactions were found to occur, affecting the properties of lithium niobate in the vicinity of the lower-electrode silicon patterns. Patterning of the lithium niobate layer had to be carried out in such a way as to avoid etching away the lower silicon electrode.

This chapter is presented in three sections. To begin with, the materials and processes are each discussed in isolation. In the second section, the complete fabrication sequence is described. Finally, there is a brief discussion of mask design.

5.1. Materials & processes

5.1.1. The silicon layers

One advantage of using silicon for the electrodes was that it was a material which had already been extensively studied: standard processes were routinely available for processing and deposition, and electrical characteristics were well documented[1,3].

Silicon layers may be grown or procured by a variety of techniques[4-11]. In this work, layers were procured by two methods: one was by plasma enhanced
chemical vapour deposition (PECVD)[5]; the alternative was to obtain wafers of single-crystal silicon-on-sapphire (SOS)[8].

This choice of materials was made largely for reasons of convenience, rather than for optimum device performance. This is justified in terms of the ultimate aim of the project, which was to demonstrate the operation of a device rather than to optimise it. It is known, for instance, that waveguides of SOS material has a tendency to exhibit relatively high optical propagation losses[12], attributed to interfacial stresses: it was expected, and experimentally confirmed, however, that with such short devices (c. 400µm), this would not be catastrophic. Little was known about the optical properties of doped hydrogenated amorphous silicon above the band gap, beyond a few studies of infra-red waveguiding in a rather different regime[13,14]: waveguiding measurements were used here also, to confirm that the material was not disastrously lossy. The most severe problem lay in the very high resistivity of amorphous silicon[15]. This limited the bandwidth of the devices in this study seven orders of magnitude below the bandwidths predicted in Chapter 3. Although a severe limitation, this was not expected to prevent the devices from functioning, and the comparative ease of procurement was deemed to outweigh the loss in performance.

5.1.1.1 PECVD hydrogenated amorphous silicon

Plasma Enhanced Chemical Vapour Deposition is a process which is widely used in the fabrication of silicon solar cells, and which has therefore been extensively studied[5,16]. The substrates are heated to about 180°C in an evacuated chamber. A low partial pressure (c. 150mTorr) of silane is introduced into the chamber, whereupon a plasma is struck by applying RF power. Typical process conditions are 10W RF (13.56 MHz) applied via a 16" dia. electrode suspended a couple of inches above the susceptor. The deposited layer may be boron-doped in situ by introducing a partial pressure of diborane into the discharge[15]. Deposition rates are of the order of 60Å per minute. The depositions for this project were carried out by BP research (Sunbury-On-Thames) and by the Department of Physics at Dundee University.
Conduction band

Fermi energy

Density of states

Valence band

Energy level diagram showing the density of states and (shaded) occupied states in an intrinsic semiconductor.

(a) interband transition: $h\nu > E_g$ dominates absorption in this regime (short wavelength)

(b) bandgap states due to impurities: $h\nu$ may be less than $E_g$

(c) bandgap "tailing": $h\nu$ may be less than $E_g$; very important for infra-red absorption in amorphous silicon

(d) free carrier absorption in doped semiconductors (n-type shown); $h\nu$ may be less than $E_g$

(e) photo-induced free carrier absorption primary photon causes interband transition (see (a)), secondary photons are absorbed by the excited carriers as in (d); secondary photon $h\nu$ may be less than $E_g$

Figure 5.1. Optical absorption processes in semiconductors
Although the electronic function of these layers in the present devices was merely to act as a conductive layer, the band gap was important for infra-red transparency. Single crystal silicon exhibits an absorption edge, at the wavelength corresponding to the band gap energy (c. 1.2μm): above this wavelength, residual absorption is due mainly to free carrier absorption by electrons in the conduction band or by holes in the valence band\cite{17}. Any residual energy levels in the band gap, however, allow bound carriers to absorb photons as well, as illustrated in figure 5.1.

The silicon layers obtained by PECVD are amorphous, as the growth temperature is too low for crystallisation. The material is also hydrogenated, by incorporating some of the hydrogen which is a by-product of the decomposition of silane. This hydrogenation is essential to the electronic properties of the material\cite{18}. The hydrogen atoms become attached to the so-called "dangling bonds" which are formed in amorphous silicon where an atom is bonded to fewer than four neighbours. The "dangling bonds" introduce electronic energy states in the band gap: hydrogenation effectively removes these states, and the band gap is at least partially recovered.

The hydrogen in these layers is not very strongly bonded, and is rapidly liberated above 410°C\cite{19}, so that any processes above this temperature is likely to cause the material to become strongly absorbing in the infra-red. This behaviour placed an upper limit on the subsequent processing temperatures of these devices.

Although the electrical and optical properties of this material are qualitatively similar to those of crystalline silicon, the quantitative behaviour is markedly different\cite{15,16}. The activation of dopant atoms depends critically on their bonding states and is therefore much less efficient in amorphous than in crystalline silicon; defect states provide carrier trapping and recombination sites; the irregular atomic distribution promotes electron-phonon scattering; these effects all tend to increase the resistivity of the material. Minimum values of the resistivity are of the order of 100 Ω-cm or more, i.e. over four orders of magnitude higher than in crystalline silicon. This yielded a minimum sheet resistance (dark) of 5 MΩ/sq for the top electrode (thickness 0.2μm): in practice, the top electrode material was doped using 10⁴ ppm boron in the gas phase, and sheet resistances were measured at 500MΩ/sq. Such high resistivity severely restricted the bandwidth of these devices, to a few hundred Hertz: however, this was deemed sufficient to demonstrate device feasibility if not the ultimate performance of the class of structures.

It was found that layers of PECVD silicon on lithium niobate substrates were apt to delaminate if heated beyond 300°C, even using ramp rates as low as 1°C per
minute. This was understood to be due to the very high thermal expansion coefficients of lithium niobate\(^2\) compared to that of silicon\(^2\). As the substrate expanded, the thin silicon layer was stretched. The tensile stress resulted in shear stress at the interface: eventually, the shear stress overcame adhesive forces and the layer peeled away.

This problem was investigated by placing specimens of lithium niobate with patterned layers of PECVD amorphous silicon in a furnace and subjecting them to various thermal cycles. It was noticed that on patterned material, delamination was correlated with certain geometrical features: this is visible on the micrograph in figure 5.2. The large area of silicon is peppered with delamination bubbles, whereas the 83\(\mu m\) wide stripe is much clearer. The smaller features, 12\(\mu m\) squares, are completely clear. This behaviour was exploited in the mask design: features in the lower electrode layer were broken down into small islands and stripes. Unfortunately, the batch of material which was used for devices turned out to be of rather inferior quality, and thermally delaminated even after being etched into small islands.
Delamination was found to be more severe with some layers than with others: in some cases, it was present after merely heating to 120°C for dehydration before the photolith process. This variation from batch to batch (of nominally identical layers) implied that pre-deposition cleaning might be a key factor.

Thermal delamination prevented fabrication of devices using PECVD amorphous silicon as a lower electrode for the present project. Having said this, the investigations described above demonstrated that the problem could be overcome by a combination of careful cleaning and patterning.

5.1.1.2. Silicon-On-Sapphire (SOS)

SOS is obtained by growing <100> silicon on <012> sapphire by low pressure chemical vapour deposition[7,8]. Although it is more expensive, and often of lower quality, than bulk or homoepitaxial silicon, SOS provides certain advantages which are indispensable to certain applications[22,23]. The individual components within a circuit are electrically insulated from one another by patterning the silicon layer into small islands, rather than by forming p- and n-wells as in bulk material. As a result, the substrate is not prone to electronic damage and consequent short circuits. In addition to this, the circuits can be driven at higher frequencies since the capacitance of an electrically conductive substrate has been removed. SOS is therefore useful where high speed and radiation-hardness are important.

In the present project, SOS offered commercially grown lower-electrode material, on an infra-red-transparent substrate with a suitably low refractive index[24].

The starting material was initially in the form of 3" diameter wafers; the silicon layers were 0.3μm thick, undoped. These were doped by boron ion implantation[1]; the implant was driven in and annealed, by placing the wafers in a furnace at 950°C in nitrogen. After 10 minutes, the nitrogen atmosphere was replaced by wet oxygen (80% water, 20% oxygen) in order to grow an oxide layer[1]. This was timed to leave just 0.15μm of silicon unoxidised. Finally, the oxide was removed by dipping in a buffered oxide etch solution (7:1 ammonium fluoride:HF) until the wafers became hydrophobic[1]. The nitrogen anneal prior to oxidation was necessary to drive the boron away from the surface, so as to avoid losing much of the boron by segregation at the silicon/oxide interface[1].
This method of thinning the silicon layer, by oxidation, was chosen in preference to a silicon etch, because it is controlled by a very well characterised and uniform process, viz. thermal oxidation\[1\], yielding a very smooth surface finish. A target thickness of 0.15µm ±10% was specified: the layers were measured by Talystep\textsuperscript{TM} after processing. The result was 0.165µm; variations across the entire batch were below the resolution of the measurement (c. 0.005µm).

The implant dose was chosen after processing two test wafers with doses of \(10^{14}\) and \(10^{15}\) ions/cm\(^2\). These were probed to measure sheet resistance, and infra-red guiding experiments were performed. The guiding experiments indicated that even the higher dose did not obliterate the optical waveguiding, so that the device material was implanted with the higher of these two doses, resulting in a sheet resistance of 500Ω/square.

5.1.2. The lithium niobate layer

The lithium niobate layer was deposited by RF sputtering. This part of the programme was discussed in detail in Chapter 4: the growth of <006> and <012> oriented lithium niobate films on SOS was achieved, and involved heating the substrates up to 370°C for a few hours. The following section describes the work which was aimed at defining a pattern on the lithium niobate, etching away areas so as to uncover the lower electrode and enable electrical contacts to be made.

Although several formulations have been found to etch lithium niobate\[25\], the etch rates are all very slow: moreover, these processes etch silicon faster than lithium niobate, so that there was no process with a selectivity high enough to provide a reliable end stop. A carefully timed ion beam milling process\[2\] was used, stopping a short distance into the underlying silicon. The ion beam milling apparatus and process are described in Appendix 5A.

An experiment was performed to measure the etch rates of lithium niobate and of electrode materials. Specimens of patterned PECVD silicon on z-cut lithium niobate and SOS were patterned with photoresist. Y-cut lithium niobate specimens and RF sputtered lithium niobate films were also prepared, for comparison. These were then ion beam milled, using argon ions at energies between 500 and 1000eV and an arc current of 0.5A. The etch rates are plotted in figure 5.3.
Figure 5.3. Argon ion beam milling etch rates versus beam energy. The data points represent measurements; the straight lines are estimates of best fit, indicating the trends for each material. The beam current density $\rho$ was maximised at each energy, and is shown by the graph on the right.

The photoresist in certain areas was found to bubble up upon exposure to the beam. Photoresist over silicon features was most seriously affected: at high beam energies, the bubbles popped open, allowing the underlying silicon to be milled. Surface profile measurements indicated that the bubbles on a particular sample would burst almost simultaneously, after a delay related to the beam energy. Samples milled at the lower energies (600eV and below) formed bubbles but these did not burst at all.

Uniformity was monitored by measuring etch rates at 1mm intervals across various samples. Uniformity across each specimen was better than 5%, although run-to-run variations were about 10% because of the difficulty of reproducing exactly the chamber and beam conditions.

Using these results, it was decided to process the devices at 500V, at which the optimised beam current density was $0.31\text{mA/cm}^2$, and the etch rates of lithium niobate and silicon were 120 and 140Å/minute respectively.

Devices were milled in several iterations, approaching the lithium niobate/silicon interface in prudent steps which were calculated using the measured uncertainties in the lithium niobate thickness and in the etch rates. At 400Å, the lithium niobate exhibited a distinctive milky appearance under a microscope, and a "bullseye" of residual material.
was often visible to the naked eye; down to about 50Å, residual lithium niobate could be detected by examining the subtle colour differences, under a microscope, between substrate areas which had been masked by the clip during deposition, and those which were not. The milling process was repeated until a point estimated to be about 50Å beyond the interface in the bullseye region, ensuring that the lower silicon was properly exposed for metallisation but was nowhere etched so severely as to detract from its electrical performance.

5.1.3. Oxide buffer layers

The uncertainties in the ion beam milling of the lithium niobate layer introduced difficulties in process control. This was not disastrous: nevertheless, a variation on the standard process was devised to overcome the problem. The process sequence involving this variation is shown in figure 5.4. A thick sacrificial layer of silicon dioxide was deposited over the first silicon layer. This oxide was then patterned in such a way as to expose the silicon in the active regions, while protecting it wherever the lithium niobate was to be removed. The protective oxide was stripped off all exposed areas by dipping the devices in HF immediately prior to metallisation. Although this technique was slightly more complicated than the standard process sequence, it was felt that it could provide greater process latitude which might be indispensible in processing larger substrates where uniformity might be more difficult to control.

5.2. Device fabrication

Device fabrication comprises two phases. The first is concerned with the process sequence as far as metallisation, using the thin film processes (deposition, patterning etc.) described above. The second phase is concerned with final preparations for device testing: dicing, chip mounting, and wire bonding.

5.2.1. The process sequences

The process experiments, described above, enabled two fabrication sequences to be put together. These are laid out in Table 5.1 and shown schematically in figures 5.4 (standard) and 5.5 (with oxide buffer).
## Table 5.1. The parallel-plate modulator fabrication sequences.

<table>
<thead>
<tr>
<th>Process step</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Deposit silicon (lower electrode)</td>
<td>0.15μm boron-doped PECVD α-Si:H or SOS. See § 5.1.</td>
</tr>
<tr>
<td>2. Photolith &amp; etch silicon (lower electrode)</td>
<td>Plasma etching, using 100 to 150 mTorr SF₆ and 100W RF²⁶. Etch rates were about 0.3μm per minute, giving etch times in the region of 30 seconds. The purpose was to remove silicon in places which were later to be covered by the top silicon layer, except in the active region. Thus the capacitance was reduced, as was the likelihood of short-circuits. Edges and ends of rib waveguides, and test features such as resistors and capacitor plates were defined. Patterning also alleviated stress and thermal delamination (see § 5.1.1.1).</td>
</tr>
<tr>
<td>3. Deposit silicon dioxide (oxide buffer process only)</td>
<td>0.5μm, deposited by evaporation²⁹. The purpose was to protect the lower electrodes against the effects of over-etching the lithium niobate and upper silicon layers.</td>
</tr>
<tr>
<td>Step</td>
<td>Process Description</td>
</tr>
<tr>
<td>--------</td>
<td>---------------------</td>
</tr>
<tr>
<td>4.</td>
<td>Photolith &amp; etch silicon dioxide (oxide buffer process only) This photolith was complicated by the transparency and low surface reflectivity of the devices, enabling the exposure beam to scatter off the aligner chuck, exposing masked photoresist from below. To counter the problem, samples were back-coated with matt black paint prior to resist coating. Etching was performed using a 1% HF solution, which did not etch the underlying silicon(^1). The purpose was to remove silicon dioxide from the active regions of devices and test structures.</td>
</tr>
<tr>
<td>5.</td>
<td>Deposit lithium niobate Lithium niobate, for the electro-optic layer, was deposited at 370°C by RF sputtering: see Chapter 4. In cases where the oxide buffer was incorporated, the lithium niobate could be patterned and etched at this stage: see Step 8.</td>
</tr>
<tr>
<td>6.</td>
<td>Deposit silicon (upper electrode) 0.2µm PECVD boron-doped α-Si:H. See above, § 5.1.</td>
</tr>
<tr>
<td>Step</td>
<td>Process</td>
</tr>
<tr>
<td>------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>7.</td>
<td>Photolith &amp; etch silicon</td>
</tr>
<tr>
<td>8.</td>
<td>Photolith &amp; etch lithium niobate</td>
</tr>
<tr>
<td>9.</td>
<td>Deposit metal</td>
</tr>
<tr>
<td>10.</td>
<td>Photolith &amp; etch metal</td>
</tr>
</tbody>
</table>
1. The first silicon layer is deposited and patterned:

   Electrode

   Substrate

2. The lithium niobate layer is deposited.

3. The top silicon electrode is deposited and patterned. Note how the lower electrode is totally protected from the etching process by lithium niobate.

4. The lithium niobate layer is ion beam milled to expose the lower silicon layer. Note the erosion of the lower silicon layer and substrate due to over-etching.

5. Metal contacts are deposited and patterned.

---

Figure 5.4. The parallel plate modulator fabrication process:
(a) omitting oxide buffer layers.
1. The first silicon layer is deposited and patterned:

```
Substrate
Electrode (silicon)
```

2. A layer of silicon dioxide is deposited and patterned. Note how the highly selective etch does not affect the silicon.

```
Substrate
Electrode (silicon)
Silicon dioxide
```

3. The lithium niobate layer is deposited and patterned. Note the erosion of silicon dioxide and substrate by substantial over-etch, while the silicon is totally protected.

```
Substrate
Electrode (silicon)
Silicon dioxide
Lithium niobate
```

4. The upper silicon electrode is deposited and patterned. Note further erosion of silicon dioxide by the silicon etching process, while the lower electrode remains totally protected.

```
Substrate
Electrode (silicon)
Silicon dioxide
Lithium niobate
Upper electrode (silicon)
```

5. The remaining silicon dioxide is stripped off to expose the lower silicon layer

```
Substrate
Electrode (silicon)
Silicon dioxide
```

6. Metal contacts are deposited and patterned.

```
Substrate
Contact - lower electrode
Region of lateral optical confinement
Contact - top electrode
```

Figure 5.5. The parallel-plate modulator fabrication process: (b) using silicon dioxide buffers.
5.2.2. Final preparations for optical testing

Specimens were either sawn and polished or cleaved so as to present clean waveguide ends for optical end-firing.

For lithium niobate and silica substrates, this was a straightforward task of cutting and edge-polishing. Devices were wax-mounted between glass slides, and cut using a diamond saw. They were then polished by grinding with successively finer grades of carborundum, and finally syton (a proprietary preparation of silica particles in an aqueous suspension)[28]. After polishing, the devices were cleaned using trichloroethylene, acetone and iso-propyl alchohol.

Sapphire substrates were found to be too hard for this polishing method. SOS-based devices were prepared by the more expedient method of scribing and cleaving. The devices were scored across the back using a diamond scriber, and then snapped in two. The cleaved edge was not easy to align precisely, as it had to be done by the naked eye through the back of the device. This improved with practice, and a number of adequate devices were obtained by this method. A scanning electron micrograph of a cleaved modulator is shown in figure 5.6.

![Figure 5.6. Scanning electron micrograph and schematic diagram of the cleaved end of a completed modulator.](image-url)
The completed specimens were finally mounted onto chip carriers, and selected devices were connected to output pins by ultra-sonic wire bonding\cite{29}. A photomicrograph of a finished device is shown in figure 5.7.

5.3. Device design: Photomasks

A mask set, named Merlin, was designed for use with the fabrication sequence described above. This section is concerned with the layout of the modulators and test features incorporated in Merlin.

It was decided to keep the modulator designs as simple as possible, to avoid compounding the difficulties already inherent in fabrication and in the optical waveguiding experiments. In particular, the modulators were all designed to be straight waveguides rather than Mach-Zehnder arrangements, so as to avoid the electrical complications which waveguide junctions would bring about, as well as simplifying the optical structure. A typical modulator layout is shown schematically in figure 5.8.
Notes.
Bond pads are not drawn to scale. In devices, they were at 200μm square (minimum), to ensure ease of wire bonding.
The lithium niobate layer was not patterned until after etching the upper silicon layer (see Table 5.1). In completed devices, the lithium niobate layer therefore extended wherever upper silicon was present, in addition to the masked area shown here.

Figure 5.8. Merlin: modulator layout

Lateral confinement in the waveguides was provided by the rib topology, formed by the overlap region in the silicon patterns. The optical mode was launched into the overlap region, and was laterally confined by virtue of the discontinuity at the edge of the overlap and the reduced slab guide index in the single layers of silicon on either side (1.8 to 2 in TM, compared to 2.4 in the overlap region). An array of Vernier alignment boxes was designed to allow adjustment and measurement of the guide width from 2 to 8μm in both x and y directions by shifting the upper silicon pattern with respect to the other mask layers, while the device layouts were designed to allow for these wide alignment tolerances without incurring open or short circuits (see figure 5.9).
Merlin comprised two sets of modulators at right angles to each other, with a range of lengths from 50 to 560μm. Two long devices (2000 and 3000μm) were also included, with both contact pads at the same end so that the lengths could be adjusted by cleaving the sample.

The design layout is shown in figure 5.10.
Modulators: 50 to 560μm

Modulators: 2 and 3mm

Test features: Capacitors

Edge test features

Resistors

Transistors

Overall chip size: 2 x 3mm (approx.)

Figure 5.10. Merlin: chip layout

Merlin included a range of test features for process monitoring:

(a) resistors, in each silicon layer, ranging from 1/100 to 100 squares;

(b) capacitors, both simple and double: the "double" versions consisted of two capacitors in series, with both external contacts to the upper silicon layer or metal, so that they could be fabricated without having to pattern the lithium niobate (see figure 5.11);
Note that the dielectric may be left as a continuous film, avoiding the need to etch it.

Figure 5.11. Double capacitors.

Figure 5.12. Potential metal step coverage problems.

(c) field effect transistors\(^1\);
(d) edge test features, to test for metal or silicon open circuits over various edges, as well as short circuits due to fillets between neighbouring tracks over various edges (see figure 5.12).
5.4. Summary

Each of the processes necessary for fabricating parallel plate modulators was examined first in isolation, then in combination with the others. Logistical problems of device fabrication were identified and overcome, and a process sequence was developed. A "sacrificial buffered oxide" process was also devised, to cope with potential nonuniformity problems in extending the process to large substrate sizes.

A mask set was designed for this fabrication process, incorporating a range of test features and modulator structures.

This work led to the fabrication of prototype modulators: assessment of these devices is discussed in Chapter 6.

References


Chapter 6: Optical experiments

Abstract

This chapter is concerned with optical waveguiding measurements of the modulator structure. As was discussed in Chapter 3, the modulation in these devices was expected to contain contributions from several effects: the electro-optic effect in the lithium niobate, free carrier injection modulation in the electrodes, via both absorption and refractive index changes. As was also mentioned in Chapters 3 and 5, a decision was made to fabricate devices using relatively thick lithium niobate layers (over 0.2μm) in order to reduce the likelihood of short circuits. This added to the complexity of the devices' behaviour, as the structures were multi-moded, supporting two TM and two TE modes. The aim of this work was to measure the device characteristics, and to try to separate out, as far as possible, the multifarious effects and interpret them in terms of the theoretical understanding developed in Chapter 3 and the materials work of Chapter 4.

As will be discussed below, the measurement of modulation involved the technique which was mentioned in Chapter 3, i.e. interference between two or more modes of the same waveguide. All of these modes were subject to modulation effects to varying degrees, for reasons mentioned above.

The chapter begins with a discussion of the theoretical aspects of the experiments: calculations are presented of the effect of the inclination of the crystallographic axes of the lithium niobate films on silicon and of the interference scheme employed in these experiments. This is followed by a discussion of the experimental arrangement which was devised for testing the devices. A number of measurements are then described, such as the frequency response and the dependence of modulation depth on the polarisation. Finally, the results are summarised and the good agreement with predicted characteristics is put forward in support of the theory developed in Chapter 3 and the lithium niobate growth work of Chapter 4.

6.1. Inclined crystallographic axes

X-ray studies (Chapter 4) indicated that the lithium niobate in the device regions was not crystallographically aligned to the waveguide axes: the material was <012>
oriented, so that the extraordinary (c) axis was tilted at about 57° to the device normal.
The lithium niobate layer was composed of crystallites with a variety of in-plane orientations: crystal c-axis projections on the device surface were not aligned to the waveguide axis. This led to a reduction in the net electro-optic coefficient (as compared to the idealised situation with c-oriented material), modulation of the TE modes as well as TM, and TE-TM mode conversion[1].

6.1.1. Electro-optic coefficients

The effective electro-optic coefficients were diminished for two reasons. First, the crystal c-axis was inclined to the applied electric field, so that the electric field component parallel to the c axis of the crystal was reduced. Second, the crystal c-axis was inclined to the direction of polarisation of the optical wave, so that the change in the extraordinary refractive index was only partially effective in modulating the optical wave. As will appear below, these effects were mitigated by the r51 electro-optic coefficient which contributed a significant degree of modulation in this inclined-axis situation.

The electro-optic calculation below is divided into three parts. First, the variational analysis discussed in § 3.1.1 of Chapter 3 is applied to the case of inclined crystallites, so as to determine the effect of a given change in the index ellipsoid[2] on the propagation constants. Secondly, the effect of the applied electric field on the index ellipsoid of lithium niobate is calculated for the structure fabricated here - i.e. with the crystallite orientations as inferred from the X-ray studies of Chapter 4. Thirdly, these two results were combined with the electro-optic efficiency analysis developed in Chapter 3 (§ 3.1.1), to yield the change in effective index of each mode of the waveguide.

Although the waveguide model, MOD2, was not able to simulate propagation constants of waveguides with inclined crystallographic axes, the electro-optic effects could be calculated using the variational expressions discussed in § 3.1.1 of Chapter 3. This method stemmed from an expression for the change Δβ in the wavenumber β:

\[ 2\beta \Delta \beta = \varepsilon_0 \omega \int \Delta \varepsilon E \cdot E^* \, dx \]  \hspace{1cm} (6.1, from 3.1)

To apply this to the case of inclined crystallographic axes, \( \varepsilon E \cdot E^* \) is replaced by components \( \varepsilon_i E_i E_i^* \), where \( E_i \) is the component of \( E \) along the crystallographic axis.
denoted by the subscript j, and is equal to \( E_x \cos \theta_{xj} \). In these expressions, x, y and z refer to the waveguide axes. Applying this to the integral, \( \Delta \varepsilon \) is replaced by \( \Delta \{ \varepsilon_j \cos^2 \theta_{xj} \} \) which contains two terms, representing both stretching and tilting of the refractive index ellipsoid. The change in \( \beta / k_0 \) is then given by

\[
\Delta \{ \beta / k_0 \} = \frac{1}{2} \varepsilon_0 \omega \int \sum \{ \Delta \varepsilon_j \cos^2 \theta_{xj} + \varepsilon_j \Delta (\cos^2 \theta_{xj}) \} \cdot |E_x|^2 \cdot dx,
\]

where the power P is normalised and where the summation is taken over all combinations of x, y, z and j. Converting, as before, to an expression involving the z-component of the Poynting vector, this yields

\[
\Delta \{ \beta / k_0 \} = \frac{\beta / k_0}{n^2} \sum \{ n_j \Delta n_j \cos^2 \theta_{xj} + \frac{1}{2} n_j^2 \Delta (\cos^2 \theta_{xj}) \} \cdot \int S_z \cdot dx \quad (6.2)
\]

for TM modes, and

\[
\Delta \{ \beta / k_0 \} = \frac{1}{\beta / k_0} \sum \{ n_j \Delta n_j \cos^2 \theta_{yj} + \frac{1}{2} n_j^2 \Delta (\cos^2 \theta_{yj}) \} \cdot \int S_z \cdot dx \quad (6.3)
\]

for TE modes. In these expressions, the summations are taken over j's, i.e. over the three crystallographic axes x', y', z'. As in Chapter 3, the expression assumes negligible z field components. In both cases, the integral is taken over the lithium niobate region.

To evaluate these expressions, \( \beta / k_0 \) and \( \int S_z \cdot dx \) were calculated using a value of 2.25 for the refractive indices of the lithium niobate layer, incurring no great error in the final result since the birefringence was small.

The \( \Delta n_j \) and \( \Delta \theta_{xj} \) terms were calculated using the electro-optic coefficients of lithium niobate. In the present case, the field is applied along the \(<012>\) direction, which lies in a symmetry plane equivalent to the negative y-axis\(^{[3]}\), as can be seen from figure 6.1. The problem can therefore be simplified by transforming to axes x', y', z', where x' is chosen to lie along \( y^* \), y' is along \( y_h \), and z' is along \( z_h \). The details of the calculation are set out in Appendix 6A.
Figure 6.1. The relationship between the hexagonal \((x_h, y_h)\) and orthorhombic \((x_o, y_o)\) axes, and the reciprocal lattice vectors \((x_h^*, y_h^*)\) of lithium niobate, and the symmetry planes (denoted by the symbol \(S\)). Calculations were performed by transforming from the conventional orthorhombic axes \((x_o, y_o, z_o)\) to a more convenient choice \((x', y', z')\) as shown.

The effect of the applied electric field on the index ellipsoid can be expressed in terms of stretching and tilting:

\[
1/n_{x'}^2 = 1/n_o^2 + E_o(r_{13}\cos\theta_o - r_{22}\sin\theta_o) \\
1/n_{y'}^2 = 1/n_o^2 + E_o(r_{13}\cos\theta_o + r_{22}\sin\theta_o) \\
1/n_{z'}^2 = 1/n_e^2 + E_o r_{33}\cos\theta_o
\]

where \(E_o\) is the electric field applied along the waveguide x axis, and \(\theta_o\) is the angle between the waveguide x axis and the crystal c axis, which was calculated in Chapter 4 to be 57°. The tilt is given to first order by

\[
\Delta\theta_o = \delta = \frac{n_o^2 n_e^2 E_o r_{51}\sin\theta_o}{n_e^2 - n_o^2} (6.4)
\]

This is illustrated in figure 6.2.
Figure 6.2. The effect of the applied electric field on the refractive index ellipsoid. The ellipsoid is tilted through an angle \( \delta \), and the principal axes are stretched by amounts \( \Delta n_x \), \( \Delta n_y \), and \( \Delta n_z \) as shown.

The changes in refractive index are given by

\[
\Delta n_x' = -\frac{n_o^3}{2} \left( r_{13}\cos\theta_o - r_{22}\sin\theta_o \right) E_o
\]  
\( (6.5a) \)

\[
\Delta n_y' = -\frac{n_o^3}{2} \left( r_{13}\cos\theta_o - r_{22}\sin\theta_o \right) E_o
\]  
\( (6.5b) \)

\[
\Delta n_z' = -\frac{n_o^3}{2} r_{33}\cos\theta_o E_o
\]  
\( (6.5c) \)
Finally, the $\cos^2 \theta_{ij}$ and $\Delta(\cos^2 \theta_{ij})$ terms in equations (6.3) were obtained from the direction cosines between the crystal and waveguide axes, which are apparent by inspection of figure 6.3 and are set out in Table 6.1.

![Diagram of crystal axes and waveguide axes](image)

Figure 6.3. The relationship between the transformed crystal axes ($x', y', z'$), ($x'', y'', z''$), and the waveguide axes ($x_{wg}, y_{wg}, z_{wg}$).

Table 6.1. Direction cosines for conversion between waveguide and crystal axes. Note: $\phi$ is the angle between $y$ and the projection of $z'$ on the $yz$ plane.

<table>
<thead>
<tr>
<th>Waveguide axes</th>
<th>x</th>
<th>y</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crystal axes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x'$</td>
<td>$\sin \theta_o$</td>
<td>$-\cos \phi \cos \theta_o$</td>
<td>$-\sin \phi \cos \theta_o$</td>
</tr>
<tr>
<td>$y'$</td>
<td>0</td>
<td>$-\sin \phi$</td>
<td>$\cos \phi$</td>
</tr>
<tr>
<td>$z'$</td>
<td>$\cos \theta_o$</td>
<td>$\cos \phi \sin \theta_o$</td>
<td>$\sin \phi \sin \theta_o$</td>
</tr>
</tbody>
</table>

Finally, equations (6.4) and (6.5a to c) may be combined with these cosine terms to evaluate the summation terms in equations (6.2) and (6.3):
For TM modes: (Ex terms only) Relative values, x10\(^{12}\)m/V

\[
E_0\{ \frac{-n_o^4}{2} (r_{13}\cos\theta_o - r_{22}\sin\theta_o)\sin^2\theta_o + 2
\]

\[
- \frac{n_e^4}{2} r_{33}\cos^3\theta_o - 53
\]

\[
- n_o^2n_e^2r_{51}\sin^2\theta_o\cos\theta_o \} - 274
\]

Total: -325

For TE modes: (Ey terms only)

\[
E_0\{ \frac{-n_o^4}{2} (r_{13}\cos\theta_o - r_{22}\sin\theta_o)\cos^2\theta_o\cos^2\phi + 0.4
\]

\[
- \frac{n_o^4}{2} (r_{13}\cos\theta_o + r_{22}\sin\theta_o)\cos^2\theta_o\sin^2\phi - 19
\]

\[
- \frac{n_e^4}{2} r_{33}\cos\theta_o\sin^2\theta_o\cos^2\phi - 128
\]

\[
+ n_o^2n_e^2 r_{51}\sin^2\theta_o\cos\theta_o\cos^2\phi \} + 274
\]

Total: +127

The "relative values" column was calculated by putting \(\theta_o = 57.3^\circ\), and using an average value of \(\cos^2\phi = \sin^2\phi = 0.5\). Values for the electro-optic coefficients were taken from the low-frequency values quoted in reference [4] (and reproduced in Appendix 3A). For comparison, the term corresponding to modulation of the TM polarisation in c-oriented lithium niobate is \(n_o^4r_{33}/2\), which scores 336 on the same scale.

It can be seen from these equations that the \(r_{33}\) term in the modulation of the TM mode is reduced by a factor \(\cos^3\theta_o\), while the TE mode acquires an \(r_{33}\) modulation term equal to \(n_e^4r_{33}\cos\theta_o\sin^2\theta_o\cos^2\phi\). It is interesting to note that with the axes inclined in this way, the \(r_{51}\) term, representing the effect of the field in altering the inclination of the index ellipsoid, is greater than any of the other terms: in the case of TM modes, dominated by \(E_x\), this adds to the stretching terms, and the overall electro-optic effect is almost equal to that which is predicted for the c-axis-normal situation (325 vs. 336, or 97%). In the case of TE modes, however, where the optical electric
field is along y, the effect is opposed to the stretching terms, and the overall electro-optic effect is much weaker (127 vs. 336, or 38%).

Finally, these equations can be substituted into equation (6.2) and values can thus be obtained for the modulation of $\beta/k_o$ per unit applied voltage for each mode for a given device structure. These figures are given in Table 6.2.

Table 6.2. $\Delta(\beta/k_o)$ per volt, for devices P44. For comparison, the corresponding figure for TM0 modulation with aligned crystal/waveguide axes is $4.39 \times 10^{-4}$ per volt.

<table>
<thead>
<tr>
<th>Mode</th>
<th>$\Delta(\beta/k_o)$ per volt</th>
</tr>
</thead>
<tbody>
<tr>
<td>TM0</td>
<td>$4.25 \times 10^{-4}$</td>
</tr>
<tr>
<td>TM1</td>
<td>$6.08 \times 10^{-5}$</td>
</tr>
<tr>
<td>TE0</td>
<td>$-5.05 \times 10^{-5}$</td>
</tr>
<tr>
<td>TE1</td>
<td>$-1.88 \times 10^{-5}$</td>
</tr>
</tbody>
</table>

6.1.2. Mode conversion

Mode conversion effects were important to these devices, since they provided potential loss mechanisms by coupling to radiation modes (i.e. scattering) as well as by coupling to lossy guided modes. In addition to this, strong coupling between the reference and modulated signals would reduce the relative modulation efficiency, by causing a modulation of the reference signal while reducing the modulation of the other.

The method used above for the electro-optic calculation can also yield an estimate of the coupling coefficients between different modes. The modal electric field induces a electric displacement in the material: this displacement may be viewed as an excitation source, coupling to other guided modes and to radiation modes$^{[1]}$. The calculation is in three steps: first, to calculate the excitation field from a knowledge of the input field (i.e. the modal distribution) and the actual (i.e. non-idealised) refractive index profile; second, to perform the overlap integral between this excitation field and the modal distribution of the secondary mode; thirdly, this overlap integral could then be used in coupled mode theory$^{[5]}$ to calculate the rate of power transfer from one mode
to the other as they travel down the waveguide, and hence the power in each output mode as a function of the input polarisation.

\[
D = \varepsilon_x E_x' + \varepsilon_z E_z'
\]

\(D\) contains a component along to \(y_{wg}\), which propagates as a TE mode.

Figure 6.4. TM-TE conversion due to off-axis crystallites of birefringent material.

TE-TM conversion, in the present situation, arises from the inclination of the index ellipsoid to the right or the left with respect to the propagation direction, as shown in figure 6.4. From symmetry arguments, it can be seen that switching the crystallite inclination from left to right results in a reversal of the phase of the induced field. Since there are, on average, equal numbers of crystallites inclined in each direction, the net power transferred should be zero to first order. Strictly speaking, there may be a slightly preferred orientation in the system, determined by the orientation of the underlying sapphire with respect to the waveguide. However, the presence of such asymmetry was not detected in the X-ray data discussed in Chapter 4.

The same symmetry argument implies that TM0-TM1 and TE0-TE1 conversion is also weak. This is because the index perturbations are, on average, zero, so that the first order coupling coefficient vanishes.

Second order effects can also give rise to mode coupling, via random variations in the predominant crystallite orientations over a large scale. This may be seen by considering \(N\) independent phase contributions \(E_i\) distributed with a standard deviation \(\sigma\) about zero: the sum, \(\Sigma E_i\), of these contributions, is distributed with a standard deviation \(\sigma \sqrt{N}\), so that the total power transfer, \((\Sigma E_i)^2\), exhibits a \(\chi^2\) distribution with a mean \(N\sigma^2\). Mode conversion was observed on certain films, by prism coupling at
visible wavelengths, as described in Chapter 4. In the case of these devices, operating at longer wavelengths, the effect was expected to be smaller because of the inverse dependence of $E_i$ on the fourth power of the wavelength$^{[2]}$ as well as the short device lengths.

### 6.1.3. Inclined crystallographic axes: summary

The relative amplitudes of phase modulation of the four guided modes are given in Table 6.3. As discussed in Chapter 3, the effects arise from phase and absorption modulation in the silicon as well as the electro-optic effect in the lithium niobate. From this table, the dominant effect is expected to be TM0 modulation, via the electro-optic effect in the lithium niobate layer.

**Table 6.3. Modulation per volt, for a 400µm modulator such as P44 D5.**

Note. The figures for absorption modulation indicate the change in signal transmission: although not directly comparable to the phase figures, it is apparent that these are small and can be neglected for the present purposes.

<table>
<thead>
<tr>
<th>Layer</th>
<th>Mechanism</th>
<th>TM0</th>
<th>TM1</th>
<th>TE0</th>
<th>TE1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lithium niobate</td>
<td>phase change</td>
<td>-40°</td>
<td>-5.7°</td>
<td>4.8°</td>
<td>1.8°</td>
</tr>
<tr>
<td>Upper silicon</td>
<td>absorption</td>
<td>0.4%</td>
<td>0.4%</td>
<td>1.2%</td>
<td>0.5%</td>
</tr>
<tr>
<td></td>
<td>phase change</td>
<td>1.3°</td>
<td>1.5°</td>
<td>4.2°</td>
<td>1.8°</td>
</tr>
<tr>
<td>Lower silicon</td>
<td>absorption</td>
<td>0.3%</td>
<td>0.4%</td>
<td>0.4%</td>
<td>1.1%</td>
</tr>
<tr>
<td></td>
<td>phase change</td>
<td>0.9°</td>
<td>1.3°</td>
<td>1.4°</td>
<td>3.8°</td>
</tr>
</tbody>
</table>

It should be remembered that the relative signs of these figures depend on the poling state of the lithium niobate, and the dopant types in the silicon layers. In the present case, the silicon layers are similarly doped, so the effects in the two layers are opposed to each other (since the induced charges are always opposite in sign) and will tend to cancel out.
6.2. Measurements: experimental arrangement

Several difficulties were associated with the comparatively long operating wavelength of the devices (1.53\(\mu\)m). As was discussed Chapter 3, this wavelength was chosen to be beyond the silicon band-gap: a fortiori, the photon energy was below that detectable by silicon photodiodes; it was also outside the range of most phosphor cards and cameras. An infra-red LED [6] was obtained for this work: even with this, the pictures were faint and the response time was slow. A germanium photodiode [7] was obtained for performing quantitative measurements.

The waveguides were very thin, with typical mode sizes about 0.5\(\mu\)m, or less than one third of the free-space wavelength. This small mode size was chiefly due to the high refractive index of silicon (circa 3.47). The minimum focus spot size from the end-firing optics, on the other hand, was of the order of a micron or more. The coupling efficiencies at both ends of the waveguide were further reduced by the high reflectivity of silicon (also due to its high index). These effects, combined with losses in the waveguide, limited the available waveguiding power to a small fraction of the laser output (circa 8mW).

The waveguides in this study were about 400\(\mu\)m long. The input and output of the waveguides were therefore quite close together and it was necessary to devise a method for detecting the guided output signal without the detector being swamped by non-coupled light from the input focus. This was achieved using the arrangement shown in figures 6.5 to 6.7 and described below.

The infra-red laser beam (InP, \(\lambda = 1.53\mu m\)) was directed onto a steering mirror (M1), and thence through a pellicle beam splitter [9] (S1). A helium-neon visible laser beam was steered by two mirrors (M2 and M3) onto the pellicle splitter (S1). The pellicle splitter consisted in a membrane of thickness just over 4\(\mu\)m and refractive index 1.457 at 0.6328\(\mu\)m. The reflectivity was therefore dependent on wavelength, polarisation and angle of incidence via the Brewster and Fabry-Perot effects [9]. The angles of incidence were chosen to give near-zero reflection in the infra-red, while reflecting appreciably in the visible. In this way, the visible and infra-red beams were combined with minimal loss of infra-red power. By judicious adjustment of S1 and M1, the two beams were made to coincide so that the visible beam could be used for coarse alignment of the rest of the optical bench.
Beyond the splitter, the beams were directed via a half-wave plate (P1) and two mirrors (M4 and M5) and focused by an objective lens (L1) onto the end of the waveguide.

The output beam was focused by a second objective (L2) through a Glan-Foucault polariser (P2) and an aperture A1 onto the detector D1. Two detectors were used: an infra-red leddicon camera and a germanium photodiode. Modulation measurements were obtained by feeding the photodiode signal through a transimpedance amplifier to a lock-in amplifier as shown in figure 6.5.

Figure 6.5. Schematic diagram of the optical bench used for infra-red waveguiding and modulation measurements
A crucial feature of this arrangement was that the detection optics (L2 onwards) were at an angle to the waveguide axis and the launching optics. This was not only done for reasons of space (parts of the device, such as bond wires and so forth, were in the way of any on-axis detection optics), but also to reduce direct cross-talk between launch and detector optics. Using this arrangement, much of the non-coupled light tended to continue propagating near the waveguide axis, missing the detector lens L2. Some non-coupled light was nevertheless collected by L2: however, the angular separation between the two ends of the waveguide, as seen from L2, enabled this light to be eliminated by placing a spatial filter in front of the detector. This is illustrated in figures 6.6 and 6.7. A further advantage of this arrangement was that it provided a coarse alignment microscope: by focusing on the input end of the waveguide, the focus and alignment of the launch spot could be adjusted to within a few microns.

Figure 6.6. Detail of apparatus for infra-red waveguiding.
6.1.2. Measurement of phase modulation

As was mentioned in Chapter 5, it was decided to minimise the complications attendant on waveguide bends and junctions, and to design only straight rib waveguide modulators, formed by narrowly overlapping silicon patterns. The disadvantage with
this type of device was that it provided only phase modulation, and a reference beam was required to interfere with the signal in order to measure the effect.

The conventional method of achieving this is to use a bulk Mach-Zehnder arrangement, in which the laser beam is divided into two paths using a beam splitter, recombining the two signals using a second splitter beyond the modulator\textsuperscript{[12]}. In the present work, however, this approach suffered from problems of poor overlap and vibrations, chiefly because of the very small mode sizes: added to the problems of working with weak signals at long wavelengths, these difficulties made modulation measurements practically impossible.

An alternative arrangement was devised, using the scheme outlined in Chapter 3, in which both TE and TM modes were simultaneously launched into the waveguide by adjusting the half-wave plate (P1) in the launch optics. In the detection optics, the (unmodulated) TE mode could then be used as a reference beam, interfering with the (modulated) TM mode after passing through a polariser set at an angle intermediate between the two polarisations. This scheme is shown schematically in figure 6.8. This arrangement avoided most of the vibration and overlap problems associated with the Mach-Zehnder arrangement, while allowing adjustment of the relative intensities of reference and modulated beams (by means of the half-wave plate and the analysing polariser) to make best use of the meagre beam power available.

Variations on this scheme are also possible, using the other modes (TM1 and TE1): the mathematical analysis given below applies to a general situation, using a spectrum of modes in both polarisations.

Let the launch optics be such as to set up modal fields

\[ H_j = p_j H_0, \]

where \( H_0 \) is the total field available, and the \( p_j \) coefficients determine the field distribution among the modes.

These fields are transmitted through the device, emerging with attenuation and phase \( \exp(-\alpha_j) \) and \( \exp(-i\phi_j) \) respectively; finally, these are recombined in the detection system, with weights \( q_j \) representing the coupling efficiencies of each mode, to give a total field

\[ H_{\text{meas}} = \sum \{ p_j q_j \exp(-\alpha_j) \exp(-i\phi_j) H_0 \}. \]
The signal intensity $I_{\text{meas}}$ is proportional to the squared modulus of $H_{\text{meas}}$, which is equal to

$$I_{\text{meas}} = H_0^2 \left\{ \sum A_j^2 + 2 \sum A_j A_k \cos(\phi_j - \phi_k) \right\} \quad \text{(6.6)}$$

where $A_j = p_j q_j \exp(-\alpha_j)$. 

Figure 6.8. The TE-TM interference modulation scheme.
In the special case where only the phases are modulated, the modulation signal arises only from the second term,

$$2H_0^2A_iA_j\cos(\phi_1-\phi_2).$$

In the case of TM0-TE0 operation, let the input polarisation and the analysing polariser be set at angles $\theta_{in}$ and $\theta_{out}$ to the "TM" plane: then $p_1$ and $q_i$ become

$$p_1 = C_{1in}\cos(\theta_{in}), \quad p_2 = C_{2in}\sin(\theta_{in}),$$

$$q_1 = C_{1out}\cos(\theta_{out}), \quad q_2 = C_{2out}\sin(\theta_{out}),$$

where the $C$ terms are fixed coupling efficiencies; substituting, the $A_1A_2$ term becomes

$$2H_0^2\sin(2\theta_{in}).\sin(2\theta_{out}).C_{1in}C_{2in}C_{1out}C_{2out}\exp(-\alpha_1-\alpha_2).$$

This implies that the signal is maximised when both polarisers are at $45^\circ$ to the waveguide plane, irrespective of inequalities between the modal losses, whereupon the sine terms are both equal to one and the peak-to-peak amplitude of the full modulation (i.e. $\phi_1-\phi_2$ varying through a full cycle) is just

$$I_{mod} = 4H_0^2C_{1in}C_{2in}C_{1out}C_{2out}\exp(-\alpha_1-\alpha_2). \quad (6.7)$$

The modulation depth $I_{mod}/I_{meas}$ can be obtained by setting $\cos(\phi_1-\phi_2)=1$ in equation 6.6, giving the maximum signal as $H_0^2\left(\sum A_j\right)^2$ so that, writing $I_{mod}$ in terms of $A_j$'s, we arrive at a familiar result,

$$\frac{I_{mod}}{I_{meas}} = \frac{4A_1A_2H_0^2}{\left( A_1 + A_2 \right)^2}$$

These result can be generalised to other two-wave interference schemes: whatever the fixed losses and inefficiencies in the system, the *adjustable* split ratios should be 1:1 at both ends of the device if the total modulation is to be maximised, whereas the *total* split ratio should be 1:1, for equal power in reference and modulated signals at the detector, if the modulation depth is to be maximised at 100%. If the fixed losses are the same for the two signals, these two conditions will coincide.

This scheme may be used for TM0-TM1 operation by aligning both polarisers for TM transmission: the split ratio is then adjusted by means of the launch objective position. The modulation amplitude is then given by this same equation with the appropriate coupling and transmission losses.
When using TM0-TM1 interference, the modulation can clearly not be detected via a polarisation change: instead, the output light is modulated in the Fourier plane and the analysing polariser is replaced by spatial filtering. In these experiments, spatial filtering was achieved by virtue of the finite aperture of the collection objective. Although an exact simulation of the experiment requires a precise knowledge of the shape of the end of the waveguide, the principle of operation may be demonstrated by considering the simplified structure shown in figure 6.9.

Essentially, the light travelling down the waveguide emerges primarily from two points, A and B, corresponding to the ends of the two silicon stripes. The microscope objective collects the emerging light within a solid angle as shown, and focuses it onto a detector. To a first approximation, the wave amplitude $E_D$ at the detector due to a guided wave with effective index $\beta/k_o$ is given by

$$E_D = \int_{\theta} (E_A + E_B e^{i\phi}) \cos \theta \, d\theta. e^{-i\omega t}$$

where $E_A$ and $E_B$ are the electric fields due to the optical energy in the two silicon layers at the point A: the phase difference between these is represented by the $e^{-i\phi}$ term, where $\phi$ is given by

$$\phi = (\beta - k_o \cos \theta)L.$$
The cosθ term in the integrand represents approximately the effect of the shape of the waveguide end, resulting in light being emitted primarily in the forward direction. The integral is taken over the values of θ which correspond to the collection aperture of the detection optics.

The integrand may be re-written

\[ \{ 2E_B \cos(\phi/2)e^{-i\phi/2} + (E_A - E_B) \} \cos\theta \]

The second term is constant, representing the effect of an unbalanced split in optical energy between A and B, whereas the first term is oscillatory, with a period of 2π in φ. The squared modulus of the integrand is the radiation pattern emerging from the waveguide: example plots are shown in figure 6.10 for TM0 and TM1. Also shown in the figure are the variations in the phases of the fields.

![Figure 6.10. Radiation patterns for the TM0 and TM1 modes of the modulator.](image)

The total field at the detector can be calculated by integrating the sum of the two radiation patterns (TM0 and TM1) over the solid angle represented by the collecting microscope objective. As TM0 is phase-modulated, its radiation pattern changes not in amplitude but in overall phase. When added to the radiation pattern of the TM1 mode, this results in energy redistribution, so that the signal collected by the microscope objective is altered. This is illustrated in figure 6.11, which is a plot of the squared
modulus of the sum of the two radiation field patterns, i.e. the radiation pattern of optical energy as a function of $\theta$, for phase modulations of $0, \pi/2, \pi$, and $3\pi/2$. The energy radiated at a given angle varies sinusoidally with the phase modulation: although the total energy emerging remains constant, the energy collected by an objective of finite aperture also exhibits the same sinusoidal behaviour.

![Diagram showing radiation pattern and field splits](image)

**Figure 6.11.** Radiation pattern of optical energy due to the interference of TM0 and TM1 modes emerging from the end of the modulator, for four values of the phase shift electro-optically induced on TM0.

6.3. Measurements: experimental results

6.3.1. Polarisation

Input polarisation was controlled by adjusting the half wave plate (P1 in figure 6.5). The position of the half-wave plate was indicated by a scale, relative to an arbitrary zero position. The TM polarisation was determined to correspond to positions marked $20^\circ$, $110^\circ$, $200^\circ$ and $290^\circ$ on this scale. Mode conversion was examined using the video camera, by launching TE modes while the output polariser was set for TM, or vice versa. No significant output was observed in either case, in agreement with the theoretical expectations discussed above.

As mentioned above, the devices were expected to function by interference between the modulated TM0 mode and other modes which were essentially unmodulated. The modulation was measured as a function of input polariser position, optimising as far as possible both the launch objective positioner and the output
polariser for each measurement. The results are shown in figure 6.12, and two distinct regimes are visible, depending on the output polariser position. These were interpreted as being chiefly due to two effects, TM0-TE0 and TM0-TM1 interference. Significant differences were found between the optimum position of the output polariser and the theoretically expected settings: these were believed to be partly due to optical imperfections in the output polariser. Another reason for the deviations, in the case of TM0-TE0 interference, was the almost unavoidable contribution from TM0-TM1 interference under these polariser settings. When operating in the TM0-TM1 regime, on the other hand, the pure TM polarisation precluded any contributions from the TE modes.

![Figure 6.12. Modulation signal versus input polarisation.](image)

Notes. The optical system was re-optimised after setting the half wave plate at each position. Data set A (+, *) was inferred to be due to interference between TM0 and TE0; Data set B (x, o) to interference between TM0 and TM1. The curves indicate theoretically expected trends.

The photodiode signal modulation is plotted as a function of output polariser position in figure 6.13. The input polariser was set at 180° which implied combined TE/TM operation. The distortion of the cycle (i.e. the lack of 180° rotational symmetry) was ascribed to optical imperfections in the output polariser.
6.3.2. Launch positioner

The graphs in figure 6.14 show the amplitude of optical modulation as a function of the launch objective position, for two settings of the launch and detection polarisers. The launch objective was shifted by means of a piezoelectric transducer. The top left hand graph shows the behaviour under purely TM polarisation, so that the mechanism is TM0-TM1 interference. The top right hand graph shows behaviour under mixed TE-TM polarisation, as discussed in the last section. Below these graphs, theoretical curves are illustrated. These were calculated by performing overlap integrals to determine the coupling efficiency between the launch objective and the waveguide modes as follows.

The launch objective provides a focused beam of light, which is represented by a distribution function \( L(x-x_0) \), where \( x \) is the vertical coordinate with respect to the waveguide axes as defined in Chapter 2, and \( x_0 \) is the position of the centre of the distribution. This field sets up a corresponding field \( L_{\text{wg}}(x-x_0) \) inside the waveguide. This distribution is viewed as a superposition of various guided and radiative modes of the structure: the proportion accorded to a given mode \( M_u(x) \) is then obtained using the well-known orthogonality of the waveguide modes\(^1\),

\[
\int E_u \times H_v^* \, dx = \delta_{uv}
\]

where \( \delta_{uv} = 1 \) (\( u = v \)) or 0 (\( u \neq v \)) and where the distributions are normalised with respect to optical power, as before.
Then the coupling efficiency $C_u$ between $L_wg(x-x_0)$ and $M_u(x)$ is given by

$$C_u = \int L_wg(x-x_0).M_u(x) \, dx$$

where $L$ and $M$ are appropriately chosen field components.

Finally, the modulation amplitude is obtained by substituting for $C_{1in}$ and $C_{2in}$ in equation (6.7).

$L_{wg}$ was obtained from $L$ by multiplying through by normal-incidence Fresnel coefficients\[^9\] so as to take into account the different reflectivities of the various strata. Tolerably good agreement between theory and experiment was obtained using the Airy function\[^9\] with a 4$\mu$m spot size (half-amplitude diameter). Although the analysis was greatly simplified, in ignoring both the precise shape of the end of the waveguide and the lateral field variations (i.e. along the y axis) which are inevitably present in practice, this was sufficiently good to differentiate between the two regimes of operation, i.e. TM0-TE0 and TM0-TM1 as discussed above.
6.3.3. DC response

The DC response, i.e. optical signal versus applied voltage, could not be measured directly, since modulation of the photodiode signal was necessary in order to distinguish it from background noise. The devices were therefore operated with a small 300 Hz signal (0.6V peak-to-peak) superimposed on a DC level. The output was measured by means of the photodiode and lock-in amplifier as described above and is shown in figure 6.5. The measurements therefore represented the gradient of the light output as a function of DC bias: the "optical signal output vs. voltage" relationship, plotted in figure 6.15, was obtained by numerical integration. The relationship is linear for small voltages, which is consistent with a linear electro-optic effect, rather than a quadratic relationship which might imply a thermal effect due to electrical heating.

Ideally, this should yield a sinusoidal curve, with a period equal to twice $V_n$, as discussed above: it was found, however, that the relationship was significantly distorted. This was interpreted in terms of several phenomena.

The depletion phenomenon, discussed in Chapter 3, was expected to cause the lightly doped p-type upper electrode became depleted at negative DC bias. As discussed in Chapter 3, the reduction of the electric field in the lithium niobate, as a function of applied voltage, depends on the charge carrier concentration and the thickness and dielectric constant of the lithium niobate layer. Electrode depletion also affected the electrical bandwidth, by reducing the effective thickness (and consequent increase in sheet resistance) of the upper silicon electrode, and increasing the thickness of dielectric (and consequent reduction of capacitance). The net result was predicted to be a marked increase in bandwidth. Taking all these effects into consideration, the predicted optical transmission vs. applied DC signal is shown in figure 6.16, assuming a modulation frequency of 300Hz. A comparison between figures 6.15 and 6.16 indicates that the observed behaviour cannot be satisfactorily accounted for by depletion effects alone.
Figure 6.15. Modulated signal versus applied voltage. Points: from measurements (see text); curve: sinusoid fitted to the negative-bias portion of the data.

Figure 6.16. Calculated device output vs. applied DC signal, showing the effect of electrode depletion.

DC leakage current characteristics were measured, and are shown in figure 6.17. The short-circuit leakage current through the device was calculated from the geometry and from test resistor measurements to be 30nA ±10% at 10V. Comparing this to the measured leakage of 25.5nA, it would seem quite likely that the device was indeed short-circuit under these conditions. Under negative bias, however, the device was lower by about a factor of two, so that the voltage drop across the silicon was halved. In addition to this, the leakage current under positive DC bias was observed to drift: typically, it took about twenty seconds for the current to reach an equilibrium value after each alteration in the bias voltage (cf. Chapter 4, § 4.4.4, figure 4.14).
This provides an explanation for the observed behaviour: under positive bias, the DC leakage current ensured that the DC field across the lithium niobate remained either zero or very small, while the 300Hz signal was not so greatly diminished, resulting in 300Hz optical modulation almost independent of DC bias. Under negative bias, a portion of the DC signal did appear across the lithium niobate, so that the optical signal modulation varied sinusoidally with DC bias, as expected for an ideal device.

A sinusoidal curve was fitted to the negative-bias portion of the curve by estimating the zero-intercepts of the measured data and of its first derivative, yielding a value for $V_n$. The result, $20 \pm 5V$, had a large uncertainty due to the relatively small portion of the curve available, as well as the phenomena discussed above, whose influence could not be precisely determined. This implied phase modulation of $9 \pm 2^\circ/V$, which is not incomparable to the predicted figure of $40 - 5.7 \approx 34^\circ/V$ from the difference between TM0 and TM1 modulation in Table 6.3. The shortfall, of a factor of 3.8, may be explained in part by the electrical leakage - as estimated above, this would account for a factor of about 2. The remaining factor, of about 2, might be explained by structural defects in the lithium niobate (the possible presence of unwanted material phases or orientations, or incomplete poling). These will be discussed in greater detail in Section 6.4. The magnitude of the electro-optic effect, however, is larger than could be explained by modulation effects in the silicon alone.

6.3.4. Frequency response

The frequency response of the devices was predicted from resistivity, dielectric and geometric data, combined using the modelling program, as described in Chapter 3. Resistivity and geometrical factors were measured from electrical probing and microscopic measurements respectively. Measurements were taken using a sinusoidal driving waveform, over the range 3Hz to 10kHz which was found to be well above the
bandwidth. A lock-in amplifier was used to measure both the relative phase and the strength of the optical modulation signal.

Results are shown in figure 6.18, where the modulation depth and phase (relative to the driving waveform) are plotted versus driving frequency. The theoretical curves are also shown. The measured bandwidth is 360Hz, in good agreement with the theoretically calculated value. The decline in optical modulation, towards lower frequencies, is consistent with the hypothesis discussed in the last section, whereby a low-frequency electrical leakage mechanism causes a degradation in performance.

![Figure 6.18. Modulated optical signal versus frequency.](image)

![Figure 6.19. Illustrating bandwidth enhancement under visible illumination.](image)
Figure 6.19 shows measurements taken under visible illumination, illustrating the enhanced bandwidth due to photogenerated carriers. The illumination was provided by a 5mW helium-neon laser ($\lambda = 0.6328\mu m$), with a beam width of approximately 2mm. The bandwidth increased from 510Hz to 730Hz, which was consistent with the measured change in the resistivity of the upper electrode material under these conditions.

6.4. Conclusions

The devices functioned, demonstrating the basic feasibility of parallel-plate waveguide electro-optic modulators. Measurements indicated no serious disagreements between the theoretical behaviour and experiment.

The DC response yielded a figure for $V_{\pi}$, which was in reasonable agreement with the theoretically calculated figure. Although a range of mechanisms no doubt contributed to the overall modulation, the magnitude of the effect was too great to be accounted for except by electro-optic modulation in the lithium niobate layer.

The discrepancy, of a factor of 3.8 between predicted and measured values of $V_{\pi}$, may be explained by a variety of factors. The contribution of electrical leakage current has been discussed above, accounting for a factor of about 2. A likely explanation for the remaining discrepancy lies in structural defects in the lithium niobate layer. Although the X-ray studies described in Chapter 4 indicated the presence of $<012>$ oriented lithium niobate, it proved impossible to state with any degree of certainty the fraction of material which had grown in this way: it is entirely possible that only about half of the material contributed to electro-optic modulation. Moreover, the electro-optic coefficient within that fraction may have been reduced by defects below the scale of the unit cell dimensions, such as lithium vacancies, which would not have been detected by X-ray studies. Reduction in the electro-optic coefficients has been observed in proton-exchanged lithium niobate, where some of the lithium ions are replaced by protons[13].

The frequency response, and the enhancement of the bandwidth under illumination, were both in agreement with theory, lending credence to the electrical modelling.

The behaviour of the devices as a function of input and output polarisations was studied. This, and the measurements of signal strength as a function of the launch
focus position, lent credence to the optical modelling and to the mechanisms predicted for optical modulation.

Altogether, these investigations of the modulator behaviour yielded a range of evidence in support of the theoretical studies and crystal growth work described in Chapters 3 and 4. Having demonstrated that the parallel plate waveguide modulator is feasible, it will be discussed in Chapter 7 whether such a device represents a useful technological advance.

References


182


Chapter 7: Conclusions

Abstract

In this chapter, the achievements of this research programme are summarised and compared to the objectives which were set forth in Chapter 1. The work was successful in demonstrating the feasibility of electro-optic waveguide modulators using thin film lithium niobate. Further work is suggested, e.g. to improve on the top electrode material, and to investigate some alternative device geometries.

7.1. Achievements of this work

The first part of the thesis, Chapters 2 and 3, provided a theoretical basis for this work, examining the behaviour of parallel-plate electro-optic waveguide modulator structures. A computer program was developed for solving the multi-layer lossy waveguide problem, using a complex Newton-Raphson algorithm. The computer model was a seminal part of the present programme, making possible the theoretical investigations and the experimental work which followed.

The investigations of Chapter 3 provided new data and an understanding of the behaviour of lossy multilayer waveguide modulators. The advantages of the parallel-plate approach were seen to comprise electrostatic and optical contributions. It was established that the lithium niobate thickness was a crucial parameter, as it affected the electric field applied to the electro-optic layer for a given voltage: to first order, the device length was directly proportional to this thickness. Minimising the electro-optic layer thickness led to severe optical losses, because of the overlap between the optical energy distribution and the electrically conductive materials needed for electrodes.

Surface plasmon effects were found to exacerbate the optical losses, by concentrating optical energy at the metal-dielectric interfaces. Although the losses associated with high order modes were found to decrease as the lithium niobate was made thicker, the fundamental and first harmonic modes remained intolerably lossy.

The use of TE modes was investigated, and showed some remarkably promising results. Although these entailed using the comparatively weak $r_{13}$ coefficient of lithium niobate, the optical efficiency, $\gamma$, could be enhanced, in some cases by over 300%, by the fact that the effective index of these structures could be made very low (c.
0.6). Even so, the losses in these structures remained unacceptably high for practical devices.

Low-index buffer layers were examined, in an attempt to reduce losses by distancing the optical mode from the metal layers. This was not successful, as the reduction in optical loss required buffer thicknesses at least equal to that of the lithium niobate, whereupon the electrostatic efficiency was drastically reduced. The reduction in electric field strength across the lithium niobate was more severe, by almost an order of magnitude, than suggested by the change in linear dimensions, because of the high dielectric coefficient of lithium niobate. Although this might, in theory, have been alleviated by seeking a buffer material combining a low refractive index with a high dielectric coefficient, a penalty would have been paid in the additional electrostatic energy stored in the buffer layer - leading to a high drive power.

This work indicated some of the more promising approaches for fabricating parallel plate modulators, and ruled out some unfeasible structures. A practical device structure was identified: by using semiconductor instead of metal electrodes, the severe problems of plasmon losses were avoided. This represented a significant departure from the original direction of the work. The move to silicon brought the devices closer to VLSI technology, which was an important step since one of the ultimate aims of this work, as was stated in Chapter 1, was monolithic integration of modulators with silicon devices. Another significant innovation was the development of the composite waveguide. A novel mode of operation was identified in these structures, interfering several modes of a single guide to obtain amplitude modulation from phase modulation.

In the second part of the thesis, Chapters 4 and 5, the fabrication of a prototype device was discussed. The lithium niobate growth work described in Chapter 4 was not new in itself, although a few novel results emerged such as the growth of y-oriented epitaxial lithium niobate on <012> sapphire. The arrangement for applying DC bias to the substrates during RF sputter deposition, not previously reported, was shown to exert a useful influence on the properties of the grown films.

In Chapter 5, the remainder of the device fabrication work was reported. A waveguiding perturbation technique was developed for measuring thickness and refractive index of thin films. Some data was presented on the structural problems encountered in combining lithium niobate with other materials such as silicon and silicon dioxide. The main importance of the work described in Chapter 5, in the context of this thesis, was the development of a fabrication process for the new devices.
The third part of the thesis, Chapter 6, was concerned with optical testing and analysis of the prototype devices. This was successful, in so far as there was a fair agreement (within a factor of 2) between theoretical predictions and experimental measurements. Although lower than bulk values, the electro-optic coefficient in the lithium niobate film was higher (by a factor of about 5) than previously reported results in sputtered thin films of lithium niobate\textsuperscript{1,2}. This may be taken as a vindication of the growth work presented in Chapter 4. The 50% shortfall, with respect to bulk figures, could be explained by incomplete poling, or by defects in the crystal lattice.

The demonstration of electro-optic modulation in these devices was a new result, and represented the achievement of the principal aim of this research programme.

7.2. Parallel-plate modulators and VLSI: feasibility

This thesis has explored the possibilities of fabricating electro-optic waveguide modulators using a thin film of lithium niobate as the active medium, with a view to integration with silicon IC technology. Although a device was successfully demonstrated, it remains to assess whether the scheme is viable and merits further development.

From the point of view of optical performance, the insertion loss due to absorption (a few dB) is not excessive, although likely to be compounded in a practical device by coupling and scattering losses. The coupling losses depend largely on the modal overlap between the modulator and the passive waveguide, and could be minimised by using an extension of the modulator structure to form the passive waveguide. In this arrangement, the silicon would be selectively doped, so that only the modulator region would be electrically active and optically absorbing. A development of this idea is to use a structure of reduced thickness (e.g. by removal of the intermediate lithium niobate or the top silicon layer in the waveguide region), so as to achieve monomode operation in the waveguide while preserving two modes in the modulator. The modal overlaps might thus be arranged in such a way that the TM0 and TM1 modes of the modulator section each coupled with 50% efficiency to the TM0 mode of the passive waveguide: conversely, the TM0 mode of the waveguide would then couple to a superposition of the two modes in the modulator section, as shown in figure 7.1. The TM0-TM1 interference scheme described in Chapter 6 could then be used, to achieve amplitude modulation in a straight-waveguide device, obviating the
need for y-branches and bends. This device is analogous to the Mach-Zehnder configuration, in so far as optical energy is split and recombined, the final output depending on the relative phase modulation between the two portions of the beam: in this arrangement, however, the light is separated in the spatial frequency domain (i.e. modes of the waveguide) rather than physically distinct paths.

![Figure 7.1. Combined modulator with passive waveguide.](image)

From the electrical point of view, the predicted device bandwidths (c. 10GHz) are probably adequate for many applications, although this may be traded against optical absorption to suit a particular application.

Fabrication of these modulators has been demonstrated on VLSI-grade substrates. This has indicated that these two technologies appear to be compatible, although it remains to fabricate and test silicon ICs alongside modulators on a single substrate before compatibility is proven. One potential problem lies in the presence of lithium, which is highly mobile and detrimental to CMOS devices.

In summary, while the prototypes fabricated in this work are clearly not suitable for high speed optical interconnects, they have provided experimental evidence in support of the modelling and materials work, suggesting that a usefully fast and compact device is feasible. As a result, further work is proposed to realise this possibility.
7.3. Further work

Programmes of further work, to complement the research described in this thesis, may be divided into two categories: the first is concerned with extending this research, by improving the performance of the devices reported here, and looking at new ways to exploit the parallel plate modulator structure. The second area of further work is aimed at exploring novel ideas suggested by some of the results reported here.

7.3.1. Device improvements

7.3.1.1. Materials

The main drawback of the demonstrated devices lies in the low bandwidth, about seven orders of magnitude below that predicted in Chapter 3. The primary reason for this lies in the high resistivity of the upper hydrogenated amorphous silicon which was employed as an upper electrode material. This form of silicon was chosen for reasons of convenience, as it was available and could be grown at a relatively low temperature.

An alternative source of material is polycrystalline silicon, grown by low pressure chemical vapour deposition (LPCVD) at temperatures of about 600°C, which is routinely used as an electrode in complementary metal-oxide-semiconductor (CMOS) technology\(^{[3]}\), with resistivities of the order of a few m\(\Omega\)-cm giving a potential bandwidth upwards of 10GHz. This material may suffer from a high degree of optical scattering, as the crystallites are known to grow with significant stresses\(^{[4]}\). Optical scattering may be considerably reduced by growing this material at a slightly reduced temperature (c. 580°C), in the amorphous phase; a high temperature anneal (30 minutes at 950°C) may then be used to crystallise the film, whereupon the resistivity may be made as low as a few m\(\Omega\)-cm\(^{[4]}\).

Another approach would be to use laterally recrystallised silicon\(^{[5,6]}\). In this method, silicon is first deposited over islands or strips of another material, with single crystal silicon beneath. An intense beam of electrons or light is then swept across the structure, melting the top silicon layer. The high thermal conductivity of silicon ensures that the heat flows across and down into the underlying silicon, rather than vertically through the islands: as a result, the thermal gradients are such that the solid/liquid interface sweeps across the islands, seeding epitaxial crystal growth from the underlying silicon.
It may be possible to find a material which combines high conductivity without the accompanying optical loss. Free carrier optical absorption[7] and electrical conduction are phenomena which depend on charge carrier density and mobility:

$$\alpha = \frac{-q_e^3\lambda_0^2 N}{4\pi^2c^3n_0m_e^*\mu} \quad \text{(see Appendix 3B)}$$

$$\rho = \frac{1}{q_e} \frac{1}{N\mu} \quad \text{(see reference [8])}$$

where $q_e$ is the electronic charge, $\lambda_0$ the free space wavelength, $c$ the speed of light in vacuo, $n$ the refractive index, $m_e^*$ the effective mass of the charge carriers, $\mu$ their mobility, and $N$ the carrier density.

Combining these, the ratio $\alpha/\rho$ depends on $N^2$. Indeed, herein lies the advantage of silicon electrodes over metallic layers, since the carrier density is lower in silicon than in silver, enabling a usefully high conductivity to be achieved with much lower optical absorption. Extending this idea further, it might be possible to find a material with an even lower carrier density but high mobility, relaxing the compromise between absorption and bandwidth. A disadvantage of this scheme lies in the depletion phenomenon, discussed in Chapter 3 and demonstrated in Chapter 6, which is exacerbated by low carrier density.

The alternative approach to this problem is to seek a material which exhibits different properties for electrical and optical purposes. These two functions involve charge transport at two separate frequencies: there is no fundamental reason for the effective carrier density and mobility to be similar in both regimes, as the optical frequency (c. $10^{14}$Hz) is much higher than the electrical driving frequency (c. $10^{10}$Hz). It might also be possible to make use of the fact that these devices require electrical conduction parallel to the plane of the device, whereas the optical electric field (in the TM polarisation) is normal to the plane. Anisotropic conductivity has been demonstrated in Langmuir-Blodgett films[9,10], with resistivities below $1\Omega$-cm. It is not unreasonable to suppose that further advances in this field may offer a material suitable for the present devices, either replacing the upper silicon layer or supplementing it.

The lithium niobate growth work of Chapter 4 was restricted to a narrow set of process conditions: the aim was to grow and measure device-quality material in the shortest possible time, rather than to characterise the growth kinetics precisely over a wide range of conditions. This has laid the basis (and prepared the apparatus) for a
thorough investigation which would be necessary for optimising the device performance.

### 7.3.1.2. Device geometry

Several advantages may be gained by modifying the device geometry. The devices described in this thesis relied on one lateral contact to each electrode. Although this benefited from simplicity of fabrication, a top contact to the upper electrode might improve the bandwidth by shortening the electrical conduction path. This is illustrated in figure 7.2. A buffer layer, of a conductive material with a low refractive index (such as indium tin oxide) could be used to distance the metal contact from the optical modes. An added benefit of this idea would be that the upper electrode in such a device need not cover the edge of the lower electrode. It was shown in Chapter 4 that electrical leakage currents were associated with step coverage. Although the currents might be reduced by improvements in lithium niobate growth or silicon etching, the use of a top electrode would circumvent the problem altogether.

![Figure 7.2. Top-contacted parallel-plate modulator.](Image)

The problem of silicon step coverage could be tackled by means of an electrical buffer layer to cover the edge of the lower silicon pattern before lithium niobate deposition, as shown in figure 7.3; an alternative to this arrangement is shown in figure 7.4, where the lower electrode extends beyond the edge of the upper electrode pattern, avoiding the step coverage entirely. In this case, a very thick dielectric buffer
avoiding the step coverage entirely. In this case, a very thick dielectric buffer layer is used to distance the two electrodes outside the active region.

Figure 7.3. Illustrating the use of a buffer layer to avoid electrical breakdown of the lithium niobate at the steps created by the edge of the lower silicon pattern.

Figure 7.4. Extending the lower silicon beyond the edge of the upper silicon, so as to avoid step coverage problems. The buffer layer reduces capacitance.

7.3.2. **Novel applications and implementations**

It was mentioned in Chapter 1 that few attempts have been made to fabricate Fabry-Perot cavity modulators, because of the difficulty of fabricating efficient waveguide mirrors. The structures developed in this work may offer the possibility of achieving this, because of the high refractive index (and hence reflectivity) of silicon and the small vertical dimension of the waveguide. With a mode size no greater than about 0.5μm, waveguide mirrors may be fabricated by etching no more than about 0.5μm through the waveguide, which is likely to be much easier than the several microns associated with waveguides of lower index materials.
### 7.3.3. Other devices

The growth of epitaxial y-oriented lithium niobate on <012> sapphire offers the possibility of an efficient coplanar-electrode modulator operating in the TE polarisation. As was mentioned in Chapter 1, the bandwidth of coplanar-electrode modulators is limited by the velocity mismatch between the microwave and optical signals. This is largely due to the high dielectric constants of lithium niobate, c. 30 along the z axis and 80 along the x and y axes, which lead to a low microwave propagation velocity. Sapphire has a much lower dielectric constant (c. 10) so that a microwave stripline on a thin film of lithium niobate on sapphire would propagate with a higher velocity, reducing the mismatch and increasing the bandwidth\[^{12}\]. Figure 7.5 shows the optical efficiency of this scheme, as a function of the lithium niobate layer thickness, together with $L_\pi$ based on the assumption of a 5V driving signal and an inter-electrode gap of 5μm.

![Graphs showing characteristics of the TE coplanar-electrode modulator](image-url)

Figure 7.5. Characteristics of the TE coplanar-electrode modulator, as functions of lithium niobate thickness, assuming an inter-electrode gap of 5μm.
7.4. Conclusions

The achievements of this work have been summarised and compared to the objectives. The parallel plate configuration has been examined theoretically, and crucial design considerations have been identified: computer modelling has shown that thin film lithium niobate modulators are feasible, in a configuration that is fast (>10GHz), compact (<1mm long), and compatible with VLSI in terms of materials and fabrication. The main aim of the research programme was achieved, viz. the demonstration of a prototype device.

Areas of further work have been identified, aimed at realising the potential capabilities of parallel plate waveguide modulators, and exploring the possibilities of new devices (the straight-guide TM0-TM1 modulator, and the coplanar-electrode device for TE modulation which is made possible by the growth of epitaxial lithium niobate-on-sapphire).

References


12. C.D. Watson, private communication.
Appendices
Appendix 2.A: MOD2 program organisation.

2A.1. Layout

The program is organised as a main unit, MOD2, which calls subroutines for reading data, calculations, and output. These are described below.
MOD2: is the main program;

- **MENU**: reads control data from the menu file (see below);
- **READSTR**: reads data from the structure file (see below);
- **LPWRITE**: changes structure variables for each loop (see below);
- **DISKOP1**: writes header data to the results file;
- **RSTEP**: steps \( \beta/k_o \) through the required range;
  calls
  - **RSOLVE**: to home in on each solution;
- **OUTPUT**: displays solutions on the screen;
- **DISKOP2**: writes solution data to the results file (see below);
- **DISKOP3**: organises data files for plotting modal fields;
  calls
  - **PLOTCLC & PLOTCL2**: to calculate the data for saving;
- **SAVEPLOT**: to write the plot data to the files.

In addition to these,
- **RCALC**: called from RSTEP and from RSOLVE, calculates \( R \);
- **CMEQ, CMMULT**: complex matrix functions, called from RCALC.

The program, when run, prompts the user to supply a Menu file name. The Menu file is written in a standard format, (see example below), using a screen editor. This menu file contains all the control information required, including the name of the Structure file, the range of \( \beta/k_o \) to be examined, the search resolution, the required output format, and information for iterative loops. The loops enable the program to step through a number of values of up to two structure variables, e.g. to obtain calculations for a range of film thicknesses. The variables allowed for these loops are the thicknesses of each layer, and the refractive indices of any region.
2A.2. Example menu file

MOD2 MENU - SOS vertical modulator

Structure file name: merlin.str

Search a range (of Neff) for guided modes:

Range: From: 3.80
To: 1.76
Search resolution: 200
Reference layer: 2
Modes (TE/TM): TM

Output option: 3
1. Beta, real & imag.
2. Beta/ko, real & imag.
3. Beta/ko (real part) & loss (dB/cm)
4. Beta/ko (real part) & log10(loss,dB/cm)
5. As 3 but EO efficiency table (MOD3 only)

Output form: 3
1. Screen only
2. Printer
3. Disk
4. Disk, + files for plotting program

Disk file name: merlin.sln

Iterations:

Loop 1: active
Layer: 2
Variable: nx
No. of loops: 2
Initial value: 2.21
Step: 0.01

Loop 2 (within loop 1): inactive
Layer: 1
Variable: t
No. of loops: 10
Initial value: 0.10E-4
Step: 0.01E-4

NB. After setting up the menu, please ensure that:
1. All entries begin in column 25;
2. The column (54) of full stops is undisturbed; otherwise the program may spout garbage.
Note. The program reserves the right to spout garbage even when the full stops ARE lined up.
2A.3. Example structure file

Lambda: 1.53

Cover: (Air)
1.00, 0 i x
1.00, 0 i y
1.00, 0 i z

Layers: 3

Layer 3: (Silicon)
Thickness: 0.20E-4
3.467, -0.00001 i x
3.467, -0.00001 i y
3.467, -0.00001 i z

Layer 2: C (LiNbO3)
Thickness: 0.23E-4
2.14, 0 i x
2.21, 0 i y
2.21, 0 i z

Layer 1: (Silicon)
Thickness: 0.165E-4
3.467, -0.0008 i x
3.467, -0.0008 i y
3.467, -0.0008 i z

Substrate: (Sapphire)
1.75, 0 i x
1.75, 0 i y
1.75, 0 i z

********************************************************************
2A.4. Example output file

Wavelength: 1.5300

Cover:
1.0000, +0.0000
1.0000, +0.0000
1.0000, +0.0000

Layers: 3
3: .2000E-004
  3.4670, -0.000010
  3.4670, -0.000010
  3.4670, -0.000010
2: .2300E-004
  2.1400, +0.0000
  2.2100, +0.0000
  2.2100, +0.0000
1: .1650E-004
  3.4670, -0.00080
  3.4670, -0.00080
  3.4670, -0.00080

Substrate:
  1.75, +0.0000
  1.75, +0.0000
  1.75, +0.0000

TM modes; beta/k search range: 3.8000 to 1.7600 in approximately 200 steps.
Output is given in the form of beta/ko (real) & loss (dB/cm)
Iterations: L. 2nx = 2.210, 0.000 to 2.220, 0.000 in 2 steps.

L. 2 nx = 2.2100
2 solutions:
1: 2.46288, 45.8 dB/cm
2: 1.93002, 60.4 dB/cm

L. 2 nx = 2.2200
2 solutions:
1: 2.46929, 45.8 dB/cm
2: 1.93086, 60.6 dB/cm
Appendix 3.A: Refractive indices and electro-optic effect in lithium niobate.

The unperturbed refractive index ellipsoid of a birefringent crystal, such as lithium niobate, may be expressed as follows:

$$\frac{x^2}{n_0^2} + \frac{y^2}{n_0^2} + \frac{z^2}{n_e^2} = 1 \quad (1)$$

The refractive index is given by the magnitude of a vector from the origin to a point on the ellipsoid: the direction of this vector being parallel to the electric displacement vector $D$ (or $\varepsilon E$). Mathematically,

$$n(D) = \sqrt{(x^2 + y^2 + z^2)} \quad \text{where} \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} \cdot D = 0$$

Taking into account the electro-optic effect, equation (1) is modified to

$$b_{11}x^2 + b_{22}y^2 + b_{33}z^2 + 2b_{23}yz^2 + 2b_{31}zx^2 + 2b_{12}xy^2 = 1 \quad (2)$$

The subscripts 1,2,3, represent the crystallographic axes $x,y,z$: the pairs of subscripts in equation (2) are usually abbreviated to give

$$b_1x^2 + b_2y^2 + b_3z^2 + 2b_4yz^2 + 2b_5zx^2 + 2b_6xy^2 = 1 \quad (3)$$

The $b$ coefficients are linearly related to the components of applied electric field. The relationship is often expressed as a tensor equation,

$$\Delta b_i = r_{ij}E_j \quad (4)$$

where $\Delta b_i$ is the electro-optically induced change in $b_i$, $E_j$ is the applied electric field component along the crystallographic axis denoted by the subscript $j$, and $r_{ij}$ is the electro-optic coefficient. In the case of lithium niobate, the crystal symmetry is such that many of the coefficients are zero and others are identical, leaving four independent coefficients. The full electro-optic tensor for lithium niobate is as follows:

$$\begin{pmatrix}
0 & -r_{22} & r_{13} \\
0 & r_{22} & r_{13} \\
0 & 0 & r_{33} \\
r_{51} & 0 & 0 \\
r_{51} & 0 & 0
\end{pmatrix} \quad \text{where} \quad \begin{cases}
\begin{array}{c}
r_{13} = 8.6 \times 10^{-12} \text{ m/V} \\
r_{22} = 3.4 \times 10^{-12} \text{ m/V} \\
r_{33} = 30.8 \times 10^{-12} \text{ m/V} \\
r_{51} = 28.0 \times 10^{-12} \text{ m/V}
\end{array}
\end{cases}$$
Note that these figures apply only to high frequency modulation. At low frequencies (below 50MHz\textsuperscript{[1]}), the crystal becomes deformed by the applied electric field, via the converse piezoelectric effect. The deformation contributes a further change to the refractive index. At lower frequencies, the coefficients are as follows:

\begin{align*}
    r_{13} &= 10 \times 10^{-12} \text{ m}/\text{V} \\
    r_{22} &= 6.7 \times 10^{-12} \text{ m}/\text{V} \\
    r_{33} &= 32 \times 10^{-12} \text{ m}/\text{V} \\
    r_{51} &= 32 \times 10^{-12} \text{ m}/\text{V}
\end{align*}

Equation (4) leads to an expression for the change in refractive indices due to an applied field. In the special case where \( E \) is applied along the crystallographic \( z \)-axis, symmetry is preserved and no tilting of the ellipsoid takes place. The change in the \( z \) component of refractive index is then

\[
    \Delta n_z = -\frac{1}{2} r_{33} E_z,
\]

with similar expressions for the \( x \) and \( y \) components.

\textbf{Reference}

Appendix 3.B: Optical effects of free carriers.

Optical free carrier effects may be calculated by considering the equation of motion of a carrier under the influence of an optical electric field. The displacement of the carrier, in relation to the applied field, leads to an expression for the polarisation, the dielectric constant, and finally the refractive index change due to the presence of the carriers.

The equation of motion of a free carrier in a semiconductor is given by

\[ E_{q_e} - m^{*} \frac{d^{2}x}{dt^{2}} - \frac{q}{\mu} \frac{dx}{dt} = 0 \]  

(1)

where \( E_{q_e} \) is the force on the charge \( q_e \) due to the field \( E \); \( m^{*} \frac{d^{2}x}{dt^{2}} \) is the mass-acceleration, \( m^{*} \) being the effective mass and \( x \) being the displacement of the particle; \( \frac{q}{\mu} \frac{dx}{dt} \) is the resistive force on the charge, \( \mu \) being its mobility.

\[ E = E_0 e^{i\omega t} \]  

so that equation (1) has solutions of the form

\[ x = (x_0 + ix_1).\exp(i\omega t); \]  

(2)

differentiation with respect to time is thus equivalent to multiplication by \( i\omega \):

substituting into (1),

\[ E_0 q + \omega^2 m^* (x_0 + ix_1) - \frac{i\omega q}{\mu} (x_0 + ix_1) = 0 \]

Splitting this into real and imaginary components,

\[ E_0 q + \omega^2 m^* x_0 + \frac{\omega q}{\mu} x_1 = 0 \]  

(3)

\[ \omega^2 m^* x_1 - \frac{\omega q}{\mu} x_0 = 0 \]  

(4)

combining these,

\[ E_0 q + \omega^2 m^* \{ \frac{q}{\mu \omega m^*} \}^2 = 0 \]

203
q = 1.6\times 10^{-19}\text{C}; \ \omega \text{ is typically } 2\times 10^{15}\text{Hz}; \ m^* \text{ is of the order of } 5\times 10^{-31}\text{kg}; \text{ so that } \frac{q}{\mu \omega m^*} \text{ is of the order of } 10^{-2}, \text{ and can be neglected. Finally, therefore,}
\begin{equation}
\begin{align*}
\chi_0 &= \frac{E_0q}{\omega^2 m^*} \\
\text{and } \chi_1 &= \frac{q}{\mu \omega m^*} \chi_0 = \frac{E_0q^2}{\mu \omega^3 m^*^2}
\end{align*}
\end{equation}

The dielectric constant \( \varepsilon \) and the polarisation \( P \) are related by
\[ \varepsilon \varepsilon_0 E = \varepsilon_0 E + P, \text{ so } \varepsilon = 1 + \frac{P}{\varepsilon_0 E}. \]

The change \( \Delta \varepsilon \) in dielectric constant is therefore
\[ \Delta \varepsilon = \Delta \left( \frac{P}{\varepsilon_0 E} \right). \]

The refractive index is given by \( n = \varepsilon^2 \) so that the perturbation \( \Delta n \) is equal to
\[ \Delta n = \frac{Nq\chi}{2n\varepsilon_0 E}. \]

In this case, the polarisation is due to \( N \) charges per unit volume, and is given by
\[ P = Nq\chi. \]

Combining equations (5) to (9), the real and imaginary perturbations to the refractive index are then
\[ \Delta n_{(\text{real})} = - \frac{Nq^2\lambda_0^2}{8\pi^2 c^2 n_0 m^*} \]
\[ \Delta n_{(\text{imag})} = - \frac{Nq^3\lambda_0^3}{16\pi^3 c^3 n_0 m^*^2 \mu} \]

where \( c \) is the speed of light and \( \lambda_0 \) is the wavelength in vacuo. The attenuation coefficient, \( \alpha \), is related to the imaginary part of the index, \( n_i \), by
\[ \alpha = \frac{4\pi k_0 n_i}{\lambda} \quad \text{(intensity attenuation)}, \]

204
where $k_0$ is the free-space wavevector, and the loss in dB/cm is then equal to
\[ \alpha \cdot \log_{10}(e) = 4.34 \alpha. \]

These equations show that the effects are influenced by the carrier effective
mass, $m^*$, and the mobility $\mu$. In silicon, there are several carrier effective masses
depending on the crystallographic direction as well as the carrier type.
Appendix 4.A: RF sputtering

The term "sputtering" covers a range of processes whereby energetic particles impinge on a target and cause the ejection of material. "RF" refers to the Radio Frequency signal which is employed to maintain a glow discharge, whence the energetic species are derived. In sputter deposition, the ejected material is allowed to impinge on a substrate surface whereupon a thin film of material accumulates. The use of a high frequency signal, as opposed to a DC discharge, enables the technique to be applied to insulating targets, since RF power can be capacitatively through a thickness of dielectric material.

4A.1. RF Plasma

A glow discharge, or plasma, can be set up by establishing a high electric field in a low pressure gas. The field causes acceleration of ionised species and electrons: these gather energy, and lead to further ionisation by colliding with neutral species and with solid surfaces. If the ionisation rate is high enough, the process can become self-sustaining. The decay of excited species gives rise to a visible glow.

Although the ion density in a typical low pressure plasma is low (typically circa 1 ion per 10^4 neutrals), the plasma is electrically conducting because of the high mobility of free electrons. Thus there is essentially no electric field within the plasma itself. The positive ions, however, are much less mobile because of their larger mass. This, and the tendency of solid surfaces to capture electrons rather than emit them, gives rise to a rectifying behaviour between the plasma and solid surfaces. In an RF plasma, a dark space or "sheath" becomes visible in these regions, because the ionisation rate and ion density are reduced as electrons are swept away from the solid surface by the negative part of the electric field oscillation. This effect is most visible on the driven electrode, where the RF field amplitude is highest.

The behaviour of the plasma potential distribution may be understood by considering the equivalent circuit. The electron current is represented by a diode with a small resistor: the ion current by another diode with a larger resistor. Capacitors represent the effect of the electrode and plasma surfaces facing each other across the sheath.
Ignoring for a moment the ionic currents, the rectifying behaviour of the electron currents causes the plasma potential to rise to a DC level equal to that of the most positive electrode in the system, i.e. the RF electrode at its positive peak. The ionic current then allows the plasma potential to drop slightly below this level during the negative RF cycle. The RF electrode is driven via a capacitor, and is therefore at a floating DC potential: after a few cycles of operation, the plasma potential declines, due to ionic current to the chamber walls; in the steady state, the plasma is within a few volts of the chamber walls and the RF electrode has a negative DC bias such as to bring its positive cycle to within a few volts of the plasma potential. The plasma potential retains a small RF component, due to a combination of the ionic discharge during the negative RF cycle and the capacitative coupling to the RF electrode.

4A.2. Process control

The growth of a film by RF sputtering, or indeed by almost any technique, involves a combination of several processes taking place simultaneously: ion and electron bombardment heating, electrical charging, ion implantation, surface and subsurface diffusion, and so forth. Although each of these effects may be well understood, they are not generally under direct control independently of each other. Instead, they are adjusted indirectly via a number of control parameters, such as RF power, pressure, and substrate table heating. Optimum process conditions are determined empirically, although it is useful to understand the most important phenomena, along with their dependence on the control parameters and their impact on the film composition and structure.

Bombardment by sputtered species is the main mechanism for film growth, and is governed primarily by the target sputtering rate. To first order, the sputtering rate is given by the total RF power, divided by the average activation energy for removing a target atom. The flux of material is also affected by geometrical factors, such as the distance between the target and the substrate, and the orientations of each.

The flux of sputtered species reflects the target composition. Even if one constituent has a higher sputter yield than the others, the effect on the flux is only temporary, as it becomes redressed after a while by a corresponding depletion in the target surface.
Process pressures are generally in region of 2 to 40mTorr, implying a mean free paths of 2.5cm down to 1mm, rather smaller than the typical target-substrate distances of a few centimetres). Most of the sputtered species undergo collisions with gas molecules before reaching the growing surface, so that their energy spectrum reflects the plasma temperature which corresponds typically to a fraction of an electron volt. At low pressures, there is an increased likelihood that high energy particles, generated in the intense fields near the target electrode, can reach the substrate without intermediate collisions. Time-of-flight spectroscopy, in a system similar to the one used in this work, has shown that electrons, oxygen atoms, ions, and molecules with an energy equivalent to the target sheath voltage are present during sputter deposition, although no absolute quantitative data has been published[1,2]. Such bombardment of the growing film by high energy species leads to surface heating, sputter etching, and oxygen implantation. At these energies (a few hundred eV), typical implant depths are of the order of a hundred Ångstroms. Unless the substrate temperature is high enough to anneal the consequent damage, implanted species are likely to give rise to interstitial defects, either directly or by substitution. The severity of this effect can be reduced by reducing the fraction of oxygen in the sputter gas mixture, although the oxygen concentration in the film may begin to suffer if the oxygen pressure is reduced too far. Tominaga et al[1,2] examined DC and RF sputtered zinc oxide films by X-ray diffraction, and found that there was an optimum sputter gas mixture using about 20% oxygen, which is the same as used in the present work.

The growing film is also bombarded by low energy atoms, ions and electrons. The atomic flux can be estimated from thermodynamic considerations, and is approximately 5x10^{17}/cm^{2}sec at 10mTorr; in most cases, these atoms leave the surface before becoming bonded, or being occluded by new material. The electronic and ionic fluxes are governed by the field across the substrate sheath, and hence by the DC bias, as discussed above.

The main effect of low energy atomic and ionic bombardment by species from the ambient atmosphere is to alter the surface composition by a small amount. M. Petrucci measured sticking coefficients of low energy oxygen atoms and radicals in MBE growth of niobium oxides[3], and found figures of the order of 1% to 10%. These figures depend on the likelihood of bond formation, so that argon atoms are less likely to remain in the film than oxygen atoms: even after being occluded, these may subsequently diffuse to the surface and return to the plasma if the temperature is sufficiently high. Evidence consistent with this has been found in several studies[4], including the present work, in which refractive indices of sputtered films have been

208
observed to increase with deposition temperature, up to a saturation value where the film is almost totally outgassed: however, it should be noted that the observed refractive index increases could also be due to densification of the film, so that this evidence is not conclusive on its own.

The surface temperature governs surface atomic mobility, which is an important factor in determining the structure of the growing film. The probability of an atom reaching a lattice site depends on the density of available sites as well as on mobility: this provides a positive feedback mechanism which leads to an abrupt transition between amorphous and crystalline growth as the temperature increases. The crystalline growth direction depends on competition between crystal facets with different surface energies: the surface with the lowest energy tends to grow fastest and becomes dominant. This can be over-ridden if the initial substrate surface exerts a strong influence on seed crystallites, promoting epitaxy where the film grows with a lattice aligned to that of the substrate. The final outcome of these competing effects can be changed by altering the surface energies of the different growth facets, via any of a large range of process parameters. The texture of RF sputtered lithium niobate has been observed in several studies, including the present work, to depend on sputter pressure and composition, RF power, DC bias, as well as substrate material.

The surface temperature is equal (to first order) to the substrate table temperature, but may be enhanced by surface heating due to high energy electronic and ionic bombardment. The temperature enhancement may be calculated from a knowledge of the power density and the thermal conductivity of the substrate. If the power density is \( p \), the thickness of the sample is \( d \), and the thermal conductivity is \( K \), then the temperature enhancement \( \Delta T \) is given by

\[
\Delta T = \frac{p \cdot d}{K}
\]

Substituting typical values \( K = 1.7 \text{ W/mK} \) for fused silica at 300°C, power density \(< 10\text{ W} \text{ per} \text{ cm}^2 \) (i.e. 10% of the total RF power delivered to the target), \( d = 1\text{ mm} \), the change in temperature is less than 0.5°C, so that in most circumstances the effect is negligible.
References


Appendix 4.B: Geometry of X-ray diffraction

1. Reflections

Figure 4B.1 illustrates X-rays, wavelength $\lambda$, being reflected by a set of planes with spacing $d$. The deflection angle is $2\theta_B$, and these parameters are related by the well-known Bragg condition,

$$m\lambda = 2d \sin\theta.$$

Rewriting this in terms of the wavenumber $k = \frac{2\pi}{\lambda}$ and the reciprocal lattice distance $d^* = \frac{2\pi}{d}$,

$$md^* = 2k \sin\theta.$$

![Diagram of X-ray diffraction](image)

In figure 4B.1, let $OA = 2k$ so that $OB = 2k \sin\theta$. The direction of $OB$ is parallel to the normal to the set of crystal planes which are responsible for the reflection. Putting these two results together, the vector $OB$ is a reciprocal lattice vector of the crystal. Thus the X-ray reflections may be constructed by means of a reciprocal lattice, centred on $O$, and a sphere of radius $k$, centred on a point on the X-ray beam in such a way that the $O$ lies on the surface of the sphere: reflections occur whenever a reciprocal lattice point lies upon the surface of the sphere. This sphere will be henceforth be referred to as $S$.

Powder photographs may be interpreted by visualising the reciprocal lattice points being smeared out into spherical surfaces centred on the origin, since the plane
spacings are unchanged but are oriented in all possible directions. In the same manner, an agglomeration of crystallites with one oriented axis (e.g. c axes parallel but no correlation of a and b axes) has a reciprocal composed of circular rings, centred on the axis of common orientation. In the case of crystallites with only a preferred orientation, i.e. where the alignment is distributed about the common axis, the rings are slightly smeared out into spheroidal annuli, as shown in figure 4B.2. The X-ray photographs therefore shows these as short arcs. The lattice points which lie on or near the preferred axis tend to be the most intense, since they are smeared out over a smaller area than the rings with large radii.

![Diagram of preferred orientation and its effect on lattice points](image)

Note. Both reflections illustrated here lie on the same sphere in reciprocal space, and therefore correspond to the same lattice spacing: this coincidence was chosen for convenience and clarity.

Figure 4B.2. Illustrating the arcs resulting from polycrystalline material with a preferred orientation of one axis. The arcs are formed by the intersection of the smeared reciprocal lattice points with the sphere S.
Appendix 4.C: X-ray diffraction photographs of rhombohedral lithium niobate

4C.1. Indexing

As a result of the crystal symmetry and the lack of defined x and y axes in polycrystalline material with a preferred z-orientation, most of the features in figure 4.15 and in the powder photograph below contain contributions from several reflections. Only one index is quoted in each case, but the indices of the equivalent reflections may be derived as follows[1].

It is helpful to rewrite an index \(<h k l>\) as \(<h k i l>\), where \(i\) is obtained from \(h\) and \(k\) such that \(h + k + i = 0\). Then \(h\), \(k\), and \(i\) may be interchanged cyclically to generate the equivalent reflections, i.e. \(<k i h l>\) and \(<i h k l>\), which may be abbreviated to \(<k i l>\) and \(<i h l>\) respectively. Thus \(<0 1 2>\) has equivalents \(<1 0 2>\) \& \(<1 1 0>\), \(<1 1 6>\) has equivalents \(<2 1 6>\) \& \(<1 2 6>\) ad so forth.

4C.2. Powder photograph

The photographs in figures 4.15 and 4.16 were taken using unfiltered Cu radiation, so that both Cu K\(\alpha\) and K\(\beta\) lines \((\lambda = 1.541\text{Å} \& 1.392\text{Å} \text{respectively})\) and underlying white radiation contribute to the patterns[1]. The powder photograph in figure 4C.1, below, was taken under the same conditions. Indexing is given in the accompanying key, for lines up to \(<2 0 8>\) (\(\theta_B = 34.25^\circ\)).

4C.3. Rotation photographs

The photographs in figure 4C.2 were taken using filtered radiation, so that only the Cu K\(\alpha\) reflections \((\lambda = 1.541\text{Å})\) are present.

Reference

Note. only the stronger $\beta$ reflections are shown.

Figure 4C.1. Powder photograph of rhombohedral lithium niobate.
Figure 4C.2. Rotation photographs of rhombohedral lithium niobate.
Appendix 5A: Ion beam milling

The ion beam milling apparatus comprises an arc chamber, a set of acceleration grids, and a substrate table assembly in a vacuum chamber. This is shown schematically in figure 5A.1.

The system is evacuated, and a small amount of argon is bled into the arc chamber. The arc chamber pressure is kept constant by means of a feedback loop, monitoring an ionisation pressure gauge and controlling a variable bleed valve. A tungsten filament, in the centre of the chamber, serves as an electron source and cathode. Thermionic electrons are accelerated towards a surrounding anode, ionising argon atoms by collision. A magnetic field is maintained in the arc chamber, to lengthen the electron paths and enhance the ionisation rate.

Argon ions diffuse through a grid in the floor of the arc chamber. Directly below the first grid, there is a second grid which is used to set up an electric field, accelerating ions downwards towards the substrate table in the main vacuum chamber. The second grid is at earth potential, along with the substrate table and the remainder of the vacuum system: the accelerating field is set up by maintaining the arc chamber at a large positive potential, between 500 and 2000V. Arc and filament power supplies are isolated and floating; remote controls enable adjustments to be made by an operator at earth potential.

Below the grids, a filament emits electrons to neutralise the ion beam. There is a third grid, above this filament, at a positive potential (typically 200V) with respect to earth, to prevent electrons from back-streaming through the accelerating field into the arc chamber and causing damage.

The ion beam current density is monitored by means of a Faraday cup incorporated into a shield above the substrates. After stable operating conditions have been established, the shield is withdrawn to allow the ion beam to impinge on the substrates. The substrate table is water cooled, and rotates to minimise non-uniformity and hot spots.

The milling mechanism is by sputtering: although reactive species may be introduced either into the arc chamber or into the vicinity of the substrates, the present work (processing lithium niobate, silicon, and sapphire) was carried out using only argon.
Figure 5A.1. The ion beam milling apparatus.
Appendix 6.A: Calculation of the electro-optic effect in <0 1 2> oriented lithium niobate

The electric field is applied along the <012> direction, which lies in the symmetry plane defined by the y* and z* axes (see figure 6.1). By symmetry, it follows that the effect of the electric field cannot tilt the refractive index ellipsoid about the y* axis. Transforming to axes x', y', z' as illustrated in figure 6.1, the index ellipsoid should reduce to the form

\[ b_1 x'^2 + b_2 y'^2 + b_3 z'^2 + 2b_5 x'z'^2 - 1 = 0. \] (1)

The transformation is effected by putting

\[ x = \frac{\sqrt{3}}{2} x' - \frac{1}{2} y', \quad y = \frac{1}{2} x' + \frac{\sqrt{3}}{2} y', \quad z = z' \] (2)

so that

\[ x^2 = \frac{3}{4} x'^2 - \frac{\sqrt{3}}{2} x'y' + \frac{1}{4} y'^2, \] (3a)
\[ y^2 = \frac{1}{4} x'^2 + \frac{\sqrt{3}}{2} x'y' + \frac{3}{4} y'^2, \] (3b)
\[ z^2 = z'^2 \] (3c)
\[ yz = \frac{1}{2} x'z' + \frac{\sqrt{3}}{2} y'z', \] (3d)
\[ zx = \frac{\sqrt{3}}{2} x'z' - \frac{1}{2} y'z', \] (3e)
\[ xy = \frac{\sqrt{3}}{4} x'^2 + \frac{1}{2} x'y' - \frac{\sqrt{3}}{4} y'^2 \] (3f)

The index ellipsoid, related to the orthorhombic axes, was given in equation (3) of Appendix 3A. Using this transformation, the ellipsoid becomes

\[ x'^2 \left\{ \frac{3}{4} b_1 + \frac{1}{4} b_2 + \frac{\sqrt{3}}{2} b_6 \right\} \]
\[ + \quad y'^2 \left\{ \frac{1}{4} b_1 + \frac{3}{4} b_2 - \frac{\sqrt{3}}{2} b_6 \right\} \]
\[ + \quad z'^2 b_3 \]
\[ + z'y' \left\{ - \sqrt{3} \ b_4 - b_5 \right\} \\
+ z'x' \left\{ b_4 + \sqrt{3} b_5 \right\} \\
+ x'y' \left\{ - \frac{\sqrt{3}}{2} b_1 + \frac{\sqrt{3}}{2} b_2 + b_6 \right\} \\
- 1 = 0. \tag{4} \]

This equation is identical, in form, to the original (equation (3) of Appendix 3A), and coefficients \( b_1', b_2', \) etc. can be defined using an obvious scheme.

In the new axes, the electric field components are \( E_x', E_y', E_z' \); in this case, \( E_y' = 0 \). Using the electro-optic tensor set out in that appendix, and applying the transformation defined in equation (2) to the electric field components, the \( b \) coefficients are as follows:

\[ b_1 = \frac{1}{n_0^2} - \frac{1}{2} r_{22} E_x' + r_{13} E_z' \tag{5a} \]
\[ b_2 = \frac{1}{n_0^2} + \frac{1}{2} r_{22} E_x' + r_{13} E_z' \tag{5b} \]
\[ b_3 = \frac{1}{n_e^2} + r_{33} E_z' \tag{5c} \]
\[ b_4 = \frac{1}{2} r_{51} E_x' \tag{5d} \]
\[ b_5 = \frac{\sqrt{3}}{2} r_{51} E_x' \tag{5e} \]
\[ b_6 = - \frac{\sqrt{3}}{2} r_{22} E_x' \tag{5f} \]

Substituting equations (5a-f) in equation (4), the coefficients \( b_1', b_2', \) etc. of the new ellipsoid are as follows:

\[ b_1' = \frac{1}{n_0^2} + r_{13} E_z' - r_{22} E_x' \tag{6a} \]
\[ b_2' = \frac{1}{n_0^2} + r_{13} E_z' + r_{22} E_x' \tag{6b} \]
\[ b_3' = \frac{1}{n_e^2} + r_{33} E_z' \tag{6c} \]
\[ b_4' = 0 \tag{6d} \]

219
as predicted by the symmetry arguments, so that equation (1) does indeed represent the transformed ellipsoid. This ellipsoid will have principal axes along the y' axis and along

\[
p_1 = \begin{pmatrix} \cos \theta \\ 0 \\ \sin \theta \end{pmatrix}, \quad p_3 = \begin{pmatrix} -\sin \theta \\ 0 \\ \cos \theta \end{pmatrix}
\]

with corresponding refractive indices \( n_1 \) and \( n_3 \), and where \( \theta \) is as calculated below. Transforming to these new axes,

\[
x' = x'' \cos \theta - z'' \sin \theta \quad (7a)
\]
\[
z' = x'' \sin \theta + z'' \cos \theta \quad (7b)
\]
\[
x'^2 = x''^2 \cos^2 \theta - 2x''z'' \sin \theta \cos \theta + z''^2 \sin^2 \theta \quad (7c)
\]
\[
z'^2 = x''^2 \sin^2 \theta + 2x''z'' \sin \theta \cos \theta + z''^2 \cos^2 \theta \quad (7d)
\]
\[
x'z' = x''^2 \sin \theta \cos \theta - 2x''z'' \{ \cos^2 \theta - \sin^2 \theta \} + z''^2 \sin \theta \cos \theta \quad (7e)
\]

Substituting into equation (1), putting the \( x''z'' \) term equal to zero, and using well known trigonometric identities,

\[
\tan 2\theta = \frac{2b_5'}{b_1' - b_3'} \quad (8)
\]

while the \( x''^2 \) and \( z''^2 \) terms yield values for the indices \( n_1 \) and \( n_3 \): using equations (6a - f),

\[
\frac{1}{n_1^2} = \frac{1}{n_0^2} \cos^2 \theta + \frac{1}{n_e^2} \sin^2 \theta
\]
\[
+ ( r_{13} \cos^2 \theta + r_{33} \sin^2 \theta ) E_z'
\]
\[
+ (- r_{22} \cos^2 \theta + r_{51} \sin 2\theta ) E_x'
\]

and

\[
\frac{1}{n_3^2} = \frac{1}{n_0^2} \sin^2 \theta + \frac{1}{n_e^2} \cos^2 \theta
\]
\[
+ ( r_{13} \sin^2 \theta + r_{33} \cos^2 \theta ) E_z'
\]
\[ + ( - r_{22} \sin^2 \theta - r_{51} \sin 2\theta ) E_x' \]  

(10)

\( \theta \) is small for practical situations. Neglecting 2\textsuperscript{nd} and higher orders in the electric field, equations (8-10) reduce to

\[
\theta = \frac{r_{51} E_x'}{1/n_0^2 - 1/n_o^2} \tag{11}
\]

\[
\frac{1}{n_1^2} = \frac{1}{n_o^2} + r_{13} E_z' - r_{22} E_x' \tag{12}
\]

\[
\frac{1}{n_3^2} = \frac{1}{n_e^2} + r_{33} E_z' \tag{13}
\]

while equation (6b) yields

\[
\frac{1}{n_2^2} = \frac{1}{n_o^2} + r_{13} E_z' + r_{22} E_x' \tag{14}
\]

Finally, these equations may be used to calculate changes in the index due to an electric field \( E_o \) along \( <012> \), by writing \( E_x' = E_o \sin \theta_{012} \) and \( E_z' = E_o \cos \theta_{012} \), where \( \theta_{012} \) is the angle subtended by the \( z \) axis and the \( <012> \) direction.