Integrated Optical Encoder

John Peter Carr

Submitted for the degree of Doctor of Philosophy

Heriot-Watt University

School of Engineering and Physical Sciences

January 2010

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that the copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without the prior written consent of the author or of the University (as may be appropriate).
Abstract

This thesis presents the work undertaken by the author towards his Ph.D. qualification at the Microsystems Engineering Centre (MISEC) in the School of Engineering and Physical Sciences of Heriot-Watt University, Edinburgh.

The work concentrates on the design, assembly and test of a novel miniaturised and monolithically integrated optical encoder for use in a high precision measurement application. Utilising microsystems manufacturing techniques the traditional components of the encoder are monolithically integrated onto a single compound semiconductor chip radically reducing the footprint and cost of assembly. Fabrication of the gratings at the wafer level, by standard photolithography, allows for the simultaneous alignment of many devices in a single process step. This development coupled with a new unique photodiode configuration has demonstrated increased performance and has significantly improved the alignment tolerances in both manufacture and set-up, which substantially reduce costs and required set-up times.

The work presented centres on a surface emitting optical encoder chip designed to be compatible with both flip-chip bonding and wire bonding assembly technologies. Feasibility of the integrated optical encoder concept has been achieved using wire bonded devices. Integrated optical encoder chips are successfully demonstrated against both amplitude and phase scale gratings. When operated in photoconductive mode and with differential amplification circuitry providing a gain of 33, DC to AC ratios of the order of 7:1, signal-to-noise ratios greater than 60:1 and Lissajous curves having peak-to-peak voltages reaching the operating limits of the test set-up have been achieved.

The results mean that an optical encoder of an order of magnitude cheaper and smaller than existing sub-10nm products can now be produced and commercialised by the company Renishaw plc, which sponsored this project.
Acknowledgments

First I would like to thank Professor Marc Desmulliez for presenting me with this opportunity and for his guidance, support and encouragement over the course of this research.

I would also like to thank the other members of the MISEC group at Heriot-Watt University who allowed me to convey ideas and for any assistance they may have provided over the years. In particular I would like to thank Mark Leonard for his invaluable tuition and continued assistance with the cleanroom facilities.

Thanks also go to the industrial arm of the project, Renishaw plc, for having the confidence to invest in the project. In particular I would like to thanks Dr Nick Weston and Dr David McKendrick for their contributions to the project and to this thesis. I reserve thanks for Graeme Cunningham for his assistance and contribution to the measurements presented in this work and for his efforts developing the electronics. Furthermore I would like to thank those at Compound Semiconductor Technologies not only for their involvement with the chip design and its manufacture but also for allowing access to their equipment and taking the time out to accommodate my visits. I also acknowledge the efforts of CST for their measurements that contribute to work presented in this document.

Finally I would like to thank my wife Verity and my parents for their never-ending support and understanding throughout without which I could not have completed this work.
**ACADEMIC REGISTRY**

**Research Thesis Submission**

<table>
<thead>
<tr>
<th>Name:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>School/PGI:</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Version: (i.e. First, Resubmission, Final)</th>
<th>Degree Sought (Award and Subject area)</th>
</tr>
</thead>
</table>

**Declaration**

In accordance with the appropriate regulations I hereby submit my thesis and I declare that:

1) the thesis embodies the results of my own work and has been composed by myself
2) where appropriate, I have made acknowledgement of the work of others and have made reference to work carried out in collaboration with other persons
3) the thesis is the correct version of the thesis for submission and is the same version as any electronic versions submitted*.
4) my thesis for the award referred to, deposited in the Heriot-Watt University Library, should be made available for loan or photocopying and be available via the Institutional Repository, subject to such conditions as the Librarian may require
5) I understand that as a student of the University I am required to abide by the Regulations of the University and to conform to its discipline.

* Please note that it is the responsibility of the candidate to ensure that the correct version of the thesis is submitted.

<table>
<thead>
<tr>
<th>Signature of Candidate:</th>
<th>Date:</th>
</tr>
</thead>
</table>

**Submission**

Submitted By (name in capitals):

Signature of Individual Submitting:

Date Submitted:

**For Completion in Academic Registry**

Received in the Academic Registry by (name in capitals):

Method of Submission
(Handed in to Academic Registry; posted through internal/external mail):

E-thesis Submitted (mandatory for final theses from January 2009)

Signature: Date:
Table of Contents

Chapter 1
Introduction.......................................................................................................................1

1.1 Position Transducers and Encoders ..............................................................2
  1.1.1 Electromechanical transducers............................................................2
  1.1.2 Induction Transducers .......................................................................4
  1.1.3 Magnetic Encoders ............................................................................6
  1.1.4 Optical encoders ...............................................................................7

1.2 Rationale for Miniaturisation ........................................................................13

1.3 Thesis Layout................................................................................................15

1.4 References.....................................................................................................18

Chapter 2
Optical encoder overview .......................................................................................22

2.1 Optical encoder basics ..................................................................................22
  2.1.1 Incremental encoders .......................................................................24
  2.1.2 Absolute encoders ...........................................................................27

2.2 Measuring methods .......................................................................................29
  2.2.1 Imaging Principle .............................................................................30
  2.2.2 Interference principle .......................................................................31

2.3 Present Renishaw optical encoder ..............................................................33
  2.3.1 Present Optical encoder readhead assembly .....................................38

2.4 A brief summary of competing products ....................................................42

2.5 Summary .......................................................................................................43

2.6 References.....................................................................................................44

Chapter 3
Literature Review – Assembly and Packaging Techniques.................................49

3.1 Integration ......................................................................................................49
  3.1.1 Monolithic integration .....................................................................49
  3.1.2 Hybrid assembly ..............................................................................50
  3.1.3 Integration as applied to optical encoders .......................................52

3.2 Assembly .......................................................................................................60
  3.2.1 Wire Bonding ..................................................................................60
  3.2.2 Flip-chip bonding ...........................................................................61
  3.2.3 Failure mechanisms of interconnect technologies .........................62
  3.2.4 Packaging solutions for flip-chip applications ...............................69
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.3</td>
<td>Summary</td>
<td>73</td>
</tr>
<tr>
<td>3.4</td>
<td>References</td>
<td>73</td>
</tr>
<tr>
<td>4.1</td>
<td>Package design concept</td>
<td>79</td>
</tr>
<tr>
<td>4.2</td>
<td>Monolithically Integrated Optical Encoder Chip</td>
<td>81</td>
</tr>
<tr>
<td>4.3</td>
<td>LED and Detector Layouts</td>
<td>84</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Pixel Layout</td>
<td>87</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Chevron layout</td>
<td>90</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Reference markers</td>
<td>92</td>
</tr>
<tr>
<td>4.4</td>
<td>Summary</td>
<td>92</td>
</tr>
<tr>
<td>4.5</td>
<td>References</td>
<td>93</td>
</tr>
<tr>
<td>5.1</td>
<td>Technology selection</td>
<td>94</td>
</tr>
<tr>
<td>5.2</td>
<td>Electroplating basics</td>
<td>98</td>
</tr>
<tr>
<td>5.3</td>
<td>Gold bath compositions</td>
<td>101</td>
</tr>
<tr>
<td>5.4</td>
<td>Bump fabrication process</td>
<td>103</td>
</tr>
<tr>
<td>5.4.1</td>
<td>Substrate preparation</td>
<td>104</td>
</tr>
<tr>
<td>5.4.2</td>
<td>Seed layer deposition</td>
<td>105</td>
</tr>
<tr>
<td>5.4.3</td>
<td>Photoresist spin coating</td>
<td>107</td>
</tr>
<tr>
<td>5.4.4</td>
<td>Exposure and development of AZ and THB photoresist</td>
<td>110</td>
</tr>
<tr>
<td>5.4.5</td>
<td>Etching of the upper seed layer metal</td>
<td>111</td>
</tr>
<tr>
<td>5.4.6</td>
<td>Electroplating</td>
<td>112</td>
</tr>
<tr>
<td>5.5</td>
<td>Summary</td>
<td>120</td>
</tr>
<tr>
<td>5.6</td>
<td>References</td>
<td>121</td>
</tr>
<tr>
<td>6.1</td>
<td>Preliminary bonding</td>
<td>125</td>
</tr>
<tr>
<td>6.2</td>
<td>Flip-chip pick-up tool modifications</td>
<td>127</td>
</tr>
<tr>
<td>6.3</td>
<td>Finite element analysis model development</td>
<td>129</td>
</tr>
<tr>
<td>6.4</td>
<td>Au to Au thermocompression bonding</td>
<td>134</td>
</tr>
<tr>
<td>6.4.1</td>
<td>Bump shear strength</td>
<td>134</td>
</tr>
<tr>
<td>6.4.2</td>
<td>Flip-chip bonding</td>
<td>135</td>
</tr>
<tr>
<td>6.5</td>
<td>Summary</td>
<td>141</td>
</tr>
</tbody>
</table>
6.6 References ........................................................................................................... 142

Chapter 7

Wire-bonded surface emitting LED optical encoder chip .............................................. 144

7.1 Chip on glass carrier ........................................................................................... 146

7.1.1 LED response ............................................................................................... 146

7.1.2 Photo detector response ............................................................................... 150

7.1.3 Thermal probe ............................................................................................. 152

7.1.4 Functional test .............................................................................................. 153

7.1.5 Summary ....................................................................................................... 157

7.2 Silicon submount ................................................................................................. 157

7.2.1 Thermal probe .............................................................................................. 158

7.2.2 Response ....................................................................................................... 159

7.2.3 Functional test .............................................................................................. 160

7.2.4 Chip on silicon carrier summary ..................................................................... 166

Chapter 8

Design and Characterisation of the Reference Marker ............................................... 167

8.1 Comparison of the incremental encoder chips .................................................... 167

8.1.1 Set-up ........................................................................................................... 167

8.1.2 Grating pitch variation ................................................................................ 169

8.2 Investigation of the reference mark schemes ...................................................... 172

8.2.1 Chip reference marker configurations .......................................................... 174

8.3 Chip reference marker results ............................................................................. 182

8.3.1 Z-type chip .................................................................................................... 182

8.3.2 A-type chip ................................................................................................... 184

8.3.3 A2-type chip ................................................................................................. 186

8.3.4 B-type chip ................................................................................................... 188

8.3.5 C-type chip ................................................................................................... 190

8.3.6 D-type chip ................................................................................................... 191

8.4 Additional observations ...................................................................................... 193

8.5 Summary ............................................................................................................. 195

Chapter 9

Conclusions and Future Work ....................................................................................... 198

9.1 Review summary ................................................................................................. 198

9.1.1 Integrated encoder chip – design and development ........................................ 198

9.1.2 Flip-chip bonding process development ....................................................... 200

vi
9.2 Future work .........................................................................................................201
  9.2.1 Reference mark and chip design development .............................................201
  9.2.2 Flip-chip bonding development ...................................................................204
  9.2.3 Advanced Packaging ....................................................................................205
  9.2.4 Further development ....................................................................................207
9.3 Concluding statements ........................................................................................208

Appendices
  Appendix A ...............................................................................................................209
  Edge emitting communications laser .................................................................209
  Appendix B ...............................................................................................................212
  Substrate emitting optical encoder ......................................................................212
  Appendix C ...............................................................................................................215
  Effect of substrate thickness .................................................................................215
  Appendix D ...............................................................................................................218
  Functional test of flip-chip bonded substrate emitting devices .........................218
    D1 Electrical measurements ...............................................................................218
    D2 Optical measurements .................................................................................219
  Appendix E ...............................................................................................................221
  Au capped Cu bumps ............................................................................................221
  Appendix F ...............................................................................................................224
  Flip-chip bonding of Cu core bumps ....................................................................224
  Appendix G ...............................................................................................................226
  Flip chip bonding parameters ..............................................................................226
List of Figures

Figure 1.1 – Circuit diagram of a simpler potentiometer (voltage divider rule) ........................................3
Figure 1.2 – Illustration of brush contact absolute encoder ........................................................................3
Figure 1.3 – Illustration incremental brush encoder contact with digitised quadrature output ..................4
Figure 1.4 – Illustrative example of (a) Vernier fringe pattern and (b) Moiré fringe pattern ...................9
Figure 1.5 – Illustration of the three channel Heidenhain optical encoder based on the interferential measuring principle to produce three sinusoidal signals shifted by 120° ..................11
Figure 1.6 – Schematic of the thesis layout .................................................................................................17
Figure 2.1 – Basic optical encoder operating principles showing: transmission mode, top, where light passes straight through from the source to the detector and below, reflection mode, where light reflects from the scale onto the detector .............................................................23
Figure 2.2 – Illustration of an incremental encoder pattern (a) on linear scale, (b) rotary scale and (c) the digital output signal [2.3] ........................................................................................................................................25
Figure 2.3 – Square-wave output signals of a two channel (quadrature) incremental encoder indicating detection of rotation as well as angular position. The electrical signals are 90 degrees out of phase and relate to the direction-clockwise (CW) if Channel A leads Channel B and counter clockwise (CCW) if B leads A ........................................................................................................................................26
Figure 2.4 – Examples of Lissajous curves having (a) frequency and amplitude ratio’s of 1 and a phase difference of 90°, (b) where the phase shift is modified to 0°, (c) where the amplitude ratio does not equal 1, and (d) where the frequency ratio is 3:1 ........................................................127
Figure 2.5 – Illustration of an absolute encoder pattern on a rotary scale with representative output pulses [2.3] ........................................................................................................................................28
Figure 2.6 – Examples of binary coded patterns (a) Pure binary code and (b) Gray code ..........................29
Figure 2.7 – Illustration of the imaging principle for a transmission type optical encoder setup ..........30
Figure 2.8 – Illustration showing the path of two light beams that interfere at the analyser plane .........31
Figure 2.9 – Illustration interference principle showing how positive and negative orders of diffraction can be used to expand the scanning area ..................................................................................33
Figure 2.10 – Illustration of the three grating reflective system used with current Renishaw optical encoder ........................................................................................................................................34
Figure 2.11 – Illustration of the basic optical path and stand off distances ..............................................35
Figure 2.12 – Illustration of diffracted rays from two slits converging on a single point at the analyser plane ........................................................................................................................................37
Figure 2.13 – Illustration of the Moiré fringe pattern with (a) 0° and (b) 90° phase shift ......................37
Figure 2.14 – Present Renishaw optical encoder readhead (a) fully packaged, (b) with metal case removed showing additional electronics and (c) an enlarged view of the cable connection and interconnects ........................................................................................................................................38
Figure 2.15 – Schematic view of discrete component layout for present Renishaw optical encoder .......39
Figure 2.16 – Four element array demonstrating light distribution and phase combination for a parallel photo detector arrangement ........................................................................................................................................39
Figure 2.17 – Images showing (a) the optical chip and grating assembled to PCB and (b) an enlarged view of the area at the chips edge highlighting the wire bonds encapsulated in epoxy .......... 40
Figure 2.18 – Images showing (a) the cavity in the PCB, exposing the index and reference gratings that are illuminated by the LED and (b) a magnified view of figure 2.13(b) showing the epoxy encapsulated LED.

Figure 3.1 – Motorola two chip accelerometer highlighting hybrid integration [3.3].

Figure 3.2 – Multi Chip Module (MCM) example (a) HDI (b) Chip On Flex (COF) both showing ablated windows for MEMS access.

Figure 3.3 – Illustration of the Holzapfel optical encoder with monolithically integrated gratings and detectors [3.11].

Figure 3.4 – Cross section schematic of micromachined photodetector arrangement.

Figure 3.5 – Isometric illustration of monolithically integrated optical encoder by Sawada et al. [3.12].

Figure 3.6 – Illustration of the Numerik Jena Chip-on-Glass encoder technology.

Figure 3.7 – Photodetector array layout positioned around central light sources. Also shown are reference detectors and their associated light sources.

Figure 3.8 – Cross section schematics showing (a) the optical chip and (b) the optical chip flip-chip bonded to a glass substrate for a COG package solution.

Figure 3.9 – Images of integrated 25μm diameter mesa VCSELs and photo detector position sensor arrangements produced at the University of Illinois [3.18].

Figure 3.10 – Wire bonding principle showing (a) an cross-section illustration of a chip wire bonded to an IC board; note that a globular encapsulant (usually epoxy based) is sometimes employed over the chip and wires for added protection and reliability, and (b) SEM photo of multiple wires bonded in fine 50μm pitch [3.21].

Figure 3.11 – Flip-chip bonding principle (a) illustration of chip flip-chip bonded on to IC module (chip usually face down) and (b) cross-section image of flip-chip bumped device [3.23].

Figure 3.12 – An example of wire bond lift off / peel failure mechanisms [3.22].

Figure 3.13 – Flip-chip failure mechanisms (a) large void (b) bump with crack in the passivation and (c) wasted Ni particle as not enough pressure has been applied to create the bond [3.6].

Figure 3.14 – Example of wafer level chip scale packaging from Tessera [3.33].

Figure 3.15 – ShellOP wafer scale packaging with BGA.

Figure 3.16 – ShellOP process step showing pad extensions and T-contacts.

Figure 3.17 – Rerouting of contacts using backside plasma etching to form via holes [3.27].

Figure 3.18 – Example of optical cap assembly.

Figure 4.1 – Illustration of packaging solutions employing (a) FCoG technology and (b) wire bonding technology.

Figure 4.2 – Schematic cross-section of the surface emitting chip (not to scale) on an InP substrate for the monolithic integration of the optical encoder components.

Figure 4.3 – Parallel detector layout for (a) wire bonding and (b) for flip chip bonding. Note gratings are not shown here.

Figure 4.4 – Set-up example for amplitude error from a single photo detector failure.

Figure 4.5 – Position error as a function of a photodiode not contributing to the total signal over several ABCD repeats. Note normalised to one Lissajous or 0.5 of the pitch.

Figure 4.6 – Optical encoder chip pixel layout design (note gratings are not shown here).
Figure 6.1 – Illustration of thermocompression flip-chip bonding process .........................................................125
Figure 6.2 – Karl Suss FC6 thermocompression bonder ..........................................................................................127
Figure 6.3 – Image showing (a) the damage to chip caused by pick up tool during the bonding process and (b) a magnified image........................................................................................................................................................................127
Figure 6.4 – Illustration of original flip-chip pick up tool (a) Isometric illustration and (b) a schematic cross section illustration during bonding ........................................................................................................................................................................128
Figure 6.5 – Illustration of the new flip-chip pick up tool (a) Isometric illustration and (b) a schematic cross section illustration during bonding ........................................................................................................................................................................128
Figure 6.6 – Zygo measurements of pick up tool showing (a) deformation surface profile (b) top view contour plot and (c) an isometric 3D representation ........................................................................................................................................................................129
Figure 6.7 – Geometry of FCoG design to be used in ANSYS FEM ........................................................................130
Figure 6.8 – Simulated bonding temperature profiles ..................................................................................................131
Figure 6.9 – Temperature distribution at the end of the temperature cycle shown in figure 6.8 for (a) isotropic view and (b) magnified side elevation ........................................................................................................................................................................131
Figure 6.10 – Temperature distribution through glass submount and bumps with no chip ........................................................................................................................................................................132
Figure 6.11 – Simulated bonding force profile ..............................................................................................................133
Figure 6.12 – Deformation as a result of thermal expansion and applied bonding force ........................................133
Figure 6.13 – Images of the modelled bump deformation (a) before and (b) after the applied thermal and force cycles ........................................................................................................................................................................134
Figure 6.14 – Individual bump on glass shear measurements for short (20µm) and tall (24µm) bumps.134
Figure 6.15 – Generic example of the temperature and force parameter controls as a function of time, as input into the Karl Suss FC6 flip-chip bonder ........................................................................................................................................................................135
Figure 6.16 – Scatter chart of shear strength values for applied load and load duration ........................................137
Figure 6.17 – Optical images of two chips after shear testing showing varying degree of interconnection ........................................................................................................................................................................138
Figure 6.18 – Chip bond pad images after shear testing after applied pressures of (a) 35g/bump pre planarity correction, (b) 25 g/bump, (c) 35g/bump and (d) 45g/bump ........................................................................................................................................................................138
Figure 6.19 – Zygo measurements showing (a) 3D interpretation and (b) surface profile plots ..................................139
Figure 6.20 – Illustration of current crowding effect in electrodeposition ....................................................................140
Figure 6.21 – Image showing (a) a flip-chip bonded encoder chip on glass submount and (b) bump cross section........................................................................................................................................................................141
Figure 7.1 – Images showing a surface emitting optical encoder chip wire bonded to (a) glass carrier having two layer metallisation and (b) to a Si carrier having only a single layer of metal ...............................................................................................................................144
Figure 7.2 – Delamination of the dielectric layer ........................................................................................................145
Figure 7.3 – Dual metallisation process flow starting with (a) the first level metal deposition, (b) resist deposition, (c) resist mask patterning, (d) selective etching (e) resist strip and dielectric application, (f) via opening, (g) second level metallisation, (h) second resist deposition, (i) resist mask patterning for second level, (j1) selective etching of the Au layer for electroplating preparation and (j2) selective etching of the second level metal to the dielectric for singulation preparation and wire bonding ........................................................................................................................................................................145
Figure 7.4 – LED configurations having (a) circular mesa with non-bridged connection, (b) lozenge shaped mesa with non-bridged contacts, (c) circular mesa with bridged connection and (d) lozenge shaped mesa with bridged contacts .................................................................147

Figure 7.5 – Plots for (a) L-I curves for lozenge and circle LEDs and (b) I-V curves of LED output measured by monitor diode where the circled area suggests large photo currents of 50mA implying a 50mW LED output .............................................................................148

Figure 7.6 – Schematic representation of (a) the ideal optical encoder chip and (b) highlighting observed transistor like effects and poor isolation between the LED and detector sections ..........149

Figure 7.7 – Representation of the optical encoder chip connected in photoconductive mode with a reverse bias voltage of 2V applied to the photo-detectors .........................................................150

Figure 7.8 – Single channel photo detector response when illuminated with an incandescent light source at 150mm ........................................................................................................150

Figure 7.9 – Measured response from a single photo detector channel after the integrated LED is switched on with 5V, 200mA drive conditions .....................................................................151

Figure 7.10 – Response from a single photo detector channel periodically illuminated by an incandescent lamp while the integrated LED is switched on then off ........................................151

Figure 7.11 – Chip on glass surface temperature measurements; (a) temperature probe locations and (b) the change in measured surface temperature at the indicated locations for increasing LED drive current ........................................................................................................153

Figure 7.12 – Integrated optical encoder chip test rig with the amplitude scale grating ........................................154

Figure 7.13 – Maximum voltage peak-to-peak Lissajous curve observed for circuit connected as described in figure 7.7 with 1MΩ resistor ........................................................................155

Figure 7.14 – Representation of the optical encoder chip connected in photoconductive mode with a reverse bias voltage of 2V applied to the photo-detectors, 1MΩ resistors and shorting of the monitor diode ........................................................................................................155

Figure 7.15 – Maximum voltage peak-to-peak Lissajous curve observed for circuit connected as described in figure 7.14 ........................................................................................................156

Figure 7.16 – Rotational axis for mechanical alignment showing (a) yaw, (b) roll and (c) pitch ................................................................156

Figure 7.17 – Comparison of chip surface temperature for increasing LED drive current ........................................158

Figure 7.18 – Response from a single photo detector channel periodically illuminated by an incandescent lamp while the integrated LED is switched on then off ........................................159

Figure 7.19 – Encoder chip on silicon showing an example of a maximum peak-to-peak voltage Lissajous curve with bandwidth limited signals (a) before saturation and (b) at the onset of saturation ........................................................................................................160

Figure 7.20 – Representation of the optical encoder chip connected in photoconductive mode with a reverse bias voltage of -2V applied to the photo-detectors, 100kΩ resistors and shorting of the monitor diode ........................................................................................................161

Figure 7.21 – 1.5Vpp signal Lissajous curve observed when the encoder chip is connected as shown in figure 7.20 ........................................................................................................161

Figure 7.22 – Signal-to-Noise Ratio contour plots for (a) lozenge shaped LED with non-bridged contacts, (b) lozenge shaped LED with bridged contacts, (c) circular LED with non-bridged contacts and (d) circular LED with bridged contacts ................................................................163
List of Tables

Table 3.1: Under Bump Metallurgy for various flip-chip processes [3.21]……………………………….69
Table 4.1: Surface emitting epitaxial structure………………………………………………………….83
Table 4.2: Advantages and disadvantages of substrate emitting and surface emitting optical encoder chip designs……………………………………………………………………………………..84
Table 4.3: Non-optimised simplified power budget requirement for 650nA per channel……………….89
Table 5.1: The advantages and disadvantages of common bump technologies………………………..95
Table 5.2: 30μm spin coat parameters for AZ9260 and THB 151N…………………………………….109
Table 5.3: 30μm resist exposure dosage and development times………………………………………..111
Table 5.4: ECF 60 gold electrolyte suggested plating ranges………………………………………….113
Table 6.1 FEA model material thermal and structural properties……………………………………….130
Table 6.2: Bond parameters and results summary……………………………………………………….136
Table 6.3: Bond parameters and results summary……………………………………………………….136
Table 7.1: Full width half maximum tolerances …………………………………………………………..157
Table 7.2: Full width half maximum mechanical tolerances…………………………………………….162
Table 8.1 – Grating pitch comparison of Renishaw and competitors encoder products………………….170
Table 8.2 – Comparison of full width half maximum mechanical tolerances for the current parallel detector arrangement against the new chevron detector arrangement with 20, 8 and 4μm scale grating pitches…………………………………………………………………………………………………..172
Table 8.3 – Comparison of different reference mark implementations………………………………….173
Table 8.4 – Full width half maximum mechanical tolerances for separately processed detector banks……………………………………………………………………………………………..194
Table 8.5 – Summary of additional observations for 1Vp-p Lissajous curves…………………………….195
Table B1 – Substrate emitting epitaxy order…………………………………………………………….213
Table E1 – Cu electroplating bath composition……………………………………………………………222
Table E2 – Ni electroplating bath composition…………………………………………………………….222
Table G1 – Karl Suss FC6 thermocompression flip-chip bonding parameters for batch 1 bumps……….226
Table G2 – Karl Suss FC6 thermocompression flip-chip bonding parameters for batch 1 bumps……….226
Glossary

$\alpha$, angle
$\alpha$, signal output
$\theta$, signal input
$\theta$, phase shift
$\theta_A$, angular separation at the analyser grating
$\theta_i$, angular separation at the index grating
$\theta_s$, angular separation at the scale grating
$\rho$, material density
$\Phi$, Phase change
$\lambda$, wavelength
$^\circ C$, Degrees Celsius
$A$, analyser grating
$A$, array channel $A$
$A$, cross-sectional area
$a$, angle
$AC$, Alternating Current
$ACA$, Anisotropic Conductive Adhesive
$AC:DC$, Alternating Current to Direct Current ratio
$ADC$, Analogue to Digital | Converter
$Ag$, silver
$Al$, Aluminium
$ANSYS$, Analysis System
$ASIC$, Application Specific Integrated Circuit
$Au$, gold
$B$, array channel $B$
$BGA$, Ball Grid Array
C, array channel C
CW, Clockwise
CCW, Counter Clockwise
Circ, Circular shaped LED
CMOS, Complementary Metal Oxide Semiconductor
COF, Chip on Flex
COG, Chip on Glass
Cr, Chrome
CST Ltd, Compound Semiconductor Technologies Limited
CTE, Coefficient of Thermal Expansion
Cu, Copper
CVD, Chemical Vapour Deposition
D, array channel D
d, thickness
DC, Direct Current
DI, De-Ionised
DTI, Department for Trade and Industry
E, error
EC, European Commission
EMI, Electromagnetic Interference
eV, electron volts
EIA, Electronic Industries Association
F, Faraday constant
f, spatial frequency
FCB, Flip-Chip Bonding
FCoG, Flip-Chip on Glass
Fe, Iron
FEA, Finite Element Analysis
FWHM, Full Width Half Maximum
GaAs, gallium arsenide
gf, grams force
h, thickness
H₂O₂, Hydrogen Peroxide
HCL, hydrochloric acid
HDI, High Density Interconnect
HWU, Heriot-Watt University
I, current
I, intensity
i, current
i, index grating
IC, Integrated Circuit
ICA, Isotropic Conductive Adhesive
InGaAs, Indium Gallium Arsenide
InP, Indium Phosphate
IR, Infra Red
I/O, Input/Output
J, current density
JEDEC, Joint Electron Devices Engineering Council
l, length
Laser, Light amplification by stimulated emission of radiation
LCC, Leadless Chip Carrier
LED, Light Emitting Diode
LIGA, Lithographie Galvanik Abformung
Loz, Lozenge shaped LED
LTCC, Low Temperature Co-fired Ceramic
M, Molecular weight
m, mass
m, diffractive order
MCM, Multi-Chip Module
MEMS, Micro Electro-Mechanical Systems
mins, minutes
MISEC, Microsystems Engineering Centre
MOEMS, Micro Opto-Electro-Mechanical Systems
MOPVE, Metal Organic Phase Vapour Epitaxy
MST, Microsystems Technologies
MQW, Multiple Quantum Well
n, integer
n, number of free electrons
n, number of detector repetitions
n, refractive index
n, scaling factor
NH₃, Ammonia
Ni, Nickel
P, Phase output
p, period of grating
Pₐ, spatial period on the analyser grating
Pa, Pascals
Pb, Lead
PCB, Printed Circuit Board
PD, Photo-Detector
Pᵢ, spatial period on the index grating
ppm, parts per million

ppr, pulses per revolution

Pₚ, spatial period on the scale grating

Pt, Platinum

PTFE, Polytetrafluoroethylene

PVD, Physical Vapour Deposition

Q, Charge

Q, Quadrature output

QW, Quantum Well

R, length of one set of photo diodes in an array

R, resistance

r, mark space ratio

Ref, reference detector

Rev/s, Revolutions per second

RIE, Reactive Ion Etch

RF, Radio Frequency

rpm, revolutions per minute

s, scale grating

s, seconds

S₀, slit 0

S₁, slit 1

sec, seconds

SEM, Scanning Electron Microscope

Si, Silicon

Si₃N₄, Silicon Nitride

SiO₂, Silicon Dioxide

SiP, System in Package
Sn, Tin
SNR, Signal-to-Noise Ratio
SRD, Scale to Readhead Distance
T, Temperature
T, thickness
t, thickness
t, time
TAB, Tape Automated Bonding
Ti, Titanium
u, distance from index grating to scale grating
UBM, Under Bump Metallurgy
UV, Ultraviolet
UV-LIGA, Ultraviolet Lithographie Galvanik Abformung
V, volts
v, distance from scale grating to analyser grating
VCSEL, Vertical Cavity Surface Emitting Laser
Vp-p, Voltage peak-to-peak
W, Tungsten
w, width
WSP, Wafer Scale Packaging
List of Publications

Patent Application:


Conference publications:


Journal Papers:


Journal Papers (Invited, pending):


Journal Papers (under preparation):


Chapter 1

Introduction

Since early man first started to communicate and build societies he has been developing measurement tools in order to trade for food, clothing and materials, to build dwellings of specific shape and size and to know the distance from one settlement to another and how long it will take to travel between the two. These measurements are found in the form of weight, volume, length and time. Measurements which are largely taken for granted in today’s world [1.1]. However, the significance of man’s ability to measure may be much more profound with some suggestion that there may be a direct relationship between the advancement of man and his ability to measure smaller and more accurately [1.2].

Naturally man’s first instruments derived from parts of the human body, forearm, hand and finger, much because they were readily available. Weight on the other hand is more complex than length. As the human body does not lend itself to weight measurement alternatives were sought so man turned to nature and noticed that seed’s and grain were of fairly uniform size and used these to determine both weight and volume by counting the number of seeds. Time was perhaps more difficult still so here man turned to the sky’s monitoring the cycles of celestial bodies.

One of the first known units of measurement is the cubit, an ancient form of ruler based on the length of the forearm from the elbow to the tip of the middle finger. Of several standard cubits discovered the most notable are the Royal Egyptian cubit, dating back to circa 1550 B.C. this believed to be the length (20.62 inches) of the forearm of the Pharaoh Amenhotep, and the Greek Olympic cubit measuring 18.24 inches. It’s believed the Greeks and Romans inherited the cubit from the Egyptians and it is they who are responsible for first subdividing the cubit into a span, palm, and finger representing a half, one sixteenth and one twenty-fourth respectively [1.3].

However, by simply observing the construction of Egypt’s Great Pyramids, built circa 4500 B.C., it is apparent that the ancient Egyptians had developed sophisticated measurement systems prior to the reign of Pharaoh Amenhotep. Despite the base an area of over 13 acres the length of the sides are out by only 13mm and are within a degree of forming a perfect square. The blocks used are constructed with such precision that no
mortar has been used yet they have stood the test of time. Such feats of engineering are impressive even by today’s standards.

Thus today’s linear measurement tools have been centuries in the making and no less important to the precision engineer of today than they were to the ancient Egyptians several thousand years ago. As mankind progresses there remains a continued need for ever smaller more complex and accurate measurement, primarily driven today by man’s desire to explore space and in the electronics industry with endeavours to achieve ever smaller transistor nodes in order to increase the speed of our computers. To put it into perspective the effort required by the ancient Egyptians in building the pyramids is estimated to be equivalent to our putting man on the moon [1.3].

One of the most important tools in precision engineer today is the position transducer. These devices are used to convert a mechanical position into a representative electrical signal. From the position information, additional data including displacement, direction of movement, velocity and acceleration can all be determined. The most common devices used to achieve this are encoders.

### 1.1 Position Transducers and Encoders

#### 1.1.1 Electromechanical transducers

The earliest position transducers in the form of encoders were of the brush contact type, operating as an electromechanical transducer, such as a potentiometer [1.4]. Figure 1.1 shows the circuit diagram for a potentiometer (voltage divider) that consists of a length of resistive element onto which brush wipers are in intimate contact. As the wiper translates along the length of the resistive element the output voltage ($V_{out}$) varies as equation 1.1. Therefore the output voltage can be related to specific position along the length of the resistive element [1.5].
The first true encoders were also of brush contact type. These consisted tracks of printed conductive segments on an insulating substrate. The brushes, formed of wipers that individually address each of the tracks are then moved along the tracks as illustrated in Figure 1.2. When the brush comes into contact with a conductive segment it acts as an electromechanical switch closing the circuit. The patterns formed on the tracks when read concurrently outputting a binary signal with each bit representing a unique position along the measuring length [1.4]. Encoders that address the position using coding techniques are known as absolute encoders, with the number of tracks representative of the number of bits available.

Alternatively incremental measurement can be on a single track having a series of conductive and nonconductive segments. As the brushes contact the conductive segments a switch is closed as before. However a counter is no employed in the electronics that counts the pulses as the wiper is translated across the track.
Bidirectional movement can be determined by using two contacts to the track positioned such that the output pulses are shifted by 90° as shown in figure 1.3. When traversing the track the signal from contact 1 will either lead or lag the signal from contact 2 and direction of motion can be determined [1.6].

![Figure 1.3 – Illustration incremental brush encoder contact with digitised quadrature output](image)

Despite the relative simplicity and the low cost of these encoders they suffered from nonlinearity issues, backlash (mechanical hysteresis induced in the wiper), slow scanning speeds to avoid contact bounce and perhaps most detrimental was the wear and tare of the tracks and brushes which result in poor reliability and increased noise.

1.1.2 Induction Transducers

To address the issue of mechanical wear and tare non-contact measuring transducers, including magnetic and optical solutions, have been developed and as such the brush contact type encoder is now rarely used. Early non-contact position measurement devices include the Linear Variable Differential Transducer (LVDT) and the resolver.

LVDT’s as position transducers were first described in [1.7]. These devices comprise a minimum of three windings (one primary and two secondary) with a magnetically permeable core that moves between the windings inductively coupling the primary and secondary coils. As the core moves relative to the windings the primary will be more coupled to one of the secondary windings than the other introducing a voltage difference that is used to determine position from the null point (the location where coupling to both secondary coils is equal and no potential difference exists). The main advantage of LVDT’s is the near infinite resolution but they require drive and
conditioning electronics that combined with the windings result in large baulky devices. Furthermore the range of movement is restricted without adding considerable cost [1.8].

In a simpler form, inductive position sensors can be constructed from a ferromagnetic material that moves within a coil. As the core moves the coil inductance increases. The main attraction with simple inductive devices is low cost. However the total sensing length is defined by the length of the coil but additional length is required for when the core is fully extracted [1.9-1.12].

Resolvers are another well known device that utilises the inductive sensing techniques [1.11-1.13]. In this device a coil is attached to a rotating axis (rotor) that is supplied with an alternating current via a transformer that consists of two coils that are aligned to the rotary axis. The primary is fixed to the stationary windings in the transformer (stator) and the secondary on the rotor. The alternating current in the rotary induces an A.C. voltage in the secondary with amplitude that is dependant on the angle of rotation of the rotor. However when a second a.c. signal with 90° phase shift is introduced into the stator the rotor effectively modulates the signal with varying phase depending on the angle. The sinusoidal signals output in from the stator are separated by 90° or quadrature. This means that as well as using the output to determine position, displacement and speed, direction can also be determined depending on whether output 1 leads or lags output 2. Resolvers are most frequently used with motors because of the similarity in operation but they are also used where more rugged high endurance devices are required.

Where resolvers are used for rotational measurement the inductosyn, which can be considered as a folded out resolver, can be used to make both linear and angular measurements [1.12-1.13]. The inductosyn consists of a fixed conductive meandering structure and a movable conductive meander structure positioned in close proximity to the fixed component. Translation of the moveable meander induces a maximum voltage when the conductive paths overlap and a minimum when the movable part aligns to gaps in the fixed meander component. The relative changes in position as a result of this motion produce sinusoidal voltage signals. By offsetting the meander structure and connecting in alternate pairs two signals phase shifted by 90° can be produced making it bidirectional displacement measurements possible.
1.1.3 Magnetic Encoders

Magnetic encoders are generally Hall effect or magnetoresistive type [1.14]. In Hall effect type when a current passes through a conductive material in the presence of a magnetic field, applied at right angles to the current flow, a voltage that is perpendicular to the to both the current and magnetic field appears across the conductor [1.15-1.16]. This is called the Hall voltage the magnitude of which is proportional to both the current and magnetic flux density. For encoder applications a series of these Hall effect devices arranged along an axis with a permanent magnet used in the scanning head. Where the magnet is in close proximity with the Hall effect device a voltage output is produced in relation to the magnetic field strength. Additional Hall effect device or magnets in the scan head can be used to create phase shifted signals form which position and displacement and be determined.

Magnetoresistive types operate in similar fashion but where a change in voltage is found in Hall effect sensors magnetoresistive materials undergo a change in resistance in the presence of a magnetic field and applied current [1.17-1.18]. The effect is particularly large in permalloys due to the magnetic properties of the material. To create a position measuring device a tape patterned with magnetic materials having alternating polarities is used against a scanning head consisting at least two magnetoresistive pickups. The pickups can be positioned to provide two 90° phase shifted signals for quadrature detection. While Hall effect encoders are generally lower cost Magnetoresistive devices generally operate with higher precision, higher speeds, greater temperature ranges and have greater sensitivity to a magnetic field than their Hall effect counter part devices. Advances in magnetic encoder technology have resulted in today’s encoders achieving resolution of around 1μm over short measuring lengths [1.19-1.20].

Despite advances in inductive and Magnetic encoders they fail to match the resolution, precision, accuracy and speed that can be achieved with today’s optical encoder technology. For these reasons optical encoders are the dominant position feedback device limiting inductive and magnetic encoders to specialist applications where optical encoders are unsuitable. Inductive encoders are extensively used in heavy industry applications where they may be subjected to high levels of vibration and shock. In extreme environments where dirt and moisture are likely to cause contamination in
optical encoders an electric field remains unaffected thus magnetic encoders prevail in such conditions.

1.1.4 Optical encoders

Optical encoders are measurement devices found in almost all sectors of industry including metrology, motion systems, electronics, semiconductor handling, medical, scanning/ printing, scientific instruments, photography, space research, and specialist machine tools. Optical encoders can be defined as a transducer device that converts a mechanical position into a representative electrical signal by means of a patterned disc or scale, a light source and photo-sensitive elements [1.21-1.24]. Over the last 60 years extensive development of the optical encoder design in order to improve resolution, speed and accuracy has taken place. This sections summaries some of the more significant events with further discussion provided in chapter 2.

One of the first optical encoders was produced by the Baldwin Piano Company, now known as BEI Technologies, in 1949 [1.24-1.25]. They developed a coded optical disc that rotated to divide light provided by a light source into pulses of a particular frequency that impinged on to a photoelectric device. A second stationary disc located between the rotating disc and photoelectric device that was attached to piano keys by wires to operate shutters was used to modify the harmonics of the impinging light pulses. The electrical signals were then used to reproduce musical notes dependant on the frequency and harmonics of the light at the photoelectric device. However it was until a few years later that Baldwin modified the technology for use as a position measurement device after the Army Signal Corp had inspected the optical disc’s and commissioned the Baldwin Company to produce a binary absolute rotary encoder device used for controlling the position of radar antenna.

The accuracy of absolute encoders was improved, by removal of bit transition errors, with the application of the reflected binary coding systems otherwise known as Gray code after Frank Gray who patented the technique in 1953 [1.26]. Further discussion of binary versus Gray code is provided in chapter 2.1.2.

Meanwhile, Roberts reports on using a two grid structure to monitor linear displacement in a respiratory system [1.27]. Roberts describes a scanning head consisting of a
photographic plate having a series of opaque lines that rides opposite a fixed grid of similar structure such that light transmitted through both grids suffers a series of interrupts creating a Moiré fringe pattern that is detected by a photoelectric cell. Relative movement between the grid results in a corresponding movement of the light pattern that is then counted to determine displacement in this linear incremental encoder device.

Around the same time researchers at the Societe Genevoise D’Instruments De Physique describe a measuring tool that uses lenses to focus a spot of light onto graduated surface [1.28]. The light is focused such that the light beam is at most equal to the width of a single graduation. The light is reflected back up the microscope and through a slit in front of a photoelectric cell. The cell is used to monitor the intensity contrast of the impinging light in the form a cyclically varying electrical signal from which the position on the graduation is determined from a predefined starting location in an early form of reflective linear incremental optical encoder.

For incremental encoders the need to implement a bidirectional measurement system was recognised. As with the other techniques discussed above this requires two signals in quadrature. This was first achieved by using two photo detectors displaced from each other by a quarter of the fringe pitch such that two cyclically varying electrical signals 90° out of phase with each other were generated from the photo detectors. Shepherd et al, introduced an indexed grating system whereby the first grating, that interacts with light passed through a collimating lens, is formed of two parts. One part indexed relative to the other by a quarter of the grating pitch such that the 90° phase shift was built into the optical system [1.29]. Furthermore, they introduced a four channel detector arrangement with photo detectors connected in pairs such that the D.C. component of the photo detectors is cancelled and the modulated component amplified providing a larger, cleaner signal about a common D.C. level [1.30]. This is particularly useful in reflection based systems where unwanted reflections cause increases in D.C. noise at the photo detectors. These encoders were used in the first Coordinate Measuring Machines (CMM), developed by Ferranti Ltd [1.31].

The patent granted to Philips Electrical Industries Ltd in 1959 discloses an incremental displacement measuring device that exploits a special case of Moiré fringe [1.32]. Where the Moiré fringe pattern is generated by superimposing two gratings of similar
pitch with one grating skewed at an angle defined by the grating pitch and the fringe pitch a Vernier fringe pattern is generated when two gratings are superimposed with one grating having a slightly extended pitch relative to the other. As one grating moves relative to the other along the X-axis the fringes will move parallel to the grating, see figure 1.4(a), while they move perpendicular to the grating in the Moiré setup, see figure 1.4(b). The main advantage is a magnification of the mechanical movement.

Figure 1.4 – Illustrative example of (a) Vernier fringe pattern and (b) Moiré fringe pattern

This document also discusses the use of a three channel system whereby the photo detectors output three signals phase shifted by 120° as defined by three gratings positioned in front of the photo detectors. The signals are then processed in the electronics to produce two signals in quadrature from which position and displacement are determined.

In the early years the uptake of optical encoders was limited mainly due to the short lifetime of the light source. At this stage absolute encoders dominated the optical encoder marketplace. However with the development of solid state light sources increasing the lifetime of optical encoders and with the increasing availability of microprocessors and improving interpolation techniques the inherent simplicity and lower cost of incremental encoder technology soon grew in popularity. Today incremental encoders account for over 90% of the market [1.24].

Through the 60’s and much of the 70’s most developments largely involved variations of the themes discussed above. Thus optical encoders would comprise various combinations of the following components; light source (bulb, LED, Laser), Lenses (focusing, collimating, polarising), filters, gratings (transmissive, reflective, Ronche, diffractive), photo detectors and electronics for amplification, signal processing and
conditioning and interpolation. From the mid 70’s onwards perhaps some of the more significant developments would include the use of holographic scale gratings such as that used by [1.33]. In this system lasers are used to project a holographic image of the scale on to a coarser detector plane such that the finest of grating pitches can be employed and thus achieving the highest resolution. However these systems are restricted to short length measurements and the manufacturing cost remains prohibitively expensive, limiting their use to niche applications.

In 1978 the National Research Development Corporation were granted a patent for a three grating system [1.34]. The system relies on the diffuse illumination of a scale grating (the second of three gratings) via a first grating whose purpose was to create a spatially periodic array of photoemissive elements. These can be treated individually as point sources. Indeed this array could have been formed by an array of source elements but the pitch becomes restricted to the smallest manufacturable light sources. The transmitted light from the first grating (index) interacts with the scale grating (a diffraction grating) to produce an interference image at the third grating (analyser). The third grating has a period similar to the first but is skewed at predetermined angle defined by the period such that a Moiré fringe pattern appears over the photo detectors. Alternatively the third grating remains parallel to the scale grating but has a slightly different pitch such that Vernier fringes are produced. This system offers many benefits over some of it predecessors including, use with fine pitch gratings thus higher resolution, improved angular alignment tolerance, improved operation with respect to the distance from the scanning head to the scale (Z-axis) and small physical size largely due to the fact that the system operates with uncollimated light removing the need for lenses.

In the mid 80’s the company Heidenhain developed an incremental optical encoder that exploits the diffraction and interference of light on fine pitch gratings, in a three grating system [1.35-1.36]. The scale is a reflective diffraction grating with a feature graduation height corresponding to a quarter wavelength of the light source. The index grating is a transmissive phase prating with a period equal to that of the scale but designed so that the -1, 0 and +1 diffractive orders have equal luminous intensity. The light source illuminates a collimating lens such a plane wave is incident on the index grating, which splits the light into the three diffractive order paths. At the scale the light is diffracted with most of the energy going into the ± 1 order paths. These waves now
reflected meet again at the index grating where they are once again diffracted and interfere as illustrated in figure 1.5. The collimating lens then directs the light onto three photo detectors that convert the light into three electrical signals.

![Diagram of Heidenhain optical encoder](image)

*Figure 1.5 – Illustration of the three channel Heidenhain optical encoder based on the interferential measuring principle to produce three sinusoidal signals shifted by 120°*

The interaction of the light with the diffraction gratings mean result in a phase shift of the ±1 order beams such that the photo detectors generate three sinusoidal signals shifted by 120°. These are then converted in the processing electronics to two quadrature signals from which position and displacement can be determined. As the signals are made up of interfering ±1 order beams a single period transition of the scale results in an equivalent transition of both ±1 order beams but in opposite directions. Thus two signal periods are generated for a single period displacement. This design removes the need for additional gratings and/or tilted gratings so relaxes the alignment tolerances required to achieve the correct differential spacing. The scanning signals are largely free of harmonics so are suitable for use with high levels of interpolation. However this system is sensitive to differential contamination of the scale as the different phases radiate in different directions.

The final development to be discussed is an optical encoder scanning head developed by Renishaw Plc in 1986 [1.37-1.38]. The purpose of this device was to overcome one of the key disadvantages associated with the previously discussed devices particularly where diffractive scale gratings are used. Clearly these gratings must be of diffractive
quality therefore the accuracy of the mark to space ratio, the definition of the edges, the depth of the graduations and the freedom of the scale from scratches and imperfections are very important. Thus such scales are expensive to produce and as such are restricted to short lengths.

In this device the diffraction takes place within the scanning head rather than at the scale grating. The scanning head comprises the index grating (diffraction grating) and an analyser grating located in different planes and parallel to each other with spacing between them defined by both the wavelength of the light and the pitch of the gratings. The analyser is subdivided into four sections having mutually offset gratings such that Moiré fringe patterns can be sensed in phase quadrature by the photo detectors. The light source illuminates the reflective scale grating and the light reflected back to interact with the index and analyser before reaching the photo detectors. A lens is situated between the analyser grating and the photo detector to overcome reductions in fringe contrast caused by deviations from the ideal wavelength.

The index and analyser gratings constitute a spatial filter that is tuned to the period of the scale, within a given pass band for increased accuracy. With the index and analyser gratings fixed relative to each other within the head and the fringes alignment fixed relative to the gratings the sensitivity to alignment errors about the Z-axis is reduced leading to substantial immunity to quadrature phase errors due to misalignment.

As the purpose of the scale grating is to provide a pattern of light sources it is not required to be of diffraction quality. In fact the scale may be relatively imperfect so long as there is a dominant periodicity falling within the pass band of the filter. This means that the scale can be made using more economical manufacturing technologies reducing the cost and allowing for substantially increased lengths of scale to be produced e.g. an embossing process used on steel tape scale can be used to provide scale lengths of up to ten metres or more. The major drawback of this device is the tighter control required in maintaining the distance between the head and the scale grating. Too large deviations can result in destructive interference in the filter and no fringes will be detected.
It is evident even from this small sample that the choice of encoder technology may be dependant on many factors such as measuring length, speed, cost, accuracy and precision. Over the years many developments have been made, but are essentially variations and tweaks to the technologies discussed above. These developments tend to address some specific issues resulting encoders that are perhaps more suited to the users needs than others. Future developments will focus on achieving higher resolution and performance from more compact optical encoder devices in order to achieve cost saving, make space savings and allow encoders to fit into spaces where previously they could not be used thus to open up the market to previously inaccessible applications, particularly in the semiconductor, space application and medical fields.

1.2 Rationale for Miniaturisation

Today’s optical encoders largely comprise discrete components: light source, reference and analyser gratings, and a photodiode array. Optical diffractive methods are utilised to achieve high resolution. There is a continual demand for more precise reliable optical encoders for current as well as new applications driving the technologies development. Until recently, much of the attention has been focused on developing the optical techniques to increase accuracy and improve resolution which can result in increased complexity of the encoding method, manufacture and the required processing electronics. Moreover the critical alignment requirements between the optical gratings and to the photodiode arrays, the bulky nature of the encoder devices and the respective sophisticated packaging mean that optical encoders can be either prohibitively expensive for many applications and unsuitable for others.

In light of the shortcomings of most current encoders, the work presented in this thesis aims to develop a prototype optical encoder head with light source(s), optical gratings and photo detector arrays, all monolithically integrated onto the same semiconductor chip. Innovative integration schemes at the semiconductor level offer several distinct advantages over multi-component solutions: unprecedented resolution for a non-interferometric system as lithographic techniques define the gratings and detectors, novel light source designs to optimise the optical quality, simultaneous alignment for multiple readheads and angular alignment of individual devices. A secondary aim is to assemble this device into a more compact package.
The work carried out here involves the design, manufacturing, assembly, packaging and
testing of the resultant device. In order to meet the above discussed aims, the project
has the following objectives:

- Monolithic integration of the encoder components onto a single chip
- Reduction of size
- Reduction of cost
- Increase of resolution performance
- Improvement of reliability
- Increase of flexibility

Such objectives are consistent with the benefits associated with those that the
MicroSystems Technology (MST) can offer by the miniaturisation of systems. An
increasing number of microsystems products have been brought to market and the
microsystems industry is now widely seen as the next significant progression to the
microelectronics sectors. The term Microsystems Technology (MST), sometimes
referred to as Micro-Electro-Mechanical Systems (MEMS), as defined by the EC
Information Technology Programme is an “intelligent miniaturised system comprising
sensing, processing and/or actuating functions” [1.39]

Utilising the advantages of microsystems manufacturing and assembly, coupled with
novel design and packaging solutions, a series of prototype optical encoders have been
manufactured and are presented in detail in this thesis. The development of a flip-chip
bonding process for these devices has also been investigated but unfortunately a flip-
chip demonstrator module could not be achieved within the project time frame.
However wire bonded surface emitting devices have been used to successfully
demonstrate the concept of a monolithically integrated optical encoder chip achieving
signal to noise ratios in excess of 60:1, DC to AC ratios of the order of 7:1 and peak to
peak voltages reaching the operating constraints imposed by the electronic test
equipment. Furthermore various reference mark regimes and designs have been
investigated for compatibility with the small form factor of the miniaturised encoder
chip. While a definitive solution has not yet been obtained, a number of designs have
shown considerable promise and require only minimal further development to ensure
compatibility. These results indicate that an encoder, an order of magnitude cheaper and
smaller than existing sub-10nm products, is feasible. Such an encoder is to be produced by Renishaw PLC in the near future.

1.3 Thesis Layout

This thesis is to report on the progress made towards achieving the aims and objectives of the project outlined above. The work was carried out at the Microsystems Engineering Centre (MISEC) at Heriot-Watt University in collaboration with Renishaw plc, based on the research park of the same University. The thesis consists of ten chapters as summarised in figure 1.1.

Chapter 2 Overview: The basic optical encoder operation is presented in this chapter alongside a detailed overview of some optical encoder operation principles. The present technology employed by Renishaw in some of their current encoder devices is highlighted and some fundamental design equations based upon this approach are introduced. A brief presentation of the assembly and packaging of a current Renishaw optical encoder is also provided.

Chapter 3 Literature review: This chapter outlines the integration methodology and provides a review of the integration techniques as applied to current optical encoders. The merits of wire bonding and flip-chip bonding are explained. Other assembly methods that also provide electrical interconnections are discussed. State of the art compact packaging solutions are briefly presented.

Chapter 4 Miniaturised optical encoder design: This chapter introduces the conceptual design of a monolithically integrated optical encoder chip. Advanced packaging solutions are proposed. More details about the chip design and component layouts are also provided. Various component geometries are described and assessed against the desired criteria. One design in particular, the chevron photodetector arrangement, is singled out, and becomes the focus for the development work presented in subsequent chapters.

Chapter 5 Bump fabrication: This chapter presents the merits of bump technologies as applicable to the first level package integration of the optical encoder chip. Further details on the selected gold electroplating technology are provided. An adapted LIGA
process is presented as a solution towards the development of a sustainable bump fabrication process with challenges and modifications discussed along the way. Results obtained through electrodeposition of gold bumps are presented.

Chapter 6 Flip-chip bonding results: The findings of preliminary thermocompression flip-chip bonding investigations are presented in this chapter. Results of early stage finite element analysis models are presented and subsequently applied to bonding profiles. A novel flip-chip pick-up tool is designed to overcome detrimental stresses induced in the chip on the application of the bonding force. Bonding parameters are investigated and optical encoder chips are successfully bonded to glass submount carriers.

Chapter 7 Wire bonded encoder chip: Optical encoder chips wire bonded onto glass and silicon chip carrier submounts are used for test and verification. First the detectors and LED’s are operated discriminately to verify operation before the influences of operating both together on the same chip are investigated. Using a basic electronic setup the devices are tested for functionality against a Ronche scale grating to assess the performance of the incremental features of the encoder. Lissajous figures are used to demonstrate the success of the proof of concept device.

Chapter 8 Reference mark design: This chapter presents further advances in the optical encoder development. Chips are wire bonded into LCC packages with a range of grating pitches used to assess suitability and verify the incremental operation and different resolutions. Several reference mark schemes using auto-correlation cross correlation and split photo detector techniques are presented. The results and implications from testing of the reference designs against phase scale grating are presented and discussed.

Chapter 9 Conclusions: This chapter offers a summary of the work presented in this thesis covering the performance of the prototype optical encoder and the fabrication processes used in their assembly. A section on future work describes areas for improvement, further characterisation and potential packaging solutions.
Figure 1.6 – Schematic of the thesis layout
1.4 References

[1.1] “A Brief History of Measurement” US Department of Commerce, Technical Administration, National Institute of Standards and Technology (NSTI). Based on NBS special publication; 304A, reissued by NIST, Sept 1974


[1.32] “Improvements in or relating to Devices for Measuring the Displacement of a Moveable Member Relative to a Stationary Member”, GB Patent 816723, Philips Electrical Industries Limited, Jul, 1959


Chapter 2

Optical encoder overview

The purpose of this chapter is to provide information on the operating principles of an optical encoder and some of the different encoding techniques employed to date. The encoding technique employed by Renishaw in some of their current encoder devices is also discussed. In addition a Renishaw optical encoder readhead is described in order to gain an insight into how the components are assembled so that the benefits of the encoder proposed in this work can be fully appreciated.

2.1 Optical encoder basics

Encoders can be defined as transducer devices that convert the mechanical position or speed of a system or subsystem into a representative electrical signal. This can be achieved by using mechanical, magnetic, optical, inductive or capacitive sensing [2.1-2.4].

Compared to competing technologies, optical encoders are compact, inexpensive and reliable. They provide high resolution while maintaining a simple electronic interface and thus dominate the market as the most widely used feedback devices [2.3-2.5].

Optical encoders, as a position sensing device, were first developed in the early 1950’s to meet a demand for higher resolution and increased lifetimes that were not achievable with other sensing methods. However it was not until the late 1960’s with advances in the semiconductor industry and the invention of the LED that optical encoders became truly popular within the industrial measurement systems industry, overtaking their brush contact type counterparts. In addition the optical encoder offered advantages such as reduced size and weight, low susceptibility to noise, less electronics and lower sensitivity to mounting tolerances [2.2, 2.6].

In their simplest form, most optical encoders operate on the same basic principle as shown in figure 2.1: a light emitted from a light source passes through (transmission method) or reflects off (reflection method) a periodic or specially patterned graduation,
commonly known as a scale grating, and onto a series of photo-detectors that generate electric currents corresponding to variations in light intensity. The encoders come generally under two classifications as illustrated in figure 2.2: angular (rotating disc or shaft), or linear (strip of patterned material), also known as rotary and incremental, for the measurement of angular and linear displacements respectively [2.2, 2.5-2.7].

Figure 2.1 – Basic optical encoder operating principles showing: transmission mode, top, where light passes straight through from the source to the detector and below, reflection mode, where light reflects from the scale onto the detector.

The optical readhead is the name given to the fully packaged optical encoder unit that scans the scale grating. The readhead includes at least one light source, index grating, analyser grating, photo sensitive elements and additional electronics. Both linear and rotary encoders can technically be subdivided into categories based on the encoding technique employed. These are; incremental encoding, whereby detection of movement is made from a count of the periodic elements relative to a reference point, and absolute encoding whereby a specially patterned graduation is used to output absolute position typically in a binary format. Although both techniques are applicable to linear and rotary encoders, in practice linear absolute encoders are rarely used and have limited performance characteristics. On the other hand the inherent simplicity and low cost associated with incremental encoders has seen them dominate with around 90% of the optical encoder market [2.2-2.5].
Incremental and absolute encoders will now be discussed in greater detail.

2.1.1 Incremental encoders

The incremental technique requires a minimum of two coded patterns defined on the scale. One pattern is a track of evenly spaced opaque, and transmissive or reflective segments along the length of the scale in the case of a linear encoder, figure 2.2(a), or around the periphery of a disc for a rotary encoder as illustrated in figure 2.2(b). As the scale moves relative to the readhead sensing unit, the transmitted light generates a sequence of pulses and subsequent square waves, figure 2.2(c). The number of pulses generated over the length of the scale can be defined as pulses per revolution (ppr) and indicates the minimum resolution of a system. The resolution is dependant on the number of lines and is therefore only limited to the minimum readable line features achievable [2.6-2.11]. The reference markers act as initiation or reset identifiers of known locations on the scale from which the current position is incremented or decremented as appropriate.

In basic mode, having a single pulse train, only the leading edge is counted. Thus a scale of 200 lines would have a resolution of 200ppr. However most incremental encoders output two signals, commonly known as A channel or B channel or Quad and Phase, in quadrature fashion as standard. The two channels are phase shifted by a quarter cycle or 90°, which is necessary to achieve bidirectional operation. The term quadrature relates to the transition of one of the data channels, occurring four times per count cycle. Generally a count cycle is considered to comprise 360° (electrical degrees) as highlighted in figure 2.3. Therefore the direction of motion is determined from whether A leads B or A lags B.
By counting both the rising and falling edges of both channels, the resolution of the system can be easily increased by a factor of 2 or 4. Now consider that the digital signal is derived from a raw analogue signal that consists of sine and cosine signals, where the number of sinusoidal cycles per unit determines the resolution, then it is possible to operate the optical encoder in analogue quadrature mode.

The analogue quadrature output, in the form of a \( \sin(\theta) \) signal, A, and a \( \cos(\theta) \) signal, B, where \( \theta \) is the phase shift within one cycle, can be operated on directly to generate higher resolution. Higher interpolation is achievable by periodically sampling the signal and digitising using an analogue to digital converter (ADC) within one cycle. An interpolation table or arctangent calculations can then be used to determine position. Although limited by the number of lines present in the system the true limit of interpolation depends on the quality of the basic signal (mark space, jitter, quadrature separation, noise). Knowing the direction of travel allows the determination of distance and hence position, relative to a known location identified by the reference mark [2.8-2.11]. This is achieved by associating a distance with each line segment and incrementing or decrementing the count as appropriate. Speed is measured simply by timing the pulses entering the controller and similarly the acceleration rate is a measure of the change of speed over time. For the case of angular measurements, the distances are associated to a predetermined degree of rotation.
The advantage of incremental encoders is the ability to encode over a finer pitch and thus achieve higher resolution compared to absolute encoders. The drawback of using the incremental technique is the loss of position data on power down or power failure in which case the system must reset to a reference mark before position can be determined once again. Clearly this must be taken into consideration when setting up an optical encoder position measurement system as loss of position data may be undesirable or indeed catastrophic in some applications.

2.1.1.1 Lissajous Curve

The two cyclically varying electrical signals produced from the photo detectors in the form of equations 2.1 and 2.2 when plotted against each other produce a Lissajous curve.

\[ x(t) = A \sin(\omega t + \delta) \]  
\[ y(t) = B \sin(t) \]  

The shape of the curve depends on the ratio of the amplitude, frequency and phase of the signals, see figure 2.4. For a ratio of 1 an ellipse will appear, and for the special case of two signals of the same frequency, \( A = B \) and with \( \delta = 90^\circ \), a circle is obtained.
Therefore the shape and size of the Lissajous curve provides information on the relative phase difference, signal amplitude and alignment. As the signals are generated from the scale grating, counting trips around the Lissajous curve provides information on the position. Furthermore the gratings used in the optical encoder have been designed to produce 90° phase shifted signals and as the light source, index grating, analyser grating and photo detectors all move together relative to the scale grating the frequency of the signals are identical. Thus circular Lissajous curves are expected from optical encoders.

Figure 2.4 – Examples of Lissajous curves having (a) frequency and amplitude ratio’s of 1 and a phase difference of 90°, (b) where the phase shift is modified to 0°, (c) where the amplitude ratio does not equal 1, and (d) where the frequency ratio is 3:1

2.1.2 Absolute encoders

Absolute encoders provide absolute position information. The loss of data at power down/failure using incremental encoders can be seen as a disadvantage in some systems or simply may not be appropriate. For example a common application of absolute encoders is in robotic arms, where position information is critical to the arms operation accuracy and safety. With an absolute encoder position, information is instantly available as a unique output is provided for every position [2.3, 2.6-2.7, 2.12-2.15].
Absolute encoders are more complicated than their incremental counterparts. The disc features multiple tracks, each having their own set of opaque and clear line segments, as shown in figure 2.5. At least one independent photo-detector is associated to each track and all data read concurrently. Absolute encoders use binary counting such that each track is executed as a binary bit and an 8-bit encoder will require eight tracks, giving 256 unique positions. Each angular position is assigned a unique signal in the form of a binary code, subject to the encoder resolution.

The trade-off with absolute encoders is that, in order to double the resolution, an additional track is required. While an incremental encoder may have 256 cycles per revolution, requiring only an incremental and reference tracks, the absolute encoder requires eight data channels. This means larger devices with thicker more expensive cabling and connections [2.12].

 Typically Gray code, figure 2.6(b) is preferred by manufacturers over pure binary code, figure 2.6(a). With pure binary code, a source of error can be found when measuring over coinciding edges of light and dark between tracks since it is not possible to guarantee that all signals will switch simultaneously [2.12-2.16]. This problem is overcome using Gray code as the edges of the light/dark segments are included in each bit. The individual groups must overlap to allow for some simple logic circuitry to derive the alignment of the edges and convert the Gray code to pure binary code.

With each position having a unique signal, the absolute encoder will not lose position information on power down. Thus, there is no need for any index marker for a “reset” position.
There also exists a third method, which implements features of both the incremental and absolute methods in a pseudorandom code. This technique has been developed and employed by Gurley Precision Instrument in their Virtual Absolute® Technology [2.17]. As with the incremental technique the virtual absolute approach requires only two tracks. However the second track normally used as the reference mark becomes a serial code similar to a bar code. Although position is not known immediately on start-up as with a conventional absolute encoder, only a very short movement is required in either direction before the exact position is known. In contrast, an incremental encoder may require up to one full revolution depending on the number of reference marks. From then on the encoder is truly absolute in the sense that it generates binary data to provide the position at that instant.

2.2 Measuring methods

With reference to figure 2.1, encoders can be either transmission based [2.18-2.21] or reflection based [2.22-2.32]. In transmission based encoders the photo detectors are positioned on the opposite side from the light source (relative to the scale). The light passes through the scale onto the photo detectors. Although, the same basic principles apply, reflection based encoders replace the clear/opaque line segments, which allow transmission, with reflective line segments. This allows the photo detectors to be located on the same side as the light source as the light is reflected back from the scale, resulting in a device with much lower profile.
Depending on the configuration of the gratings used, the scanning process over the photo detectors commonly employs one of two measuring methods – the imaging principle or the interference principle.

2.2.1 Imaging Principle

In this method, two gratings with near identical patterns are positioned face to face in close proximity to one another. The fixed grating is used to project an image onto the scale graduation. The relative movement of the scale against the fixed grating generates a series of light and dark modulations, corresponding to the projected image and the scale motion, which are detected by the photodiodes.

Figure 2.7 illustrates the imaging principle as applied to a transmission type encoder. A collimating lens will often be found between the light source and the scanning grating. The scanning grating is positioned a short distance from the scale such that the light beam will pass first through the scanning grating and then the scale grating such that a Moiré effect may be observed. This is also referred to as the “Moiré principle”, when light is projected through two closely positioned gratings [2.33].

![Figure 2.7 – Illustration of the imaging principle for a transmission type optical encoder setup](image)

The grating patterns on the scale and the scanning grating consist of transparent and opaque line segments. The scale is illuminated by images projected from the light passing through the windows on the scanning grating. The patterns on the scanning grating match those on the scale and are aligned in parallel. Photodiodes are then placed
behind the scale. The incident light beam is modulated in a sinusoidal pattern as the scale moves relative to the scanning grating. The photodiodes therefore detect the variation in light intensity, which is then converted into varying electrical current. The position of the windows/images is such that the resulting detected signals are phase shifted by 90° (figure 2.3). Appropriate electronic circuitry can then be used to determine the direction of motion.

2.2.2 Interference principle

This method makes use of the diffraction of light (diffractive interferometry) at the grating to produce the measuring signal. Movement of the scale causes light in the higher order of diffraction to experience phase shifts that are proportional to the displacement. The different orders of diffraction are superimposed and brought into interference so that the detection of intensity variation rather than phase difference can be achieved at the photo detectors. This is commonly achieved using a two grating system in similar fashion to the imaging principle above but instead using diffraction gratings [2.22, 2.25, 2.26, 2.31, 2.32, 2.35 and 2.36].

![Figure 2.8 – illustration showing the path of two light beams that interfere at the analyser plane](image)

However, a triple grating system has become increasingly popular with encoder manufacturers [2.21, 2.23, 2.28-2.30, and 2.35-2.38]. The three gratings method exploits the diffraction of light to again form interference fringes that move perpendicularly to the scanning direction over the photo detectors. The third grating can then be constructed such as to create a 90° phase shift and thus direction can be detected. The arrangement of the photo detectors within this method allows for an increased tolerance of pitch and planar misalignment of the readhead to the scale and
also allows for high interpolation techniques to be applied. To illustrate this principle, consider light travelling from a near point source illuminating the scale grating. Much of this light is simply reflected directly but some is diffracted into different diffraction orders. This means that light incident on a certain point on the analyser plane can come from a multitude of paths.

Figure 2.8 shows the travel of just two beams, tracing the -1 order of one beam and the +1 order of the second. In this case the scale acts as a diffraction grating. The beams are diffracted by the scale grating and enter a third grating positioned at the analyser plane. From the third grating the diffractive orders are partially overlapping and interfere, which leads to a change in intensity onto the photo-sensitive device located beyond the analyser plane.

The detector array is constructed to have the same pitch, but with a narrower width, than the interference fringe formed on the surface of the photodiodes. Consequently the bright and dark portions of the interference light coincide with the photo-sensitive device. When the scale moves by half a scale pitch relative to the readhead, the interference pattern translates by one fringe. In other words, the movement of a 20μm grating by 10μm results in the 20μm interference pattern moving by one fringe, thereby doubling the resolution. This method also means that the measurements are impervious to translation in the out of plane axis (Z), the distance between scale and readhead. The interference fringes are generated from the same location on the scale but it will be different source rays that are illuminating that location thus there is no effect on the interference fringe pattern generated.

Further development of this scheme utilises both the positive and negative order diffracted beams as illustrated in figure 2.8. Here the analyser plane is expanded about the point source and three beams are shown tracing the positive and negative diffracted orders to their intersection on the analyser plane. This offers the benefit of improved sensitivity (against particle contamination) as a result of obtaining the interference signal light from multiple positions. This example also offers advantages against attaching errors of the optical system (mechanical position, thermal expansion, light source instability, etc.) because the light is merely deviated as the parallelism between the analyser grating and the scale is maintained. This is also true for environmental
temperature fluctuations, which may cause fluctuation of the wavelength of the light source.

![Figure 2.9 – illustration interference principle showing how positive and negative orders of diffraction can be used to expand the scanning area](image)

Employing a reflection method to utilise first order diffractions in a three grating system is therefore an attractive option. The reflection method allows the readhead to be arranged on the same side of the scale, thus construction into a tight compact design, while utilising the first order beams provides high resolution and improved sensitivity.

In the next section we further discuss the three grating system with respect to current encoder products offered by Renishaw.

**2.3 Present Renishaw optical encoder**

The latest optoelectronic products from Renishaw for measuring the relative displacement of two members employs a three gratings system comprising a reflective scale graduation, which is scanned by a readhead. The readhead includes: (a) a light source for producing rays incident on the scale, (b) an index grating for diffracting readable rays into fringes in at least one order of diffraction, (c) an analyser grating for converting fringes into light modulations at a rate which is a function of the relative movement between the scale and the readhead and (d) arrays of photo sensitive elements for converting the light modulations into a representative electrical signal.

Typically the analyser pitch is set to extend or curtail the overall grating length relative to the length of the index grating to generate a Moiré fringe pattern over the photo detectors positioned behind the analyser grating. The analyser grating and photo detectors are configured to produce four phase shifted cyclically varying electrical
signals, combined in quadrature fashion, from which the magnitude and displacement of the relative movement between scale and readhead is determined. [2.39, 2.40].

The basic functions of each grating are as follow (see figure 2.10):

- **Index Grating** (grating i): this grating contains alternate light transmitting and opaque lines whose angular pitch and circumferential spacing are identical to the portion of the lines on the scale grating with which the index grating is in register. The purpose of this grating is to project a marked free space pattern onto the scale.

- **Scale Grating** (grating s): the scale grating consists of a sequence of periodically spaced reflective segments, which reflect a periodic light pattern onto the analyser grating. In this case, however, the scale behaves as a diffraction grating and the projected image is a result of the interference of different diffraction orders. The introduction of the additional index grating also allows the reflected pattern to achieve twice the resolution of the analyser grating i.e. 10µm scale translation gives 20µm fringe translation as output.

- **Analyser Grating** (grating A): the analyser (converter) grating has a slightly different pitch to the index grating and interference fringes. The purpose is to interact with the light and beat out a pattern (convert) in the form of a Moiré
fringe pattern, which is projected onto the photo-detectors [2.39]. Thus beams generating patterns having different spatial frequencies are of no consequence.

Figure 2.11 illustrates the basic optical path found in this set-up and the position of the gratings. The spacing between the index and analyser grating determines the distance of the readhead from the scale. The most basic illustration of the optical path is when the distance from the index grating to the scale, $d_1$, is equal to the distance from the scale to the analyser grating, $d_2$, as shown in figure 2.11.

\[ \frac{f_2}{f_1} = \frac{2d_2}{(d_1 + d_2)} \]  
(2.1)

Where $f_1$ and $f_2$ are the spatial frequencies of the first and second gratings, and $d_1$ and $d_2$ are the distances to and from the second grating as shown in figure 2.10. The spatial frequency is inversely proportional to the grating period and can be defined as $\frac{1}{p}$, where $p$ is the period on the grating. Therefore substituting for $p$ and arranging to find the index period having set the scale period we have

\[ p_i = \frac{2d_2p_s}{(d_1 + d_2)} \]  
(2.2)

where $p_i$ is the index period and $p_s$ the period on the scale. In this design $d_1$ is approximately equal to $d_2$ thus,

\[ p_i = p_s \]  
(2.3)

Similarly the period at the analyser fringe plane would also have a period equal to the scale, such that $p_i = p_s = p_a$. However, to generate Moiré fringes at the detector, the analyser grating is set to either extend or curtail the index grating over the range of the detector area. Therefore the actual analyser period, $p_a$, is...
\[ p_a = \frac{R}{\left( \frac{R}{P_i} + 1 \right)} \]  

(2.4)

where R is the length of a single diode array set forming the quadrature output.

Looking again at the beams travelling through the system, we consider this time a beam to be a bundle of locally coherent rays. Two of these rays passing from the point S0 on the index, with an angular separation, \( \theta_i \), are plotted in figure 2.12. Note that, for illustrative purposes, the diagram has been folded out horizontally and the rays would actually be reflected back to the analyser in the same plane as the index. \( \theta_i \) relates to a (spatial period) \( P_i \) on the index given by:

\[ \theta_i = \frac{\lambda}{P_i} \]

(2.5)

where \( \lambda \) is the optical wavelength and \( \frac{1}{P_i} \) is the associated spatial frequency. At the scale, these rays are diffracted. The angle of diffraction is:

\[ \theta_s = \frac{n\lambda}{P_s} \]

(2.6)

where \( n \) is an integer and \( P_s \) is the period on the scale. In the diagram the +1 order of one ray and the -1 order of the other are plotted to its intersection. Here the rays are coherent and produce a tiny length of interference fringe pattern on the analyser with a period,

\[ P_A = \frac{\lambda}{\theta_A} \]

(2.7)

The result is the coherent illumination of the scale in a localised area. The size of the area depends on the pitch of the index grating and the distance to the scale. Diffraction of this beam from the scale results in an image projected onto the analyser grating due to the different diffraction orders as shown in figure 2.12 (from S0).

A second set of rays is now passing through another slit S1 illuminating the same area. The rays behave in exactly the same way and the scale diffracts the rays to converge on the analyser plane. However, the new rays will not be coherent with the first, as is the case of an LED used as a source, so they will not interfere but instead produce the same interference pattern at the analyser plane. Therefore, as they both produce the same pattern their intensities rather than their amplitudes will be added together.
Movement of the interference fringe pattern relative to the analyser grating, upon the linear or rotational movement of the scale graduation, by a displacement equivalent to one interference fringe pitch, will result in the corresponding shift of the resultant Moiré fringe produced by the analyser grating. Figure 2.13 provides an illustrative example of a 90° phase shifted Moiré pattern.
2.3.1 Present Optical encoder readhead assembly

Three discrete components, an LED, a photo detector chip and corresponding optical gratings form the core of current optical encoder readheads. Additionally some processing and interconnect electronics are included within the top level package (a folding metal case), as shown in figure 2.14, the level of which depending on the onboard complexity applied within the readhead. The schematic shown in figure 2.15 illustrates the general layout and relative positions of these key components.

The reflection technique is used and a window allows the transmission of the optical signal to and from the scale grating. An 8 to 16 wire cable, is attached for power supply and signal transfer, the number of wires required reflecting the level of the electronic complexity included.
Manipulation of the analyser gratings configuration can result in parallel (or linear) fringe patterns generated from the radially extending fringe pattern [2.39]. Thus a photo detector with an array structure suitable for both angular and linear encoders can be manufactured. This provides a number of advantages, in particular changes in the pitch or configuration of the index or scale can be made with a simple photo detector.

The present detector array has a total of 60 parallel photo sensitive elements (15 for each phase) measuring 1500 x 40µm wide seperated by 10µm. Along with additional coupling metallisation this results in a photo detector chip with dimension of 4 x 3.5 x 0.5mm. The elements of the array are grouped into four sets A, B, C and D to correspond to the signal phase shifts generated by the position of the detectors against
the analyser grating namely sin, cos, -sin and -cos. A typical light distribution signal across a photodiode array where the respective elements are coupled and amplified in either phase (P) or quadrature (Q) output is also demonstrated in figure 2.16. Using four signals to generate the quadrature outputs is advantageous in that the difference amplifier serves to increase the amplitude i.e. A-C = sin - sin and B – D = cos - cos. Furthermore this acts to cancel out common mode noise. This detector chip is then glued down, photo sensitive side up, to a PCB and contact pads wire bonded out as shown in figure 2.17.

![Images showing (a) the optical chip and grating assembled to PCB and (b) an enlarged view of the area at the chips edge highlighting the wire bonds encapsulated in epoxy](image)

Using standard lithography the optical gratings, index, analyser and reference, are fabricated by the selective etching of a thin layer of chrome which has been deposited on glass to form the scanning graticule part. The scale pitch is 20μm so from equation (2.4) we find the analyser grating can have a pitch of either 18.18μm or 22.22μm (18μm and 22μm in practical terms). The glass is then bonded to the photo detector chip and
support structure using a refractive index matching adhesive gel. Some additional epoxy is applied around the edge of the glass and over the exposed edge of the detector chip to provide added support and to help protect the wire bonds. The alignment between the photo detector array and analyser grating is of critical importance to the operation of the encoder. This process is therefore time consuming and labour intensive both contributing substantially to the overall device cost.

The light source is an off the shelf LED positioned on the opposite side of the PCB from the detector chip and scanning graticule. A cavity is cut out of the PCB board, as shown in figure 2.18, for illuminating the index and reference gratings. The LED is slotted mechanically into position such that the majority of the optical power is focused towards the detector area as illustrated in figure 2.9. The LED is held in position by a rather crudely applied epoxy as shown in figure 2.18(b). Although alignment is not critical the time and labour required for the formation of the cavity in the PCB and assembly of the LED is again adding to the overall cost of each device.

Figure 2.18 – Images showing (a) the cavity in the PCB, exposing the index and reference gratings that are illuminated by the LED and (b) a magnified view of figure 2.13(b) showing the epoxy encapsulated LED
The complexity of the additional electronics and PCB, shown in figure 2.14(b), varies from application to application. However the simplest of four channel system has a minimum cabling requirement of seven or eight wires, two or three for power delivery and four for the ABCD signals extraction. The cabling itself can be costly and can be the same cost as the rest of the assembled readhead in some cases. The cabling can be inflexible restricting the encoders use in some applications.

2.4 A brief summary of competing products

In addition to the various three grating systems Renishaw also offers products that utilise the filtering technology, as discussed in chapter 1.1.4, as well as absolute and magnetic encoders. However these do not lend themselves so readily to miniaturisation. Renishaws high precision optical encoders can offer resolutions down to 5nm when used with interpolation electronics of the order of x 2000.

Like Renishaw their competitors in the optical encoder market naturally offer products employing various measurement methods in order to cater for wider markets and some specialised requirements. As discussed in chapter 1 these encoder technologies are loosely based around a few core themes with adjustments generally allowing one advantage at the expense of another. In the high precision optical encoder arena some noteworthy competing devices are offered from Sony, Heidenhain, Numerik Jena and MicroE.

The Sony encoder measures displacement by monitoring the phase change of a laser beam as it passes through a hologram grating. Submicron resolution is achievable. However measuring lengths are of less than 1m and as this is a transmissive system the encoders are inherently baulker than their reflective based counterparts.

Most of the Heidenhain high precision products use interferential scanning techniques to produce for channel and three channel outputs similar to the techniques discussed in chapter 1.1.4. Their highest precision encoders offer resolutions down to 4nm. However despite not yet being listed in their catalogues the LIP 200 series of devices have been reported to reach picometre resolutions [2.36]. The diffraction gratings used with these devices have a pitch of 0.512μm and are expensive to produce so measuring
lengths are limited to a maximum of around 200mm. The grating pitch and quoted resolution also suggests that incredibly high interpolation factors are being employed.

The MicroE encoders use a measuring method similar to the imaging principle discussed in chapter 2.1.2 whereby the light passes through a first grating that projects and image onto a reflective scale grating. The encoders are cheap to produce and with a 20μm pitch grating can reach resolution of 1.2nm, which for a two grating system means interpolation of around x 16000. A tape scale measurand is offered up to 30m but for high accuracy a glass scale must be used which limits lengths to a maximum of 1m. This system suffers from very poor dirty immunity such that fingerprints on the scale can cause significant errors and is therefore only suited to very clean environments.

Numerik Jena offers a neat compact optical encoder that uses chip on glass technology and Epiflex connectors. The measuring principles are similar to the triple grating principles discussed in this chapter where by diffused illumination form a light source penetrates an index grating, then stikes a scale grating and is reflected back from the scale, which has a slightly different pitch to the index grating. The light passes through the index for a second time and reaches an array of p-i-n photo detectors. The index is rotated slightly with respect to the scale such that the angular offset generates a moiré fringe pattern, which moves perpendicularly to the measuring direction, over the photo detectors. With a grating pitch of 20μm Numerik Jena offers x 100 interpolation to achieve resolutions of 0.1μm, which is modest compared to those mentioned above. It is unknown whether higher levels of integration can be used with this device in order to reach the resolution achieved in the above mentioned devices. None the less this device is interesting particularly due to the compact nature of the design and is therefore discussed further in chapter 3.1.3 with packaging in mind.

2.5 Summary

The basic operating principles of generic optical encoders and the optical techniques employed by Renishaw in some of their encoder products have been presented in this chapter. A brief overview of the manufacture and assembly procedures used for the packaging of a Renishaw encoder is also discussed.
The alignment of the analyser grating to the photo diode detector array is highlighted as a critical procedure vital to the accurate operation of the encoder. Therefore this assembly process can become expensive. Other important assembly issues such as costs and numbers of wire within the cabling, and the LED assembly are also raised. These points are crucial aspects to consider for meeting the goals of a successfully miniaturised optical encoder.

2.6 References


Chapter 3

Literature Review – Assembly and Packaging Techniques

There is a continual demand for more precise reliable optical encoders for current as well as new applications. Until recently, this demand has been met by development of optical techniques to increase accuracy and improve resolution which can result in increased complexity of both the encoding method and the required processing electronics, but this has now reached a state whereby substantial effort and cost is required to achieve only small gains and such rewards can no longer be justified. Much attention is now focused on the manufacturing of the optical encoder device. In particular the assembly and packaging of the encoder components to reduce size and cost, while improving performance.

In this work the integration of the key optical components is considered as are advanced assembly and packaging techniques. Therefore, this chapter discusses the merits of both monolithic and hybrid integration and provides a review of assembly and packaging techniques as applied to optical encoders. Furthermore flip-chip and wire bonding have been identified as interconnect, assembly and packaging technologies suitable for advanced optical encoder.

3.1 Integration

Integration falls into two categories: 1) monolithic integration, whereby all system functionalities are brought together on a single chip and 2) hybrid integration, where integration is achieved by combining a number of chips into a single module. In this section both monolithic and hybrid integration technologies are discussed with particular emphasis on the integration in optical encoders.

3.1.1 Monolithic integration

Monolithic integration enables the realization of all required system functionalities onto a single chip. The electronics and MEMS elements are brought together in one manufacturing process. Optoelectronic integration is no different, but the process must
be able to integrate both optical and electrical components on a chip [3.1-3.4]. In this chapter monolithic integration refers to the bringing together of the light source(s), photo-detectors and transfer gratings onto a single chip.

Monolithic integration can reduce or eliminate the need for pick-and-place, wafer bonding and other assembly processes necessary in hybrid assembly. This leads to improved alignment accuracies which can be critical in many optical systems. Furthermore monolithic integration on a substrate such as silicon can offer better heat dissipation, which can directly improve the power efficiency of individual components such as lasers and LEDs. This thermal advantage may also lower the operating power requirements, which, in turn, improves the reliability of the device, due to the reduced operating temperatures over a longer period of time [3.5]. Monolithic integration provides an extremely compact device, and reduces assembly costs and electronic noise. However, these advantages must be traded off against increased complexity and costs.

Typical devices considered for monolithic integration include those using capacitive sensors (where parasitic capacitance effects render hybrid designs impractical) and large arrays of MEMS devices.

3.1.2 Hybrid assembly

The hybrid approach achieves many of the benefits of monolithic integration by combining a number of different chips, usually from different wafers and process technologies, on a common platform, as demonstrated in figure 3.1. In essence the hybrid design keeps the drive and processing electronics separate from the sensing, actuating or emitting structure/device. The hybrid approach is attractive for MEMS integration and packaging because of the ability to support both MEMS and microelectronics on a common substrate without compromising the respective fabrication processes; optimum yields can thus be achieved. Therefore, the hybrid option often provides a more cost effective approach to integration [3.1-3.3, 3.6].
A number of assembly approaches for hybrid integration exists including tape automated bonding (TAB), wire bonding and flip-chip bonding. Although a significant amount of literature has been published on TAB processing, it has not significantly found its way into MEMS processing. On the other hand wire bonding and flip-chip bonding have been quite extensively used.

An alternative approach is to use an overlay over the die with via interconnections fabricated on top of the die as shown in figure 3.2. Two categories of direct metal deposition exist: patterned overlay and patterned substrate. In patterned overlay, the die are embedded in the substrate and the metal connections are patterned above the die. In the case of patterned substrate, the metal is deposited first and the die is then attached on the surface.

An example of direct metal deposition is the General Electrics High Density Interconnect (HDI) process. The die are housed in holes milled in a substrate. Kapton sheets are glued over the top and vias created by laser ablation and metal deposition. The process is repeated until the connections are complete (figure 3.2(a)) [3.3].

Figure 3.2(b) shows an extension of the HDI process known as Chip On Flex (COF). In the COF the overlay is prefabricated and the chips are then attached face down onto the COF overlay with adhesive. Once bonded, a plastic substrate is formed around the die. Connections are made through vias as before. In cases where access to the MEMS/ MOEMS devices is required, windows can be created as a final step.
These processes resolve some issues relating to wire bonding such as cross talk and positioning of bonding sites but do not match the compactness offered by monolithic integration.

### 3.1.3 Integration as applied to optical encoders

The requirement for high precision optical encoders in displacement measuring is well recognised. These encoders are generally bulky and can consist of a large number of optical components such as light source(s), gratings, photodetectors, lenses, beam splitters and mirrors [3.7]. However, low cost compact encoders are preferred for an increasing number of systems with development towards miniaturisation often key to opening encoders to new applications. Advances in optical techniques, such as those discussed in Chapter 2, have gone some way to achieving more compact readheads but device miniaturisation is still difficult due to the number of optical components and the precision required in their positioning [3.8].

Although the integration of optical components is most relevant for achieving compact designs for optical encoders, the integration of the additional electronics can also play a
key role. Aubert et al. explain the monolithic integration of photodiodes and electrical functions, including detection, amplification and some digital processing onto a single chip, using standard CMOS technology [3.9]. Application of such integrated processing electronics can be found in many of today’s encoders, see for example Heiden Hain [3.10]. However, in their design, the discrete optical components, light source, photodetectors and gratings, still require separate assembly. In particular the index/analyser graticule requires precise alignment with the photo detectors, which is a costly procedure. Precise alignment will also be required when assembling the LED so as not to skew the orientation, such that the optical paths from the index to detectors are not significantly affected.

Holzapfel et al. demonstrate the integration of the index and analyser gratings with the photo diodes as shown in figure 3.3 [3.11]. The diodes are formed in a semiconductor on which provision is made for the oxide, metal and passivation layers. The photodetector arrays are formed using standard doping techniques and the metal layer is appropriately patterned to form the gratings. In the centre the bulk semiconductor material is removed by anisotropic etch leaving a cavity for the positioning of an LED. Using flip-chip technology the semiconductor chip is bonded onto a base plate like a PCB, such that an LED is located in the cavity behind the index grating. Despite a requirement for some additional assembly steps, the integration of the grating and photo detectors greatly improves the alignment between the grating and detectors. It also ensures that the stand-off height from the scale grating is constant and fixed for all devices.

Figure 3.3 – Illustration of the Holzapfel optical encoder with monolithically integrated gratings and detectors [3.11]
Hane et al. also approach the compact optical encoder challenge of a micro machined photo detector with grating and detectors integrated into a single chip as shown in figure 3.4 [3.7]. The transmission grating, a series of parallel thin beams, is fabricated from a silicon wafer. A photo diode is installed on each beam using conventional silicon microfabrication techniques. An extended light source is bonded to the backside so that the grating is illuminated from the underside. The major drawback of this design is the cost associated with the manufacturing of the photodiode beams. The high precision requires Reactive Ion Etching (RIE) processing, a costly procedure for low volume and prototyping manufacture. Care must be taken during manufacturing to ensure that stresses in the beams do not cause any adverse effects, such as buckling, during beam release. The device operates using a two-grating principle which suffers from greater sensitivity and lower resolution compared with a three-grating system.

An integrated optical encoder with a monolithically integrated laser diode light source, photo diodes, waveguide and microlenses has been presented by Sawada et al. as shown in figure 3.5 [3.12, 3.13]. The devices use a U- or V-shaped edge emitting laser diode that illuminates a scale grating via a set of waveguides and microlenses. The light is reflected back from the scale onto photodetectors via a further set of microlenses.

![Figure 3.4 – Cross section schematic of micromachined photodetector arrangement](image)

Fabrication of this microencoder is performed using surface processing technologies on a GaAs substrate. Although the emitter and detector components are formed using standard edge emitter/detector processes the fabrication of the on-chip lenses in this orientation is more difficult. The shape of the lenses determines the beam focus and may have some repeatability issues. The waveguides are created by a repeated layering
process and the etching of the mirror is accomplished using RIE. Resolution of the order of 10 nm has been reported for this device.

Figure 3.5 – Isometric illustration of monolithically integrated optical encoder by Sawada et al. [3.12]

In this design, the assembled device has a footprint of approximately 1mm$^2$, much less than that of conventional optical encoders. Each component utilizes the mirror side walls, which are fabricated using a dry etched process, requiring tight manufacturing tolerances to avoid optical losses and ensure equal optical path lengths from the two beams. Furthermore this design is highly susceptible to jitter and requires accurate positioning when mounting against the scale grating.

To tackle this issue of sidewalls Miyajima et al. have proposed a surface emitting optical encoder working on the same operating principle [3.8, 3.14]. Although predicted to become fully monolithic in future developments the example presented uses individually assembled components. The photo detectors are fabricated on a silicon substrate onto which a submount containing the vertical cavity surface emitting lasers (VCSEL) and micro lenses is assembled. Future developments include the monolithic integration of the light source and the micro lenses with the photo detectors. A resolution of 0.2µm is reported for this device.

The last two devices discussed offer the most compact design. However they utilise monochromatic light sources in a single grating system relying on a small spot size and
accurate alignment to focus on the detectors. In comparison a three-grating system allows for greater tolerance by producing a larger fringe pattern and thus an increased detection region. Given equal dimensions, the three-grating system also provides a doubling of the resolution compared to a single or double-grating system.

Perhaps the most relevant and successful integration is provided by Numerik Jena’s Epiflex measuring module and their LIK series encoders as shown in figure 3.6 [3.15-3.17]. This design addresses the demands for better performance, improved accuracy and reliability while providing a smaller more flexible product. The Numerik Jena encoder uses patented Chip-On-Glass (COG) technology to produce an encoder that is only 6mm thick and weighs only 3.2 grams. By using COG Numerik Jena has successfully integrated the light source (an LED in this case), the photodetectors and some electronic components onto a common borosilicate glass substrate, facilitating thereby an extremely compact construction.

The glass substrate, measuring 8 x 14 x 1mm, has gradations etched into a polished stainless steel strip on the bottom side facing the scale forming the index and analyser gratings so that a three-grating system is employed. The optical chip, containing the LEDs and photo detectors, and an application specific integrated circuit, ASIC, with the electronic components are assembled to the glass substrate using flip-chip techniques. The space between the optical chip and the ASIC created by the bump stand-off heights are filled with a suitable underfill material. Electrically conductive tracks are laid on the substrate top side in order to provide electrical connection of the various optical, electrical and opto-electrical components. A flexible flat cable at the side of the substrate, connected also connected to the conductive interconnect tracks, is used for signal transfer.

A major contribution to achieving miniaturisation is the integration of the LEDs and photo detectors onto a single chip. Two LEDs, each centrally located within a square of several incremental photo detector elements positioned directly next to each other as illustrated in figure 3.7, are used to form a multi scanning arrangement. The arrangement of the photo detectors around the light source is such that the sum of the optical paths is as identical as possible. The relative phase positions (0°, 90°, 180°, 270° in this example) on the various photo detectors as a result of the scanning motion are presented in figure 3.7.
The respective fields generating these phase positions are the result of appropriately patterned scanning/analyser gratings corresponding to a respective photo detector rather than the position or orientation of the diode itself. Such a configuration allows for a larger scanning field which reduces the sensitivity of the optical path to particles. In addition two reference scanning units are employed on either side of the measuring unit, each with individual light sources, to further improve performance and reduce sensitivity.

The technical data presented for the LIK 21 encoder module also indicates that a grating period on the scale of 20µm with an accuracy of ±1µm. A digital output achieves resolutions down to 50 nm.

The main drawbacks to the Jena design are twofold. Firstly, there still remains the issue of assembly of the optoelectronic chip to the glass carrier such that the critical alignment between gratings and photodetectors is achieved. Although flip-chip technologies go some way to improving alignment capabilities the procedure is still complex and costly. Secondly the integration of the light source and photo diodes onto a single chip is not performed from monolithic manufacture. In most COG designs it can be difficult to mount the light source near to or directly onto the glass. A contact is likely required on the die surface. Here Jena have etched a cavity into the silicon photo detector substrate, such that the LED die can be assembled in the cavity using a chip-in-chip arrangement. This keeps the source and detectors coplanar enabling a wire bond with minimum loop height, as shown in figure 3.8.
Figure 3.7 – Photodetector array layout positioned around central light sources. Also shown are reference detectors and their associated light sources.

The resultant chip is then flip-chip bonded to the glass substrate via stud bumps that ensure clearance to the wire bond. Although compact and flexible, allowing for assembly of various semiconductor LEDs (Si, GaAs, InP, InGaAs), it does require the individual assembly of each LED within the detector substrate. Furthermore, without access to automated wire bonding equipment, performing the wire bonding procedure itself poses a challenge to maintain repeatable, reliable bonds that do not exceed the maximum loop height.

Figure 3.8 – Cross section schematics showing (a) the optical chip and (b) the optical chip flip-chip bonded to a glass substrate for a COG package solution.

More recently, a position sensing device has been reported by Giannopoulus et al., of the Micro and Nanotechnology Laboratory at the University of Illinois [3.18], as shown in figure 3.9. The integration of a VCSEL light source with PIN photo detectors onto a
single substrate chip measuring less than 1mm$^2$ is demonstrated. The VCSEL mesa measures 25μm in diameter with 9μm diameter oxide aperture. The VCSEL acts as a point source emitting at 850nm to illuminate a corrugated grating (scale) of period 4λ and amplitude of 2λ, giving a resolution approximately equal to the wavelength. The reflected light is measured by the PIN detector diodes positioned around the VCSEL with various layouts and geometries presented.

![Figure 3.9 – Images of integrated 25μm diameter mesa VCSELS and photo detector position sensor arrangements produced at the University of Illinois [3.18]](image)

Characterisation of the device emission and detection capabilities have been performed. Although the devices have been fabricated, the position sensing capabilities have not been demonstrated. However simulated results are provided for three layout geometries modelled against the grating described above at a 63μm stand-off. A first geometry suggests signal variation of 58% but the overall power and thus the signal is very small compared to the two other geometries, which have an order of magnitude increase in power but only a 5% and 14% variation in the signal.

Besides the signal strength to signal variation problems some other fundamental issues exist. Firstly the stand-off distance (distance between the scanning unit and the scale grating as they are moved relative to each other) of 63μm is incredibly small and highly impractical in real applications. Secondly the layouts are such that only two signals with a phase shift of 180° are measured. This means, that although the device maybe suitable for measuring position over a long range, the actual direction of travel cannot be determined. The accuracy is further compromised by the absence of a reference scheme.
3.2 Assembly

MEMS/MOEMS packaging should not only protect the components but also provide interconnections for electrical signals and, in some cases, access to and interaction with the external environment [3.19]. The optical encoder, for example, requires protection from the environment but also the interaction with the external scale grating. In this section, two electrical interconnection technologies, namely wire bonding and flip-chip bonding, along with failure mechanisms associated to each technology, are discussed. In addition, techniques employed to optimise packaging for achieving compact designs are explored. Furthermore a brief review is made of packaging encompassing optical windows.

3.2.1 Wire Bonding

Wire bonding is a 40 year old mature technology with a sound infrastructure which has been carried over from the IC to the microsystems industries. The chips are attached to a substrate as illustrated in figure 3.10, using an adhesive material such as solder perform or epoxy, and fine wires bonded between the contact pads on the chip and substrate to form the interconnections. Two techniques are commonly used for wire bonding: thermocompression bonding and ultrasonic bonding. In thermocompression bonding gold wires are usually used to create the metal to metal contacts via the application of heat and pressure. Aluminium wires are often used in ultrasonic bonding. The ultrasonic process is employed in cases where the high temperatures of thermocompression bonding can have detrimental consequences. The most common metallurgical systems are Au-Au, Au-Al, Au-Cu, Au-Ag, Al-Al, Al-Ag, Al-Ni and Cu-Al [3.2, 3.6, 3.20-3.22].

The wire bonding process is fast when fully automated and fine pitch such as that shown in figure 3.10(b) can be achieved. Thus this technique is very flexible to new devices and bonding patterns. For these reasons wire bonding is often chosen at the early design stage but may not be employed as the final interconnect method because of design restrictions whereby all bond pads are required to be at the chip periphery and highly automated flexible wire bond equipment can incur high capital cost. The process may also be adversely affected by the die attach material, allowable wire lengths and
bond pad cleanliness. The quality of the bond itself will be process dependent and is a function of the heat and pressure applied in thermocompression bonding or the energy applied during ultrasonic bonding. These issues are discussed under failure mechanisms in section 3.2.3.

3.2.2 Flip-chip bonding

Although first introduced in the 60’s, interest from industry remained relatively weak until recently, with a growing demand for finer pitch, higher density and more compact solutions driving interest in this technology. Consequently recent years have seen a lot of attention on the research and development of flip-chip bonding and some significant advances made in the use of different materials and techniques. Flip-chip bonding involves the bonding of a die, top-face-down, onto a package substrate, as illustrated in figure 3.11. Electrical interconnections are made using a chosen bumping method between bond pads on the die and metal pads on the package substrate. The spacing between the chip and the substrate can be accurately controlled by careful construction of the bond pads and interconnects, which are generally less than 200µm, as shown in figure 3.11(b), and so offer low inductance.

Unlike wire bonding, which requires the pads to be positioned around the chips periphery, flip-chip allows the placement of pads anywhere on the die, thus increasing

Figure 3.10 – Wire bonding principle showing (a) an cross-section illustration of a chip wire bonded to an IC board; note that a globular encapsulant (usually epoxy based) is sometimes employed over the chip and wires for added protection and reliability, and (b) SEM photo of multiple wires bonded in fine 50µm pitch [3.21]
the number of I/O connections and reducing the real-estate footprint. Once flipped, thermocompression, thermosonic, Infra Red (IR) reflow or UV curing processes take place to melt the interconnection metals to the pads. An added advantage can be found when performing the flip-chip process under reflow conditions. Increased alignment accuracy is achieved due to the self-alignment of the bumps during the cooling down phase of the bonding process as the surface tension pulls the chip into alignment [3.2, 3.6, 3.20-3.24]. Unfortunately the work in this thesis was unable to take advantage of this self-alignment process and relies on the positioning accuracy of the flip-chip bonding equipment.

![Flip-chip bonding principle](image.png)

**Figure 3.11 – Flip-chip bonding principle (a) illustration of chip flip-chip bonded on to IC module (chip usually face down) and (b) cross-section image of flip-chip bumped device [3.23]**

There are many construction and technological solutions applied in production. The various combinations attainable have associated advantages and disadvantages and the selection is tailored for specific requirements such as the composition of the bump (metal, solder, paste or adhesive, ridged, soft or elastic), how the bumps were formed (electroless or electrolytic plating, printing technology, evaporation or sputtering and solder-jet dispensing) and how the interconnections are made (thermocompression, ultrasonic, UV curing or IR reflow).

### 3.2.3 Failure mechanisms of interconnect technologies

Careful consideration must be put on the metallurgy involved with the interconnections, in particular the adhesion properties of one metal to another and of the seed layer to the substrate or packaging. Critical compatibility issues include conductivity, bondability, hardness, corrosion resistance, dimensions (for both bonding and load issues), shear and tensile strengths, and coefficients of thermal expansion. Good design can prevent many
of the potential failure mechanisms such as flexure fatigue, bond to bond pad shear fatigue, bond pad substrate shear fatigue, voiding, corrosion, noise, vibration fatigue, resistance change and bond pad cratering. The encoders are to be placed into a number of environments where they will be subject to many of the conditions, which may cause some of these failure mechanisms to occur. For example, vibration occurs as a result of a scale disc rotation and temperature fluctuations as a result of factors such as mechanical friction producing heat or operating environment. The application field may be of consequence: for example if the encoder is to be used in space or subsea applications, different operating principles may apply. Therefore consideration of the potential failure modes, reliability issues and causes has to be considered.

3.2.3.1 Wire bonding

Wire bonding, as a mature interconnection technology, has benefited from numerous reliability studies. Although many mechanisms for failure are identified, they can generally be overcome or reduced by careful design and tight process controls.

The first failure mode stems from the cleanliness of the bond pads. Impurities on the bond pads are a major source of metal to metal bond adhesion failure. Contaminants include:

- process debris: etching, outgassing, resist stripper, solvents
- plating contaminants
- organic contaminants
- others that corrode or inhibit bonding

These causes of failure can be minimised by ensuring that the bare wafer/device is only exposed to a cleanroom environment. In addition, the bond pads can be cleaned before bonding by plasma or ultraviolet ozone cleaning. Failures during or as a result of the bonding procedure can be caused by a large number of factors including:

- Cratering: this is caused by damage to the substrate in the area under the bond pad. Common causes are high ultrasonic energy, too high or too low bonding force, impact velocity of tool to substrate, and wire hardness.

- Wire fracture and lift-off: this failure occurs when cracks form in the base of the wire bond due to weakening from overworked ultrasonic welding. Common causes are a sharp edged bonding tool, vibration before or during tool extraction,
too sharp a gradient of the wire loop or rapid tool movement. An example of lift-off is provided in figure 3.12.

- Wire length: a common problem in wedge bonding; too short wires results in increased force on the initial bond area when bonding the other end, too long wires provoke shorting between pads. Common causes are incorrect wire tension, faulty wire clipping mechanism, dirty wire path and incorrect feed angles.

- Peeling: when the wire rather than breaking fully lifts of the surface. Usually caused by poor design parameters and tool degradation.

The issue of bond failures really comes down to good design and careful process control. Enough is known about these potential failures to work the solutions into the process. Good maintenance of tool equipment also helps.

Failures may also occur during the lifetime of a device, so accelerated life tests are often implemented to assess the reliability of connections. Wire flexure fatigue and wire lift-off for example may not occur until thermal cycling tests or during operating life time. Lift-off as a result of poor mould compound conditioning or a compound that may be out of date can occur during the thermal cycling process (usually between -55°C and 125°C). Flexure failures can occur during the operation life time as a result of temperature changes causing expansions and contractions which, in turn, cause small defects to propagate through the wire leading to eventual failure. Local thermal cycling at the wire base, caused by friction as a result of different coefficients of thermal expansion between the pad and the wire, can have the same consequence. Other reliability failures include:

- Corrosion: this failure occurs by the separation of one or both ends of the wire allowing it to move and cause short circuits. Corrosion occurs due to the presence of moisture and contaminants that increase the resistance of the wire bond until device failure.

- Lead frame corrosion: this is again due to the presence of moisture and contaminants induced by surface plating to protect the base metal.
• Metal migration: This is a wear out mechanism that is essentially an electrolytic process whereby ions migrate from the anode region to the cathode region. The result is an increase in leakage current between the bridged regions or the formation of a short circuit.

• Vibration fatigue: Vibration fatigue is not a function of direct vibration to the system. This failure relates to frequencies that might induce resonance and thus damage bonds. Generally this occurs during ultrasonic cleaning so it is advisable to use a cleaner with a frequency range greater than the minimum required for resonance of the wire bond.

• Electronic noise: A lack of oxide layers under the bond pad results in poor insulation of the substrate, electrical leakage failures are the outcome.

Many of the drawbacks of wire bonding are addressed by the flip-chip bonding technology and include connection density, low inductance, better noise control, smaller device footprints and lower packaging profile [3.20-3.22]. However issues directly relating to the interconnections of flip-chip bonding (section 3.2.3.2) still cause enough concern such that wire bonding becomes the preferred approach.

3.2.3.2 Flip-chip Bonding
A key consideration is the use of flux in some solder alloys. Fluxes contain weak acids that become active at the reflow temperature to effectively clean the solder surface. Following reflow there is a flux residue that requires removal. However this residue can interfere with optical signals in optoelectronic assemblies, reducing or eliminating
the optical signal. Therefore this method of flip-chip bonding may not be suitable for some micro optical systems.

The alternative is to seek a compatible fluxless process. Combinations of certain materials allow for fluxless reflow whereby a temporary tack is put in place which acts like a glue replacing part of the flux function. The reflow is then performed in a nitrogen inert oven. The main problems associated with this process are the cost and maintenance of the nitrogen chamber. Thermocompression bonding or thermo-sonic bonding offer the advantages of a simple fluxless dry process, low processing temperatures, high operating temperatures, no lead content and very fine pitch connection. In addition thermosonic bonding also aids the simplification of assembly steps, an increase in the number of choices for contact materials, and a reduction in assembly times, pressures and temperatures. Conductive adhesives can be used in a fluxless process, however, they may suffer from outgassing issues related to underfill materials as described below [3.6, 3.20-3.22].

Underfill materials, usually epoxy based, are often employed mainly in solder reflow processes, to deal with reliability issues associated with coefficient of thermal expansion mismatch between the die and substrate and to reduce the stress placed on the joints during the bonding process. However, outgassing issues relating to underfills generally make them incompatible with devices such as micromirrors where even small amounts of contamination can be detrimental. Underfills are also incompatible for devices with micro structures on the surface. Again thermocompression bonding is an attractive alternative due to the lack of organic materials used in the process but the device must be able to withstand the elevated temperatures required.

Although flip-chip assembly allows for greater I/O density, shorter leads, lower inductance and better noise control from its interconnections, careful consideration must still be given to the metallurgy of the bonds. Many of the issues related to the interconnections of the wire bonding process transfer to the flip-chip bonding process. As previously discussed, there are various flip-chip bonding techniques and, as such, the configuration of the bump must comply with the applied technology.

For solder bumping the main bumping processes are evaporation, electroplating and printing solder pastes, see table 3.1 for details. The flip-chip soldering process requires
the deposition of an Under Bump Metallurgy (UBM), which must provide good adhesion to the bond pad, act as a diffusion barrier during the solder process, provide good wettability of the solder material and prevent oxidation. The UBM depends on the bumping process used. It is also important to consider the compatibility of the solder material with the substrate material i.e. some substrates may not be compatible with the temperatures required for the reflowing of the solder material.

Possible failure mechanisms include corrosion at the joints or metal layers and metal migration caused by electric fields or thermal gradients in the structure. The main reliability concern is the thermal fatigue of the solder joints. Thermal fatigue depends on the difference in CTE’s between chip and substrate, the solder material and the height of the joint as well as a range of environmental temperature factors. These problems can be exacerbated by defects such as voids and cracks as shown in figure 3.13. Underfill is often used to drastically reduce the effects of thermal fatigue on reliability.

The preferred bump material for thermocompression or thermosonic bonding is gold electroplated or stud bumps. The bumps are fabricated using techniques based on conventional wire bonding procedures, thus expensive equipment is not necessary. The substrate must be able to cope with temperatures around 300°C for thermocompression and 150°C -200°C during bump formation.

The reliability issues are very similar to those of bonding with solder. Damage is most likely to be caused by the cooling down phase after the high bonding temperatures. As gold has a higher melting point than solder, the fatigue damage due to thermal cycling would not be a problem if the adhesion strength between bump and pad were not exceeded. Again the use of underfills would significantly increase the reliability of the joints.

Conductive underfills come in two forms: Isotropic Conductive Adhesives (ICA’s) or Anisotropic Conductive Adhesives (ACA’s). The initial bumping requirements are similar to those for thermocompression bonding with the addition of the conductive adhesive material. ICA’s can also be used as the bump material but care must be taken with the pad material used to avoid oxidation.
When applying ICA’s shorting must be prevented as the adhesive will wet to both metallic and non-metallic surfaces. The chip placement is time dependant on whether the ICA is thermosetting or thermoplastic. Thermosetting ICA’s can be applied then stored for bonding later. Thermoplastic ICA’s should have a minimum time span between application and bonding.

The risk of short circuits using ACA’s is minimal because the joints are only conductive in the z-direction. This also means that co-planarity is required between substrate and component. Also most ACA’s are thermoplastic and attention must be paid to the combination of applied pressure and temperature for a specified bump height. Too much pressure on large heights may destroy the conductive particles. Not enough pressure applied to short heights may result in larger contact resistances or failed contacts as shown in figure 3.13(c).

Contamination from moisture in the adhesive layer and in the interface between substrate and adhesive is the common failure mechanism for conductive adhesives. This can have detrimental effects on the adhesion and mechanical strengths, decreases transition temperatures, causes swelling, and corrosion of the bumps and/or pads. Corrosion as with the other bonding methods, increases electrical resistance and can eventually lead to open circuits. There is also a lot of analysis performed on the failure effects with regards to underfills, which will not be discussed here (for more details see refs 3.21 and 3.22).

One of the major disadvantages of flip-chipped components is that testing still poses some technical and infrastructure obstacles due to the inaccessible locations of the bonds. Therefore testing prior to bonding is desirable as inspection after bonding and rework are extremely difficult. The test and burn-in of these known good die is also problematic because of power distribution, cooling and contact problems, and throughput particularly at the wafer level is challenging. Traditional testing of individual chips, using probe card or sockets for example, also run the risk of causing damage to the chip [3.22].
<table>
<thead>
<tr>
<th>UBM</th>
<th>Evaporated</th>
<th>Plating (A)</th>
<th>Plating (B)</th>
<th>Solder Paste</th>
<th>Electroless Nickel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adhesion Layer</td>
<td>Cr</td>
<td>TiW</td>
<td>CrCu</td>
<td>Al</td>
<td>Ni</td>
</tr>
<tr>
<td>Solder Diffusion Layer</td>
<td>Phased Cr-Cu</td>
<td>Cu stud/mini bump</td>
<td>CrCu</td>
<td>Ni</td>
<td>Ni</td>
</tr>
<tr>
<td>Solder Wettable Layer</td>
<td>Cu</td>
<td>Cu</td>
<td>Cu</td>
<td>Cu</td>
<td>Au</td>
</tr>
<tr>
<td>Oxide Prevention</td>
<td>Au</td>
<td>Au</td>
<td>Au</td>
<td>Cu</td>
<td>Au</td>
</tr>
<tr>
<td>Suitability for 63SnPb</td>
<td>No</td>
<td>Poor</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Use with Probed Wafers</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Mixed</td>
</tr>
</tbody>
</table>

Table 3.1: Under Bump Metallurgy for various flip-chip processes [3.21]

Figure 3.13 – Flip-chip failure mechanisms (a) large void (b) bump with crack in the passivation and (c) wasted Ni particle as not enough pressure has been applied to create the bond [3.6]

One solution is to assemble the die onto temporary reusable carriers, which simulate a single chip package including electrical connections that allow traditional IC final test and burn-in infrastructure to be used. Other solutions under investigation include burn-in self test, scan test of inner chip connections, and burn-in and test substrates.

3.2.4 Packaging solutions for flip-chip applications

As the flip-chip technology is a face down bonding procedure potential problems exist particularly for optical devices where access to the optical components is required. Accurate alignment is also vitally important. This is essentially a packaging related problem. Some issues include the flip-chip transfer of MEMS/MOEMS onto transparent substrate [3.25], the use of via interconnects described by [3.26], the
The backside process of a substrate is a developing technique used to ease wafer scale packaging [3.27, 3.29-3.35]. Although many techniques exist it is those developed by Shellcase [3.30 and 3.31], Tessera [3.33] and by Leib et al [3.27] that are of particular interest. Tessera’s wafer level chip scale packaging solution allows Ball Grid Array (BGA) bumping by laying out metal traces on a compliant layer connecting external bump pads to etched areas in a passivation layer exposing contact pads as shown in figure 3.14. The contact pads run along the periphery of the die and are redistributed to the BGA via the metal traces. However it is unclear, from the information obtained to date, how the contact pads are accessed and which side of the semiconductor layer these pads are on.

Figure 3.14 – Example of wafer level chip scale packaging from Tessera [3.33]

Shellcase’s patented ShellOP and ShellBGA technologies employ a method that is not to dissimilar from Tessera. The technology utilises an innovative glass-silicon-glass sandwich structure to enable image-sensing capabilities through the actual packaging structure.

The silicon containing an optical sensitive area is laminated between two sheets of glass and completely encapsulated in epoxy. The electrical contacts are routed to the backside of the silicon leaving the front completely free and exposed for light sensing. The electrical contacts can be laid out to form a BGA type package as shown in figure 3.15.
The routing of the electrical contacts to the backside is an important feature for such a Wafer Scale Package (WSP). It is made possible by the manufacture of notches in the scribe line which go all the way through the backside glass to reveal the cross sections of the pad extensions on the silicon top side. A metal layer, lithographically patterned to individual leads, to contact the pad extensions is deposited, forming a “T-contact” as illustrated by figure 3.16. The leads connect the pad extension layer with the external surface of the package. Once passivation layers and the solder balls are formed the packaged wafer is diced into individual units.

Lieb et al also tackles the issue of allowing full exposure of an optical sensitive topside area through the packaging by routing the electrical connections to the backside, as shown in figure 3.17 [3.27]. In this case rather than notching a second glass attachment, the silicon itself is etched directly, from the backside, using a plasma etching process to reveal a topside insulating layer (typically SiO₂ or SiN), as shown in figure 3.17. Standard contact pads manufactured on the topside of this layer are then accessed from below by removing the insulating layer.

An organic layer is applied for additional passivation and to act as a diffusion barrier. This layer also thermo-mechanically decouples the leads and bond pads from the device. Once the appropriate metallisation has been performed a second organic layer is applied but opened up over the bond pad sites, thus acts as to fully insulate the leads.
The final step is the dicing of the wafer into individual units. As with the previous examples the topside contact pads are distributed around the periphery of the die. The via contacts are used to redistribute the contacts on the bottom side in a BGA fashion. More generally the MOEMS chip(s) are first assembled to a carrier substrate, which can typically be an ASIC or ceramic package, and the optical window is attached over the top as a cap or lid, as shown in figure 3.18. The optical window, usually glass, is attached either directly (glass lid) or the glass window is encompassed within a cap or lid assembly.

![Figure 3.17 – Rerouting of contacts using backside plasma etching to form via holes [3.27]](image)

![Figure 3.18 – Example of optical cap assembly](image)

There are numerous methods used to achieve such packages that involve techniques or combinations of techniques including but not restricted to, FCB, laser welding, laser assisted bonding, ultrasonic bonding, RF sealing, microwave bonding, anodic bonding, epoxy (thermal, UV curable etc), mechanical interlocking, encapsulation, plastic or polymer adhesion [3.36, 3.37].
3.3 Summary

In summary, monolithic integration of optical components onto a single chip, for a specific type of optical encoder, has been reported to produce a small, cheap device with 20nm resolution. However the production of repeatable and reliable devices is in question and the devices have excessively tight sub-assembly alignment tolerances. In the meantime more robust encoders are reported using hybrid solutions but by employing a monolithic ideology i.e. optical components onto a single chip. However the compromise is a larger device requiring complex assembly and packaging procedures that drive up costs.

As a consequence the assembly procedures, wire bonding and flip-chip bonding, are discussed. Flip-chip bonding can facilitate compact assembly but may lead to complications further down the packaging line. The higher level packaging approaches will become evermore important as the assembly and packaging can contribute from 60%-90% of the overall cost so advantages gained through integration maybe lost if appropriate packaging procedures are not available.

3.4 References


Chapter 4

Designs of a Miniaturised Optical Encoder

Further to the problems and difficulties discussed in chapters 2 and 3 optical encoders are also hindered by the critical alignment requirements between the optical gratings and to the photodiode arrays, the bulky nature of the encoder devices and the requested sophisticated packaging, such that optical encoders can be either prohibitively expensive for many applications or unsuitable for others.

Many of these shortcomings can be addressed by employing Microsystems techniques to integrate the key components; optical gratings, light sources and photo detectors, onto a single semiconductor chip. The advantageous of monolithic integration of the optical encoder device include; reduced size, reduced cost, simultaneous alignment of the gratings to the photo detectors over an entire wafer and a reduction in optical interfaces and electrical path and interconnect lengths combine to reduce noise and increase performance.

The successful development and implementation of such an encoder chip requires close collaboration between optoelectronic manufacturers, optical designers and packaging specialists. This work was part of a DTI project with three partners: Compound Semiconductor Technologies Global (CST) Ltd which provided the design and manufacturing know-how of the optoelectronic chip, the company Renishaw plc which brought expertise in optical design and extensive encoder design knowledge and Heriot-Watt University (HWU) via the Microsystems Engineering Centre (MISEC) which offered skills in microsystems design and packaging. The designs of the chip epitaxy structure were therefore carried out by CST and manufactured according to the specifications provided by both Renishaw and HWU.

In this chapter the design concepts, configurations and layouts for an integrated optical encoder chip and the implications these have on the optical design as well as future packaging solutions are discussed.
4.1 Package design concept

The final packaging of the optical encoder chip is largely influenced by the chip design and configuration. In order to appreciate where these influences affect future package design, a concept solution is first offered by way of example before the various aspects of the chip designs are discussed.

In Chapter 3 the merits of both wire bonding and flip chip bonding (FCB) were discussed. Figure 4.1 shows an illustration of a potential packaging solution when one of these technologies is utilised. The Flip Chip on Glass technology (FCoG) employed within a System in Package (SiP) is of particular interest. In this technology, the optical encoder chip along with additional processing electronics chip(s) are flip-chip bonded directly onto the optical glass window as shown in figure 4.1(a). As well as benefiting from the advantages of FCB a chip bonded directly to the glass window offers a reduced metrology loop when mounted in a real system, providing improved mechanical tolerances. This advantage is not readily achievable with a wire bonded solution. The glass itself contains conductive tracking for power and signal transmission and is bonded, FCB or otherwise, onto the packaging substrate.

Figure 4.1 – Illustration of packaging solutions employing (a) FCoG technology and (b) wire bonding technology
Although introduced by IBM in the 60’s the adoption of FCB has until recently been relatively slow but is now being driven by the demand for increased contact density and more compact devices. On the other hand wire bonding is much more common and thus a more mature and readily available technology. A FCB solution for the optical encoder chip requires development particularly for relatively large InP chips so for this reason a wire bonding solution, shown in 4.1(b) is also considered. In each example Low Temperature Co-fired Ceramic (LTCC) is shown as the packaging substrate. However, this could alternatively be a Printed Circuit Board (PCB) carrier, or other appropriate carrier. LTCC is an attractive solution because it offers multiple layers of interconnection with the option of embedding passive components or indeed embedding die whilst keeping the device footprint to a minimum [4.1, 4.2].

The package could also contain some processing electronics to perform electrical amplification on the signals output from the photo detectors and some cabling for power delivery and signal communication with the downstream evaluation unit. Traditionally each signal requires an individual connection. In this example we propose to reduce the cable connections by time multiplexing the signals down a set of twisted wire pairs. A more ambitious scheme would be to utilise the benefits of optical transmission, such as no electromagnetic interference (EMI). This could be achieved by including an additional light source to the chip, such as an edge emitting laser, to couple I/O optical signals to an optical fibre. The signals could then be passed from the processing electronics back to the encoder chip and then transmitted down the optical fibre in time multiplexed fashion, as with the twisted wire pairs. The small increase in encoder chip footprint is offset by a further reduction in cable contacts required to the ASIC or package. Furthermore fibres are light weight, flexible and lower in cost, compared to conventional copper cables. To this end some chips incorporating an edge emitting laser are also produced and the performance of the chip encoder and laser evaluated. These results can be reviewed in appendix A. Further advanced packaging options are discussed in Chapter 9.

The impact of chip design and configuration is significant in terms of the packaging solutions that can be implemented and, as such, packaging options must be considered early in the process to ensure compatibility. In the following sections, the advantages offered by a monolithically integrated chip design over traditional encoder assemblies
are highlighted and the influences of the design options on both operation and packaging are discussed.

### 4.2 Monolithically Integrated Optical Encoder Chip

The optical encoder chip comprises a light emitting diode (LED) and photo detectors fabricated on a InP substrate. Additionally the index and analyser gratings are also manufactured on the wafer using standard lithographic processes. The integration of the optical components lends to the reduction in the overall size, weight and cost of the optical encoder readhead. A major contribution to cost reduction is the ability to systematically align all index gratings to the LED(s) and in particular the critical alignment of the analyser grating over the photo detector arrays across the entire wafer in a single process step. This alignment removes the need for individual placement and alignment of the components, which typically takes around three to five minutes per device during traditional assembly, thus resulting in substantial savings on time and labour.

![Figure 4.2 - Schematic cross-section of the surface emitting chip (not to scale) on an InP substrate for the monolithic integration of the optical encoder components](image)

Figure 4.2 presents a schematic cross-section illustration of the construction of the optical encoder chip including LED and photo-detector elements. The epitaxial structure, detailed in table 4.1, is deposited on the semiconductor substrate using metal
organic phase vapour epitaxy (MOPVE) deposition techniques and comprises two polarised p-type and an n-type regions sandwiching laminated active areas to form junction surfaces for light emission and photo detection. The p-type and n-type layers are electrically connected by gold deposited film electrodes and an isolation layer deposited to ensure contact to only desired locations. A specially designed LED structure is proposed as the light source to obtain enhanced emission to offset the reduced sensitivity of detection resulting from using standard pn-junction photodiodes. By controlling the cavity size, the shape of the output beam is controlled to direct more of the emitted energy into the ideal direction through the index grating. As can be seen from figure 4.2 and table 4.1 the emitting region and detector regions are located at different levels within the epitaxial structure. There is approximately 4μm difference in height between the contact surfaces, such that the LED resides above the absorbing region, which allows for stray light emitted in the direction towards the substrate to be absorbed. However, this also introduces complications in the flip-chip bump development process as discussed in chapter 5. It had been envisaged that this absorbing region beneath the LED could be used to monitor the LED output by measuring the back light absorbed. However this was found not to be possible as the results presented in chapter 7 show.

This structure represents the design of a surface emitting LED encoder, which allows for a choice of substrate materials. In a discrete component assembly a Gallium Arsenide (GaAs) LED is typically used as the light source of wavelength in the order of 850nm, with detection using a Silicon (Si) photo cell. However, as can be seen from table 4.1, an InP substrate is chosen for this design, that emits at a wavelength of 1300nm. InP is selected for one primary reason: to provide an alternative solution in which a simple reordering of the epitaxy layers results in a substrate emitting configuration, the details of which can be found in Appendix B. For a substrate emitting design to succeed the substrate must be transparent at the optical wavelength. GaAs is not transparent at 850nm therefore by switching the substrate material to InP and increasing the operating wavelength to 1300nm such a design becomes feasible. Choosing InP as the substrate material allows for both surface emitting and substrate emitting configurations to be investigated with minimal disruption to the manufacturing process.
On the downside InP is more expensive than GaAs thus increasing the cost per die. The material is also far more brittle and more susceptible to damage during flip-chip bonding. This issue is exacerbated in the case of substrate emitting chips as these require thinning to less than 250µm. However the design of the surface emitting LED device can be transferable to other materials such as GaAs.

Table 4.2 provides a comparative summary of the advantages and drawbacks of the surface emitting and substrate emitting designs. The development of the substrate emitting version of the chip was however suspended due to poor performance primarily resulting from internal reflection issues in the substrate. Further details of the substrate emitting design and test results are presented in appendix B through D.
<table>
<thead>
<tr>
<th>Substrate emitting encoder chip</th>
<th>Surface emitting encoder chip</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ Contact pads located on the opposite side of the chip to the optical path</td>
<td>- Contact pads must be located out with the optical path.</td>
</tr>
<tr>
<td>+ Pad distribution over the entire footprint possible</td>
<td>- Peripheral pads result in increased footprint or reduced detection area.</td>
</tr>
<tr>
<td>+ Can be bonded directly to an ASIC (FCB)</td>
<td>+ Can be bonded direct to the optical window reducing the metrology loop (FCB)</td>
</tr>
<tr>
<td>- Substrate must be optically transparent</td>
<td>+ Adaptable to other substrate materials</td>
</tr>
<tr>
<td>- Internal reflections</td>
<td>+ Absorption region located below emission region help prevent stray light</td>
</tr>
<tr>
<td>- Requires chip thinning</td>
<td>- Optically transparent underfill may be required</td>
</tr>
</tbody>
</table>

Table 4.2: Advantages and disadvantages of substrate emitting and surface emitting optical encoder chip designs

4.3 LED and Detector Layouts

Initially three detector array configurations were considered with various LED options. The three chip variants included a parallel line detector arrangement chosen to represent the miniaturisation and integration of the present readhead arrangement as illustrated in figure 4.3 and two new arrangements namely a pixel detector array (figure 4.6) and a chevron detector array (figure 4.9).

![Parallel detector layout](image)

Figure 4.3 – Parallel detector layout for (a) wire bonding and (b) for flip chip bonding. Note gratings are not shown here.
As discussed in chapter 2 current Renishaw optical encoders employ a silicon (Si) photo detector chip comprising a parallel linear array of sixty photo detector elements measuring 1500 x 40µm with a 10µm spacing. The LED is set to the side of the detector array and located further back from the optical window. It is angled such that the index grating is fully illuminated and in the direction towards the analyser grating ensuring coverage of the detector array. However, by applying this arrangement in the integrated chip, the LED cannot be offset. The index grating is manufactured directly over the LED but this cannot be angled to focus the beam towards the grating; thus the coverage of a relatively large detector array becomes somewhat compromised. Furthermore inspection of figure 4.3 indicates that in order to cater for the pad dimension and pitch restrictions a wire bond design (a) requires a chip some 60% larger in surface area, compared with a flip chip bonding design (b). Although with access to more sophisticated wire bonding equipment it would be equally feasible to wire bond to design (b).

<table>
<thead>
<tr>
<th>Photo-Diode</th>
<th>Wave Form</th>
<th>Amplitude (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td><img src="image" alt="Wave Form A" /></td>
<td>nSinθ</td>
</tr>
<tr>
<td>B</td>
<td><img src="image" alt="Wave Form B" /></td>
<td>nCosθ</td>
</tr>
<tr>
<td>C</td>
<td><img src="image" alt="Wave Form C" /></td>
<td>-nSinθ</td>
</tr>
<tr>
<td>D</td>
<td><img src="image" alt="Wave Form D" /></td>
<td>-nCosθ</td>
</tr>
<tr>
<td>Input</td>
<td><img src="image" alt="Input" /></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4.4 – Set-up example for amplitude error from a single photo detector failure

It also becomes apparent that the positioning of the LED in this parallel arrangement is wasteful of the emitted energy that propagates in the directions away from the detectors. Therefore higher efficiency could be achieved by positioning additional detector elements about a centrally located light source. Furthermore, sensitivity to adverse effects as a result of contaminant particles causing obstructions in the optical path, such as dust, can be reduced by not only increasing the number of locations where the signals
are collected from but also redistributing the locations where each channel collects the signal to minimise error.

Consider the example system shown in figure 4.4 where θ is a given input value into a system comprising four signal channels A, B, C and D. If no defects are present then all amplitudes are equal and the difference between θ and α is zero. In other words no error exists and the four quadrature signals are balanced giving accurate positional information. Now consider the situation where the photo diode channel is affected by a defect such that D now equals \( D = -\frac{n}{n-1} \cos \theta \), where \( n \) is the number of detector repetitions per channel. This could also be thought of in more general terms as a percentage loss to a particular channel due to the defect. Nonetheless the difference between θ and α is no longer zero and an error exists in the system as a consequence.

![Figure 4.5 – Position error as a function of a photodiode not contributing to the total signal over several ABCD repeats. Note normalised to one Lissajous or 0.5 of the pitch](image)

Figure 4.5 shows the effect of the number of discrete detector locations, \( n \), on the position error of a system. Note the x-axis for normalised pitch length reads only from 0 to 0.5 rather than 0 to 1. This is because two Lissajous cycles are present over one length of grating pitch. By way of example this indicates that a sub-divisional error of up to 2nm may be measured from a 20\( \mu \)m pitch scale over a single Lissajous cycle with only sixteen detector locations. This improves to 0.8nm when increased to forty. From
figure 4.5 it can be surmised that an increase in $n$ to minimise the number of affected diodes will result in an increased tolerance to defects.

4.3.1 Pixel Layout

Figure 4.6 – Optical encoder chip pixel layout design (note gratings are not shown here)

Taking the above discussion into consideration a first new layout arrangement is presented in figure 4.6. The concept comprises of two banks of photo detectors, with the individual detector diodes arranged in a pixel like fashion, positioned either side of a centrally located LED. Electrical signals, that are common in phase, are combined where possible on the chip as indicated by the diagonal lines connecting square detector pixels with subsequent connection being performed of chip. Square pixels have been employed but other geometries such as circles or hexagons could also be used.

Figure 4.7 – Illustration of a diode subcell unit where the letters A, B, C and D indicate the relative phase shift to the left as we move down from the top row
The detector blocks are further considered as subcell units of length R, as highlighted in figure 4.7. The individual detector diodes measure 100 x 100μm with a 30μm spacing. Once again, the dimensions are not limited to these and smaller pixels and more subcells could easily be implemented. Within each four-by-four subcell array there are four detector diodes contributing to each of the four phase shifted (0°, 90°, 180°, 270°) signals represented by the letters A, B, C and D, respectively. Given these dimensions the chip footprint is reduced by 50% compared with the parallel flip chip arrangement. However this also means an approximate reduction in photo detector area of 70%. However as discussed in section 4.2 the integrated chip incorporates a specially designed LED structure. Therefore the aim is to compensate for this loss of detector area with an increase in optical power at the source. As a rough guide a simplified optical power budget is provided in table 4.3. Neglecting the application of index matching optical coatings the table provides an estimate of the optical power requirements just after the LED to achieve a photo current of 650nA equivalent to that of current Renishaw encoder products.

Numerous assumptions, such as the coupling efficiency and sensitivity, have been made to perform the calculation. Considering this and bearing in mind the reduced photo detector area along with many other factors not considered in the table an estimated LED output power in the region of 650μW to 800μW may be required in the worst case scenario.

The phase associated to an individual photo detector diode is determined by the interaction of the light with the overlying analyser grating pattern. In this arrangement the phase shift is considered for each subcell row, such that the grating pattern applied is shifted by a quarter period to the right (or left if desired) for each array as the rows are transcended from top to bottom, as highlighted by figure 4.8. This ensures 90° phase shift and rotation of the array enables on chip interconnection of as many detector diodes as possible via the diagonal interconnects. Where interconnection cannot be performed on chip, without compromising the layout or expanding the chip footprint, the electrical signals are further combined on the substrate assembly.
### Table 4.3: Non-optimised simplified power budget requirement for 650nA per channel

<table>
<thead>
<tr>
<th>Path</th>
<th>Efficiency (%)</th>
<th>Efficiency Cumulative (%)</th>
<th>Optical power after element (µW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power in InP substrate from LED</td>
<td></td>
<td></td>
<td>205.9 µW</td>
</tr>
<tr>
<td>Transmittance from substrate to air</td>
<td>72.44</td>
<td>72.44</td>
<td>149.17</td>
</tr>
<tr>
<td>Index Grating</td>
<td>50</td>
<td>36.22</td>
<td>74.58</td>
</tr>
<tr>
<td>Transmittance from air to glass</td>
<td>94.9</td>
<td>34.37</td>
<td>70.78</td>
</tr>
<tr>
<td>Transmittance from glass to air</td>
<td>94.9</td>
<td>32.62</td>
<td>67.17</td>
</tr>
<tr>
<td>Reflectance from scale grating</td>
<td>50</td>
<td>16.3</td>
<td>33.58</td>
</tr>
<tr>
<td>Fraction of energy in 1st order diffraction path</td>
<td>38.3</td>
<td>6.25</td>
<td>12.86</td>
</tr>
<tr>
<td>Transmittance from air to glass</td>
<td>94.9</td>
<td>5.93</td>
<td>12.21</td>
</tr>
<tr>
<td>Transmittance from glass to air</td>
<td>94.9</td>
<td>5.63</td>
<td>11.58</td>
</tr>
<tr>
<td>Analyser grating</td>
<td>50</td>
<td>2.81</td>
<td>5.79</td>
</tr>
<tr>
<td>Transmittance from air to substrate</td>
<td>72.44</td>
<td>2.04</td>
<td>4.2</td>
</tr>
<tr>
<td>Channels per array</td>
<td>25</td>
<td>0.51</td>
<td>1.05</td>
</tr>
<tr>
<td>Optical power per detector</td>
<td></td>
<td>0.51</td>
<td>1.05 µW</td>
</tr>
<tr>
<td>Coupling efficiency</td>
<td>90</td>
<td>0.46</td>
<td>0.944</td>
</tr>
<tr>
<td>External quantum efficiency</td>
<td>81</td>
<td>0.37</td>
<td>0.76</td>
</tr>
<tr>
<td>Responsivity (typical value)</td>
<td>0.85A/W</td>
<td>0.32</td>
<td>0.65</td>
</tr>
<tr>
<td>Photocurrent</td>
<td></td>
<td></td>
<td>650nA</td>
</tr>
</tbody>
</table>

Table 4.3: Non-optimised simplified power budget requirement for 650nA per channel

Figure 4.8 – Illustration of the analyser grating shifted by a quarter period as we move down the rows to implement the phase changes of the detected signals
The resultant chip design offers compactness having a footprint of approximately 4mm$^2$. Collection of signal from numerous locations means that adverse effects resulting from contaminant particles, such as dust, causing obstruction in the optical path have a reduced impact.

4.3.2 Chevron layout

The second concept design is presented in figure 4.9 and again employs two banks of photo detectors positioned on either side of a centrally located LED. However, in this case, the pixel arrangement is replaced by a series of skewed photo detectors. Essentially a skewed photo detector replaces a number of pixel diodes that are connected in together via a diagonal track with a single photo detector element. Therefore the layout retains the advantages of the pixel regime but the photo detector diode geometries results in an increased overall detector area, approximately 70%, for the same given footprint, thereby generating more photocurrent per mm$^2$ of usable detector area. The area contributing to each phase shifted array A, B, C and D is equal and once again the final combination of electrical signals is performed on the substrate assembly.

![Figure 4.9 – Chevron detector layout. (Note gratings are not shown here)](image)

In this case the analyser grating must cover the entire skewed detector plane while implementing the quarter period shift to ensure the correct phase detection. This is
achieved by slightly offsetting the grating in relation to the period and the length of the array such that the angle of rotation for each feature is,

\[
\tan \theta = \frac{p_a}{R}
\]  

where \( p_a \) is the period of the analyser grating and \( R \) the length of one ABCD range. Figure 4.10 shows the superimposed images as a result of interaction with a skewed analyser grating. It is shown that an angle matching that of the photo detectors is achieved.

An additional advantage of this design is the alignment of the index and analyser gratings. By using a back side alignment process the alignment procedure becomes a trivial issue as the tolerances for the linear (X-Y) alignment of the gratings is much larger compared with previous designs. The angular tolerances are the same but these are already easily achieved.

Further advantages can be found from flipping one of the detector banks to achieve the chevron layout. This improves tolerance to defects that have the same orientation as a diagonal detector stripe so that it only completely eliminates the photo detector stripe on one side rather than both. Furthermore, as well as improving the manufacturing alignment tolerances, the mechanical tolerances for the readhead set-up are also improved upon, resulting in reduced set-up times.
4.3.3 Reference markers

Reference markings on the measuring scale grating at defined positions generate a reference pulse signal for detection at some reference detector elements. Various coded markings can be employed to provide absolute position information on the scale. A reference marker is an essential feature of any encoder product and the inclusion of a reference scheme is considered, but only after the incremental function had been verified.

A number of schemes are investigated, including auto-correlation, cross correlation and split diodes. The implications to the design are the addition of process steps during chip manufacturing. This could include an increase in the number of bond pads (dependant on scheme) and an increase in the chip footprint again dependant on the scheme applied and whether the reference detectors can be located inside or outside the periphery of the incremental detector bank. Consequent changes to the chip layout are applied as appropriate and these are discussed in greater details in Chapter 8.

4.4 Summary

The monolithic integration of the optical components onto a single InP semiconductor chip, which is suitable for both flip-chip and wire bonding to a carrier substrate or package, is presented as a solution for miniaturisation of the optical encoder. Signal transfer by time multiplexing of the signals using an optical regime or twisted wire method is considered. The chip structure and novel photo detector layout geometries have been discussed.

Of the proposed arrangements the most attractive layout for the photo detector array is the chevron configuration. This layout has all the multi-sensing benefits of the pixel arrangement but with an increased current per mm² for a given chip footprint. One of the most important developments is the integration of the optical grating onto the chip. All gratings are aligned to all chips on the wafer in a single processing step. The chevron layout offers a particular advantage as the X-Y tolerances for linear alignment in both grating manufacture and encoder assembly are much larger so greatly simplify
the process. The alignment process has been a critical and costly procedure in the past which is significantly reduced through this novel architecture.

Both wire bonded and flip-chip bonding are suitable as a final packaging solution. A flip-chip bonded solution offers additional advantages including reduced path lengths (lower inductance and losses), lower noise and provides the most compact device. Furthermore, employing a flip-chip on glass solution specifically provides a reduction of the metrology loop, when mounting the packaged encoder device to the subassembly. Therefore the ultimate aim is to achieve a flip-chip on glass solution. However to achieve this some bespoke development of the flip-chip bonding procedures is required while wire bonding is readily available. Therefore in the following chapters the development of a flip-chip solution is discussed while initial verification of the chip design and performance is achieved with wire bonded chips.

4.5 References


Chapter 5

Bump Fabrication

As discussed in chapter 4 the prototype optical encoder chips have been designed for compatibility with both flip-chip and wire bonding technologies. As the wire bonding procedure is relatively straightforward, wire bonded chips are used for verification and characterisation (results in chapters 7 and 8), before a suitable flip-chip bonding solution is developed. The first phase of providing a flip-chip bond solution is the fabrication of the bump interconnects.

5.1 Technology selection

Table 5.1 provides a brief summary of some of the advantages and disadvantages associated with common bump technologies. These technologies utilise three basic interconnection processes. The first interconnection process is the controlled collapse of solder bumps to wetted pads. In this process, the temperature is raised such that the solder is in a liquid state allowing it to wet and react with the bond pad. The stand-off distance between chip and substrate is controlled by the volume of solder, pad dimensions and surface tension. Multiple die can be flipped onto an IC along with other surface mount components and the whole board is placed into a reflow oven so that all components are welded at the same time. One of the major attractions of this technology is the self alignment properties induced as a result of the surface tension between pads and solders [5.1-5.4].

The second process is the solid state bonding between the bump and pad metallisation. Bonds are formed by applying a combination of heat and pressure to overcome the surface inter-diffusion barrier using either thermocompression or thermosonic bonding techniques. In this case stud bumps or electroplated bumps are used. In many applications high temperature and bonding loads are undesirable. In such circumstances conductive adhesives are used in conjunction with electroplated bumps or stud bumps. However there are no self alignment properties and more accurate placement equipment is required [5.1-5.3].
A third approach is a combination of both of the above where a solder cap reflow is applied with electroplated bumps or stud bumps. In this case a low melting point
eutectic solder is applied either directly on the bump, on the bond pad or both. This enables lower bonding temperature and forces to be applied [5.1-5.4].

In order to select an appropriate technology the limitations connected with chip design and application is considered along with the accessible process equipment. The new optical encoder chip has bond pads of 100 x 100μm and, given the correct UBM, such pads are compatible with all technologies. However, the tight pitch, down to 30μm, places constraints on the bump formation techniques available. Conventional approaches, such as solder dispense and stencil printing, are either incompatible or prohibitively expensive to employ. Electroplating or evaporation of solder bumps is possible but another problem arises with the use of flux. Flux is used as an activation agent in the bonding process to remove residual oxides. A residue is left behind, which is an undesirable by-product in the case of optical devices that can provide a source of contamination and is thus not suitable for use in optical encoder applications. Fluxless reflow solder processes are available but these require additional process steps and pre-clean treatments incurring additional assembly and capital costs [5.5-5.6].

The use of conductive adhesives with electroplated or stud bumps is also ruled out because epoxy bleeding leading to shorts, insufficient conductive particles and varying conductivity. This limits the minimum pitch attainable to dimensions greater than 200μm [5.7, 5.8].

Solid-state bonding on the other hand is well suited for the flip-chip bonding of the optical encoder chip. With access to the Karl Suss FC6 flip-chip bonder installed at Heriot-Watt University the thermocompression bonding technique was selected. As well as thermocompression, the FC6 also offers IR reflow functionality. However, reflow processes are typically performed in reflow ovens with an inert atmosphere, such as nitrogen. The lack of inert atmosphere on the FC6 may result in contamination and poor joints. Therefore the reflow process is not considered for this work.

The preferred bump material for thermocompression bonding is gold [5.2, 5.9]. In industry gold bumps are usually combined with conductive adhesives to reduce applied bonding forces. This is beneficial for large chips containing thousands of connections as they are more susceptible to damage at high bond forces. Also, it has already been established that the use of conductive adhesives are not suitable. However, the optical
encoder chips have a low bond count, (fifty or less), in which case direct gold-to-gold solid state bonding provides a reliable flip-chip technique. Additionally this is a lead-free process.

Alternative intermetallics reactions at the bond interface e.g. tin [5.10, 5.11] or indium [5.12] for example, have been investigated with gold bumps to lower the bonding temperatures and loads. However these materials, as with solder, require costly and often complicated multistep UBM processes and may also require some pre-treatment before bonding. This is eliminated with the use of direct gold-to-gold bonding [5.13].

Two methods for gold bump fabrication are stud bumping or electroplating. Stud bumping requires the use of a modified wire ball bonder to place a ball onto the pad before shearing of the wire. As such, this method suffers many of the limitations relating to pitch and size associated with wire bonding. The size of the ball is dependant on the thickness of gold wire used; 25μm wire results in bumps of approximately 75μm diameter and 50μm height [5.8]. For thermocompression bonding a smoother surface morphology gives more reliable joints and thus, to achieve coplanarity, the bump is deformed in process known as coining, shown in figure 5.1b, to remove the tail left over by the wire shearing step, figure 5.1a [5.14]. The coining step can decrease the stud height to 20μm and increases the diameter of the bump. As a large degree of deformation has already taking place, a level of compliance is lost during the joining steps.

![Figure 5.1– Images of ball stud bumping (a) on attachment and (b) after coining [5.14]](image)

Although technically suitable for the bonding process, the stud bumps are ultimately not appropriate for bonding the optical encoder. Referring back to chapter 4 and figure 4.2, the structure of the optical encoder chip leaves a non-planar contact profile of 4μm
between pads. As the height is controlled by the wire diameter and deformation during coining, this height discrepancy cannot be compensated for with stud bumps deposited on the substrate to which the chip is to be bonded. Stud bumping on the chips and coining could be a viable option. However, as InP is brittle, relative to Si, the forces required to achieve the necessary degree of deformation, which are in excess of those required for the bonding process, will cause cracks and fractures damaging the chip. Similarly, temperatures cannot be raised above the annealing temperature (450°) in order to avoid any thermal damage. The solution therefore lies in the electroplating of gold bumps that can be grown to different heights.

### 5.2 Electroplating basics

The art of electroplating, also known as electroforming or electrodeposition, has been used for over 150 years. First developed for producing decorative finishes, such as with jewellery, electroplating has since found new and widespread applications in the electronics industry such as interconnects, transmission lines, via filling, reflective and protective finishes and moulding structures.

In electroplating, a material, usually a metal or metal composite, is deposited by connecting a negative charge on an object, which is immersed in a solution containing a salt of the metal to be deposited. The positively charged ions (cations) in the salt are attracted to the negatively charged object and move through the solution where they are reduced to a metallic form at the surface of the object, while the negatively charged ions (anions) move to the anode [5.15].

In its simplest form, an electroplating system takes the form of the plating bath illustrated in figure 5.2. This consists of the following components [5.16]:

**Electrolyte:** The electrolytic solution is an electrically conductive aqueous, non-aqueous or molten medium that contains suitable metal salts, providing the ions to be deposited. Additives may also be included to improve the quality of the electrodeposition (adhesion, surface finish, uniformity, brightness, etc.) and, ideally, should not become part of the deposit.
Electrodes: A minimum of two electrodes (cathode and anode) are required. The main driving force for the ions is provided by applying an electric field across these electrodes. The positive ($M^{n+}$) and negative ($X^{n-}$) ions deposit on the cathode and anode, respectively. Due to poor stochiometry and adhesion at the anode, and the fact that most metal ions are positive, cathodic deposition is more popular in the electroplating process.

Power supply: The power supply can be (1) direct current a constant voltage which leads to potentiostatic deposition; (2) direct current at constant current which leads to galvanostatic deposition; or (3) a current or voltage waveform or pulse. There is a variety of electrical waveforms that can be used for electroplating, examples of which are provided in figure 5.3.

![Schematic showing a basic electroplating bath set-up for cathodic deposition](image)

Figure 5.2– Schematic showing a basic electroplating bath set-up for cathodic deposition

More advanced baths also include filters, heaters and temperature sensors, some form of agitation apparatus and pH monitors.

On application of an electric field, the $M^{n+}$ ions move towards the cathode and the cathodic reaction can be expressed in simple terms as,

$$M^{n+} + ne \rightarrow M$$

In the event that the electrolyte contains more than one ionic species that can be simultaneously deposited the process can be written as,

$$M^{n+} + ne \rightarrow M, \ N^{n+} + ne \rightarrow N,...$$

or
so illustrating the ability to plate alloys and compounds from a multicomponent systems. However multicomponent systems that may be suitable for the application under investigation, such as Au/ Sn, have a limited commercial offering

![Electroplating Current Waveforms](image)

**Figure 5.3** – Examples of typical electroplating current waveforms (a) D.C. current, (b) forward pulse current, (c) pulse reverse current and (d) a complex waveform example

Faraday’s laws states that substances deposited from the electrolysis are proportional to their chemical equivalent weights. This can be related to the gram-equivalent weight using the Faraday constant, F, (96500 C) and the atomic weight of the deposited metal, M, such that the mass of deposited metal, m, with the amount of applied charges, Q, is expressed as

\[ m = \frac{MQ}{nF} \]  \hspace{1cm} (5.1)

where \( n \) is the number of electrons required in the reduction of one mole of metal. The charge can be determined from the product of the current, I, and the deposition time, t,

\[ Q = \int I \, dt \]  \hspace{1cm} (5.2)

Furthermore the mass can also be related to the metal density, \( \rho \), and the volume of the deposit (thickness, h, multiplied by the area, A):

\[ m = \rho hA \]  \hspace{1cm} (5.3)

By substituting equations (5.2) and (5.3) into (5.1) and rearranging to find the growth rate from which it is also possible to estimate the material deposition.
\[
\frac{h}{t} = \frac{MI}{nF\rho A}
\]  

(5.4)

where \(I\) is the applied current and \(t\) the electrodeposition time.

Equation (5.4) merely provides a good approximation of the deposited mass because part of the current may be used to induce chemical changes not related to the electrodeposition e.g. the current may also be used in the decomposition of the electrolyte or the creation of hydrogen, for example [5.17].

5.3 Gold bath compositions

Electroplated gold deposition in the electronics industry generally divides into two categories: soft or hard gold plating. Soft gold, also referred to as pure gold, is typically used for surface finishes, bonding, joining or in applications where high electrical conductivity is at a premium. For hard gold deposits, additional metal, such as Ni, Cu or Fe, is added to the electrolyte, which results in the co-deposition of a gold/transition metal alloy. Less than 1% transition metal is required to significantly alter the properties of the gold deposit, raising the hardness to more than twice that of soft gold. Hard gold is typically found in applications where robust mechanical contacts are used, such as in push-pull connectors [5.15, 5.18].

A vast number of electroplating solutions exist in industry and the plating solutions can categorise into to five main groups: 1) alkaline gold cyanide for gold and gold alloys, 2) neutral gold cyanide for high purity gold plating, 3) acid gold cyanide for bright hard gold, 4) non-cyanide, generally sulphite based and 5) miscellaneous [5.19]. The development of so many different formulations and systems over the years has been predominantly driven by cost and safety. The cost aspect is not just related to the price of gold itself but also on the performance of the bath in relation to plating speed, volumes of deposit, longevity, control and maintenance as well the capital and overhead costs.

Cyanide baths are extensively used in the electronics industry for both soft and hard electrodeposition. They come in numerous formulations and include alkaline, neutral and acid solutions. The main advantages of cyanide baths over non-cyanide baths are
superior stability, increased deposition rates and long life. Cyanide is also cheap and the gold salts are easily manufactured. On the other hand cyanide baths have poor compatibility with many standard photo resists often causing underplating of the resist layer. However the main disadvantage is that cyanide is highly toxic so health and safety issues associated with its use and disposal are a major concern.

For this reason non-cyanide based solutions have been developed and are predominantly sulphite based. Sulphite baths are less stable than their cyanide counterparts but are still stable enough under alkaline conditions. Although deposition rates in sulphite systems are slower, such systems are compatible with most standard photo resists. They have a better throwing power resulting in more uniform deposits, excellent geometries, low surface roughness, no underplating and most importantly are non-toxic [5.4, 5.15, 5.18, 5.20].

Thiosulphate baths have also been investigated for Au electrodeposition but instability of the thiosulphate ion in the absence of free sulphate has restricted their use as a practical plating solution. On the other hand the stability of the thiosulphate complex is considerably better than that of the sulphite complex, which has led to the investigation of mixed sulphite – thiosulphate solutions. These mixed solutions have been shown to be compatible with most photoresists and produce smooth coplanar micro bumps. However, poor adhesion to the substrate, due to a lower overpotential range compared with sulphite systems, has also been reported [5.15, 5.18, 5.20].

To perform direct solid state bonding, the uniformity of the bump geometries is important and the gold must be soft enough to allow suitable deformation to compensate for any small deviations in height, otherwise performance and reliability might be compromised. Hence a soft electroplating process is essential. The electrolyte chosen for the bump fabrication process, discussed in section 5.4, is ECF 60 purchased from Metalor Technologies UK.

ECF 60 is a sulphite based gold plating solution that has been shown to be compatible with most photoresists while the low stress depositions promote good adhesion and the high throwing power resulting in thick (5μm-30μm) deposits with good height uniformity and low surface roughness. Additionally ECF 60 is cyanide free and thus offers the benefits of safer handling and easier maintenance [5.20].
5.4 Bump fabrication process

Figure 5.4 summarises the bump fabrication processes under investigation. The process utilises UV-LIGA manufacturing techniques to create a resist mould into which the bumps are formed by electrodeposition. The term LIGA is an acronym for the German Lithographie Galvanof ormung and Abformung, which means lithography electrodeposition and moulding [5.21-5.23]. The process can be traced back to IBM in the 1970s with some modifications being introduced at the Karlsruhe Nuclear Research Centre in the 1980s. With the advance of microsystems technologies UV-LIGA became popular in academic research institutions as a cheap alternative to X-Ray LIGA with a trade-off against the achievable aspect ratio, typically 10:1 to 30:1 for UV-LIGA while aspect ratios of 100:1 are readily reached using X-Ray LIGA with 500:1 structures being achievable [5.21-5.24].

In this section the process steps shown in figure 5.4 are discussed in more detail, as are the difficulties encountered and the solutions developed at each stage. While a focus on gold bumps is maintained, two alternatives aimed at reducing the cost by replacing the core of the bump with either nickel or copper have been considered. Such a bump would have an electroplated gold cap finish to allow gold-to-gold bonding to take place. However, due to failures during fabrication and at assembly, the Ni and Cu core bumps were disregarded in favour of pure gold bumps as a solution for flip-chip bonding of the optical encoder chip. Details of this bumping procedure can be found in appendix E and F.

In chapter 4 the benefits of a flip-chip on glass bonding for a surface emitting optical encoder chip have been discussed. Glass wafers are cheap, readily available, easy to handle and allow for visual inspection through the backside (particularly useful for inspecting alignment accuracy) making them ideal substrates for test vehicles. The flip-chip technology developed is also intended to be used, at a later date, with the subsequent packaging material, such as LTCC or silicon.
Figure 5.4 – Process flow for the electroplating of coplanar and non-planar bumps

5.4.1 Substrate preparation

Before any processing can be performed the substrate wafer requires cleaning to ensure that the surface is free from potentially detrimental contaminants. Common methods of cleaning include chemical, electro-chemical and acoustic cleaning. For this process ultrasonic cleaning proves sufficient. Wafers are placed in a carrier, which is subsequently placed in a beaker containing de-ionised (DI) water and a decontamination cleaning agent, Decon 90. The beaker is then placed in the ultrasonic bath, as shown in
figure 5.5 for 2 hours at 50°C to ensure thorough cleaning. Once removed from the holder the wafers are rinsed with DI water to remove any remaining traces of cleaning agent, the surface water is then removed using a pressurised nitrogen gun before being dried on a hotplate at 100 °C for twenty minutes to remove any excess moisture.

![Figure 5.5](image)

Figure 5.5 – A wafer carrier is placed in a beaker, filled with a DI water/ Decon 90 cleaning solution, and placed into an ultrasonic bath

### 5.4.2 Seed layer deposition

![Diagram of Glass Substrate, Ti Adhesion Layer, Au](image)

Figure 5.6 – First generation bump process step 2: seed layer deposition

The seed layer is a metal used to provide electrical continuity across the wafer to facilitate electroplating and may also include additional layers to promote adhesion to the substrate as illustrated in figure 5.6. Chemical Vapour Deposition (CVD) and Physical Vapour Deposition (PVD) are the most commonly used deposition techniques. PVD is primarily used to deposit thin metal layers with two common techniques of sputtering and electron-beam evaporation, both of which can be employed here. In this step, the seed layer not only provides the under bump metallurgy but will form the interconnecting tracks, after patterning, between the bumps and probe pads during the testing procedure.
The seed layer deposition is performed using the electron beam system shown in figure 5.7. The wafers are loaded into the vacuum chamber which is then pumped down to pressures below $10^{-6}$ bar. The electron gun is used to accelerate electrons towards a metal target at around 10keV, locally melting the material. The evaporated metal deposits on the wafer via simple line of sight [5.25].

While evaporation of metals such as Ti, Ni and Cu has been relatively straightforward, the Au deposition proved extremely difficult, primarily due to a lack of deposition control and small volume of the Au target. This had the effect of poor adhesion as shown in figure 5.8. Therefore, Au seed layer tracks were sputtered at CST. During sputtering, the target material to be deposited is at a high negative potential and is bombarded with positive argon ions. The sample material is sputtered away mainly as neutral atoms via momentum transfer and the ejected atoms deposit onto the substrate placed at the anode [5.26]. Additional benefits of sputter deposition included better thickness control, better coverage and greater substrate adhesion.
For the deposition of the gold bump seed layer, a thin 60nm layer of Ti is first deposited to promote adhesion between the glass substrate and a 500nm Au layer. Later it may be necessary to also include a barrier layer of Platinum, Pt, or Tungsten, W, to prevent diffusion of the Au into the Ti.

5.4.3 Photoresist spin coating

The photoresist is applied over the seed layer as illustrated in figure 5.9. A minimum of two photo resist spin coating procedures are required. The first resist deposition is used to define the pattern of the tracks and pads via the selective etching of the seed layer. The resist is applied to the wafer using a spin coater as illustrated in figure 5.10. The thickness of the resist layer is governed by the volume and viscosity of the resist and the speed at which the spin cycles are performed. Typically 1ml of resist is dispensed per inch wafer diameter. The resist is dispensed onto the centre of the wafer to help achieve a uniform distribution. The spinner is then ramped through a sequence of cycle parameters including speed, time and ramp accelerations in order to reach the desired thickness, such as those shown in table 5.2 for a 30µm thick deposition. Often an edge bead can form around the edge of the wafer, which can be undesirable in future process steps. Techniques exist to minimise this effect including adding a further series of spin
parameters designed to reduce the edge bead or the application of a stripping agent in very small doses, applied only to the wafers edge, whilst the wafer is spinning.

Once spun, the wafer was placed on a hotplate and ramped through the temperature cycle shown in figure 5.11 to cure the resist, which promotes suitable adhesion to the substrate and reduces contamination of the mask (resist sticking to the mask) in the contact exposure step [5.27 – 5.29].

The deposition of the second photoresist defines the mould used for the electroplating the bumps. The thickness of the resist is set to exceed the height of the bumps being electroplated. An alternative approach would be to deliberately over plate the bumps and polish back down providing highly coplanar bumps. As gold, by comparison to other commonly plated metals, is relatively soft care must be taken, particularly when mechanically polishing, to avoid smearing and or dislodging the bumps. However, as previously discussed height differences of 4μm are required, thus, an overplating and polishing approach is not a valid solution.
Furthermore the minimum thickness of the bumps and hence resist deposition can be largely dictated by the underfill should this be required. The underfill is an epoxy resin commonly used to improve the mechanical integrity of the bonded joints. It is typically dispensed at the corners of bonded chips and relies on the capillary action to draw in the epoxy. The coverage depends on both the viscosity of the epoxy and the stand-off height (defined by the bumps) between the chip and substrate. Underfill for gaps of less than 10μm is extremely difficult to achieve. Therefore to provide compatibility, bumps of 20μm and 24μm are targeted to ensure that gaps greater than 15μm are maintained. Thus, the resist thickness must be 25μm or greater. Two types of photoresist have been considered: AZ 9260 positive photoresist and THB 151N negative photoresist. In order to achieve 24μm bumps thicker resist coatings of 30μm are applied using the spin parameters detailed in table 5.2. Each achieves near vertical side walls with good resolution.

<table>
<thead>
<tr>
<th></th>
<th>AZ 9260</th>
<th>THB 151N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dispensing of the resist</td>
<td>70 rpm</td>
<td>70 rpm</td>
</tr>
<tr>
<td></td>
<td>30 Rev/s</td>
<td>30 Rev/s</td>
</tr>
<tr>
<td></td>
<td>20 sec</td>
<td>20 sec</td>
</tr>
<tr>
<td>Spread cycle for uniform wafer coverage</td>
<td>400 rpm</td>
<td>1000 rpm</td>
</tr>
<tr>
<td></td>
<td>100 Rev/s</td>
<td>200 Rev/s</td>
</tr>
<tr>
<td></td>
<td>20 sec</td>
<td>20 sec</td>
</tr>
<tr>
<td>Spin off cycle for final photoresist thickness</td>
<td>1000 rpm</td>
<td>2000 rpm</td>
</tr>
<tr>
<td></td>
<td>200 Rev/s</td>
<td>200 Rev/s</td>
</tr>
<tr>
<td></td>
<td>10 sec</td>
<td>30 sec</td>
</tr>
</tbody>
</table>

Table 5.2: 30μm spin coat parameters for AZ9260 and THB 151N

![Temperature vs. Time](image1.png)

**Figure 5.12 – Baking cycle for 30μm thick resist (a) AZ 9260 and (b) THB 151N**
Once again the resist is cured by placing the wafer on a hotplate. The bake times, as recommended by the manufacturers, increase for the thicker resist as shown in figure 5.12.

5.4.4 Exposure and development of AZ and THB photoresist

In this process step, positive and negative photoresists have been considered. UV exposure of the negative resist initiates the cross linking of the polymers and the resist becomes insoluble in organic or water based developers. This provides the advantage of a more mechanically stable mould, which is highly resistant to acid and alkaline aqueous solutions, making it more robust during the electroplating process. However this same reaction makes stripping of the negative resist extremely difficult, requiring special chemicals and processes, whereas positive resists can be easily removed using acetone. Furthermore, in order to achieve the required bump height, it is shown in section 5.4.6 that, when using positive resist, a single photoresist deposition step (for creating the structures for electroplating) is sufficient, but with negative resist an additional stripping and reapplication procedure is required adding considerable time and effort to the process.

The resist is exposed to UV light through a photo mask using contact lithography on a Tamarack mask aligner as shown in figure 5.13. Table 5.3 presents the optimised
energy dosage and development times for both resists. Figure 5.14 shows how the negative resist is developed compared with the positive resist. In contrast to the positive case, where by the photoreaction creates carboxylic acids weakening the bonds, the photo reaction promotes cross linking of the polymers strengthening the bonds. Thus it is the unexposed areas rather than the exposed areas that are removed in the developer. Hence, a negative image of the photo mask is required to define features on the wafer.

![Illustration of UV exposure and development steps for (a) positive resist and (b) negative resist](image)

**Figure 5.14 – Illustration of UV exposure and development steps for (a) positive resist and (b) negative resist**

<table>
<thead>
<tr>
<th>Photoresist</th>
<th>Energy dosage (mJ/cm²)</th>
<th>Development time (mins)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AZ9260</td>
<td>1800</td>
<td>5</td>
</tr>
<tr>
<td>THB 151N</td>
<td>650</td>
<td>30</td>
</tr>
</tbody>
</table>

**Table 5.3: 30μm resist exposure dosage and development times**

### 5.4.5 Etching of the upper seed layer metal

Typically two selective etch steps are required. The first occurs between the two resist deposition stages described above. Selective etching of the Au metallic layer is achieved using a gold etch solution that contains iodine and potassium iodide supplied by Rockwood Electronic Materials. This step forms the tracks and bond pads on the substrate while maintaining a conductive seed layer via the under lying Ti layer, as illustrated in figure 5.15.
The resist is then stripped and a second photo resist layer is deposited on the wafer, to create the mould for electroplating the bumps. Stripping of the AZ resists is performed with acetone and takes only a few moments to be completely removed. On the other hand, stripping the negative THB151N can take over an hour at elevated temperatures (60°C-70°C) depending on thickness and bake times applied. The stripping agent is ECK108 also attacks other materials when applied for prolonged periods.

The second etch step is the final process before dicing and takes place after the electroplating has been performed. The exposed conductive Ti seed layer is etched using an ammonia (NH$_3$) hydrogen peroxide (H$_2$O$_2$) mixture, typically 1 part NH$_3$ to 10 parts H$_2$O$_2$, with the Au layer serving as a mask.

**5.4.6 Electroplating**

A custom plating bath, shown in figure 5.16, has been used for the gold electrodeposition. The bath comprises of a two litre Perspex beaker (bath), a platinised titanium electrode (anode), a hot plate with a 60mm long magnetic stirrer and closed loop temperature feedback control. Additional features include a DC power supply and a polytetrafluoroethene (PTFE) 3 inch wafer holder with cathodic connectors. The gold plating solution is ECF 60 purchased from Metalor Technologies.
Using equation (5.4) and information from table 5.4 the bath was set up to achieve a target thickness of 20µm. However an abnormal growth protrusion is observed on the bump as shown in figure 5.17. The degree of protrusion varies across the wafer and extended up to 15µm above the bump surface plateau, which was in the range of the 20µm target height. Plating non-uniformity of gold beyond 4µm-5µm in height is a known problem at CST, who have observed similar effects. However their application requirements have never exceeded these levels.

Using the same resist structure and suitably adjusting the parameters to match the bath and material conditions, the experiment was repeated for plating Ni and Cu features. The Ni and Cu features can be used as core structure of the bumps onto which a thinner layer of gold could be electrodeposited. Also, Cu improves the conductivity of the bumps and by using less gold costs can be reduced. The results obtained were in good agreement with the projected thickness, namely, 18µm ±2µm and 22µm ±3µm for the Ni and Cu plating, respectively. More importantly no protrusion growths are observed. Therefore the problems observed in Au plating are considered to originate more from

<table>
<thead>
<tr>
<th>Description</th>
<th>Unit</th>
<th>Optimum</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gold Concentration</td>
<td>g/l</td>
<td>10</td>
<td>5-12</td>
</tr>
<tr>
<td>Potassium Sulphate</td>
<td>g/l</td>
<td>35</td>
<td>25-50</td>
</tr>
<tr>
<td>Temperature</td>
<td>°C</td>
<td>55</td>
<td>40-60</td>
</tr>
<tr>
<td>pH</td>
<td></td>
<td>9.5</td>
<td>8.5-10.5</td>
</tr>
<tr>
<td>Current Density (Vat)</td>
<td>A/m²</td>
<td>30</td>
<td>10-50</td>
</tr>
<tr>
<td>Current Density (Barrel)</td>
<td>A/m²</td>
<td>2</td>
<td>1-3</td>
</tr>
<tr>
<td>Agitation</td>
<td>m/min</td>
<td>5</td>
<td>3-7</td>
</tr>
</tbody>
</table>

Table 5.4: ECF 60 gold electrolyte suggested plating ranges
equipment setup conditions rather than the actual deposition process itself. Both the Ni and Cu plating were performed in dedicated plating tanks containing over 15 litres of electrolyte. The second most notable differences are: a) the agitation method: a air jet system is employed in each case as opposed to the magnetic stirrer used for the Au plating, and b) the contacts made to the wafer.

![Figure 5.17 – Zygo measurement showing Au protrusion growth (a) a top view contour plot and (b) an isometric 3D representation](image)

As shown in figure 5.18, the original Au plating setup used only two contacts (CST Ltd have only a single contact in their system). On further study it was found that the level of protrusion growth and non-uniformity worsens as the distance from the contacts increases. The growths occur along the same mould edge closest to the bottom of the bath. These observations suggest poor current distribution to the lower part of the wafer and possible problems with agitation. It is important to maintain an even current distribution to avoid current crowding effects and the agitation is important to the mass transport of the ions but care must be taken to avoid turbulent flow which generates air pockets preventing plating and may also lead to pitting.

![Figure 5.18 – Wafer holder for use with Au electroplating](image)
These concerns were addressed by firstly increasing the number of contacts to four with 90° separation as shown in figure 5.19(a), secondly the magnetic stirrer was modified with a paddle like attachment shown in figure 5.19(b). The paddle rotates with the magnetic stirrer effectively increases the stirrer area to promote larger volume displacements for a given speed such that a more uniform flow over the wafer surface is created.

![Image](a) ![Image](b)

**Figure 5.19 – Modifications to the Au plating bath setup including (a) a new wafer holder with increased contacts to the wafer and (b) a paddle attachment to the magnetic stirrer**

The electroplating procedure for a 20µm target thickness was then repeated with these modifications in place. The results, shown in figure 5.20, display a significant improvement in the plating profiles. The spiking effect is no longer present and Au bumps of 17µm ± 2µm are achieved across the wafer. Figure 5.20(c), shows a surface profile contour plot of a magnified image. The local discrepancy over the bump top surface is ±1.5µm. Therefore it is expected that, by finding the optimum plating densities for this set-up, the surface profile can become a lot smoother and thus the thickness distribution over the wafer will also be improved.
The electroplating trials are performed using the two photo resists discussed in the previous section. While no adverse effects are observed with the negative resist moulds, the positive resist moulds were prone to cracking and splitting, as shown in figure 5.21, this effect is commonly known as crazing [5.30].

To achieve the required height differences in the bumps one of two process paths, P1 or P2, as shown in figure 5.22 can be followed. However due to the cross linking nature of the exposed areas of negative resists only the right hand path, P2, is applicable as a re-exposure of the resist for further development can not be performed.
For DC plating at a current density of 3mAcm$^{-2}$, substituting plating duration estimates of 30 minutes, 140 minutes and 160 minutes achieve 4µm, 20µm and 24µm growths, respectively. The durations of the bump plating procedures were to $t_{P1}$ (positive resist) $\approx$ 235 minutes, $t_{P2}$ (positive resist) $\approx$ 465 minutes and $t_{P2}$ (negative resist) $\approx$ 605 minutes.

Considerable time and hence cost savings can be made through the use of positive photoresist, and overcoming the crazing problem is of benefit. The crazing effect is attributed to localised tensile stressing in the resist. The crazes eventually break down to form cracks into which electrodeposition will occur. With in the electroplating setup there are two main contributors to stressing conditions; these are the deposition rate (current density) and the bath temperature, which causes swelling of the resist. In either case deviation from the recommended optimal plating conditions will result in poorer quality deposits.

To limit the time that the sample remains under stress conditions the applied current density remains as recommended, such that the deposition rate is largely unchanged,
and a reduction of the bath temperature from 55°C to 40-45°C is implemented. Following the process path outlined by P1 in figure 5.21 bumps with a height difference of 4µm are achieved. However, the deposition appeared dull and a very rough surface profile was formed on the bump, as shown by the scanning electron microscope images (SEM) in figure 5.23.

Figure 5.23 – (a) SEM images showing bumps with 4µm height difference (b) SEM image to highlight the surface roughness of the bump and (c) Zygo measurement of the bump height

An alternative measure, intended to improve the thermal, chemical and physical stability of the resist, was to increase the duration of the bake out to 6.5 minutes form 3 minutes. Thus, more solvent is evaporated in promoting greater thermal cross linking of the resist matrix. The disadvantage is that the solubility of the resist is reduced so requiring additional development time to ensure all resist is completely removed after exposure. Therefore, at the expense of around 30 minutes additional processing time with the P1 path, a more rigid photoresist is created allowing the bath temperature to be set to 48°C-50°C. Once again the desired step height of 4µm is achieved with bumps of 18µm and 22µm. The deposits now appear much brighter and the surfaces much smoother as shown in figure 5.24.
Despite this improvement the onset of crazing occurs and plating has taken place within cracks in the resist walls. Figure 5.25 is an SEM image highlighting a fin type growth up the bumps edge. In the short term these fin growths are unlikely to cause any problems in the bonding of the encoder chips. However, should the need arise to reduce the bump pitch and gap the potential for the fin growths to create a short circuit between bumps increases.

Further study is required to refine the process and optimise the relationship between bath temperatures and resist bake times to achieving defect free bump growth. Additionally the influences of current density and agitation on the induced stresses in the bump mould may also require investigation.
5.5 Summary

The rational for the development of a gold bump electroplating process for use in solid-state flip-chip bonding is discussed. A fabrication approach that adopts UV-LIGA techniques to create a resist mould into which the electroplated features are formed is presented and shown to be compatible with multiple resist. AZ9260 positive photo resist is the preferred choice for ease of use providing a much shorter process cycle.

After overcoming some processing issues, the known uniformity problems with plated features over 5µm high electroplated Au bumps have been tackled. Au bumps of greater than 15µm thick have been achieved. In addition bumps have been formed with a 4µm thickness difference. This is an important feature that compensates for the non-planar locality of the contact pads on the optical encoder chip allowing uniform bonding pressure to be applied.

However, some minor deficiencies remain and the most cost effective approach requires further research to continue the optimisation of the bath conditions so that the most effective plating results can be produced. Nevertheless the bumps that have been fabricated to date are useable for initial flip-chip bonding experimentation, the details of which are discussed in the following chapter.
5.6 References


[5.27] Microchemicals datasheet for “Baking Steps for Photoresists”,

[5.28] Microchemicals datasheet for “Rehydration of Photoresists”,

Chapter 6

Flip-Chip Bonding Results

In chapter 3 and 5 the advantages of a flip-chip bonding solution are discussed. In particular bonding of the InP optical encoder chip onto a glass carrier submount in FCoG solution is preferred. Thermocompression bonding has been chosen as the bonding technique for the bump bonding of the electroplated gold bumps described in chapter 5. This chapter presents the results of the work carried out towards a flip-chip bonding solution.

Despite some promising results, with FCoG optical encoder chips demonstrating 100% connectivity and good mechanical performance with shear test values greater than 1000gF, a functioning device operated against a piece of scale grating has not been achieved during the timeframe of this thesis. To meet this urgent requirement a wire bonded solution is implemented and wire bonded optical encoder chips tested while development of the flip-chip bonding solution was ongoing.

6.1 Preliminary bonding

The thermocompression bonding procedure involves bringing two surfaces together into intimate contact, without any intermediate layer, under applied heat and pressure to allow diffusion to occur at the interatomic scale with substantial joint deformation as illustrated in figure 6.1. Organic contaminant barrier layers, formed from oxygen or compound gases, at the surface are known to exist. Heat is therefore required along with sufficient mechanical deformation to provide the energy to overcome this barrier layer and ensure contact of fresh gold surfaces for good bonding to take place. Therefore, in such a thermocompression bonding process, the interfacial temperature determines bond quality while insufficient deformation will result in weak or no bond. Conversely, too much deformation can cause irreversible damage to the chip and/or substrate [6.1-6.3].
Thermocompression bonding is performed using a Karl Suss FC6, (figure 6.2.). Cu bumps with a 5mm Au cap were initially flip-chip bonded onto a chip pad formed with 300nm sputtered Au. However these experiments proved largely unsuccessful due to two primary causes of failures. Firstly the pad sputtered pad layer was relatively thin, hard and non-compliant with the levels of deformation required for successful flip-chip bonding. Secondly the Cu core of the bumps hardened under the applied pressure during bonding causing the bump to pierce the die bond pad. On the other hand reduced pressures failed to bond at all. A summary of the Cu/Au bump bonding process is provided in appendix C.

Although unsuccessful, these initial bond trials provided some valuable insights into the bonding of large InP chips that led to the development of a custom chip pick up tool as discussed in section 6.2. This also led to a change of material and design of the bump/bond pad. Pure electroplated gold bumps replaced the copper core and the bond pad thickness was increased by electroplating a 1μm thick gold layer on top of the gold sputtered pads.
6.2 Flip-chip pick-up tool modifications

A large proportion of the initial bonding attempts failed as a result of damage to the chip induced by the thermocompression bonding procedure. A study of the damaged chips revealed cracks and fracture lines that followed the edge of the vacuum slot on the chip pick up tool, as indicated in figure 6.3, raising concerns over the distribution of applied pressure during the bonding cycle.
The FCB pick up tool is made of silicon carbide, SiC, with a raised platform in the centre onto which the chip is placed, as shown in figure 6.4. To hold the chip in position a vacuum slot, in the form of a race track of 1mm wide by 2mm long, is located directly underneath the chip. The chip has a surface area of 2mm by 2.5mm thus 40% of the chip, and many bumps in this area, overhang the chuck. The applied load is therefore non-uniformly spread over the chip resulting in stress concentrators appearing along the edges of the vacuum slot that led to the cracking of the chip.

The current race track vacuum slot has been replaced with a new custom pick tool design which is a perforated diaphragm over a 1mm vacuum slot, as shown in figure 6.5. The capillary holes are 80µm in diameter so this design has no significant sharp edges in contact with the die, and therefore no stress concentrators.
element analysis, FEA, showed that, by forming the tool in stainless steel and slightly increasing the diaphragm thickness, a robust tool could be manufactured at low cost. Therefore a stainless steel tool was made resulting in no significant deformation being observed.

![Image](image.png)

**Figure 6.6 – Zygo measurements of pick up tool showing (a) deformation surface profile (b) top view contour plot and (c) an isometric 3D representation**

The newly developed pick up tool was used in the assessment of Au to Au flip-chip bonding processes discussed in section 6.4.

### 6.3 Finite element analysis model development

An initial analysis to develop an FEA model to aid in the optimisation of the flip-chip bonding process has been undertaken. Although the work described is at an early stage, the rationale of this activity was to develop a robust model that could be used to evaluate pre-bonding conditions for various materials and bump layouts for future optical encoder chip designs.

In its current form the model, shown in figure 6.7, consists of a glass submount (BK7 type properties) with simplified thin Au surface metallization features located about the central bonding areas. Similarly, the InP chip has thin Au features to represent the metallization of the surface of the chip. The thermal and structural properties for these materials are provided in table 6.1.
The 60μm diameter, 20μm tall Au bumps and the 1μm thick Au bond pad geometries are formed as a single object. The 4μm difference in height between some connections has so far been neglected. Therefore a true representation of the bond interface has not included and the model assumes the complete initial contact between pad and bump has already been established with 100% alignment accuracy. Furthermore the simulations performed to date assume a fully constrained surface layer at the glass to chuck interface and a uniformly distributed load applied to the chip surface.

A transient thermal analysis is performed to assess the heat affected zone associated with the bump area. The profiles of the thermal inputs applied to the glass submount chuck surface and the chip arm contact surface are shown in figure 6.8. Where the first bond trials made initial bump to pad surface contact at temperatures in excess of 150°C, this model assumes initial contact at room temperature. The reasons for the contact at elevated temperatures are twofold: it reduces the risk of thermal shock to the chip and
submount and also minimises the time for temperature ramping during the bonding process. Although not shown here, additional models suggest that this mode of operation would only provide a very slight increase in the final temperature at the upper glass surface over the given bond cycle duration with no influence on the bump temperature reached.

![Figure 6.8 – Simulated bonding temperature profiles](image)

Figure 6.8 – Simulated bonding temperature profiles

Figure 6.9 – Temperature distribution at the end of the temperature cycle shown in figure 6.8 for (a) isotropic view and (b) magnified side elevation

The results shown in figure 6.9 suggest that the bump temperature is largely influenced by the temperature applied to the InP chip. This is further emphasized in figure 6.10, where the InP chip, and hence the thermal load to the chip, have been removed leaving only the glass submount and Au bumps. In this case an applied thermal load of 300°C to the submount sees the temperature at the bumps reach a similar temperature to the top glass surface of approximately 290°C. While others such as Ang et al, report higher temperatures applied to the submount in order to allow lower temperature to be applied to the chip arm, aimed at protecting the chip from adverse thermal effects, these results indicate that such a process would not be compatible for this flip-chip on glass
application and the desired bond temperature would need to be applied through the die [6.2].

![Temperature distribution through glass submount and bumps with no chip](image)

Figure 6.10 – Temperature distribution through glass submount and bumps with no chip

During the manufacturing process the chip is annealed at 450°C and thus, to avoid causing any thermal damage, a maximum temperature of 350°C is considered safe, while ideally lower temperatures would be used if possible.

A second analysis that incorporates the transient thermal results as a thermal condition for a steady state structural model has also been performed. A pressure of 40Pa is applied uniformly to the surface of the chip after following the profile shown in figure 6.11. The deformation of the features from initial conditions as a result of the temperature change and applied force is shown in figure 6.12. With the bottom glass surface assumed to be completely constrained we first see thermal expansion, particularly in the glass, before a compression at the chip/ bump area on the application of the 40Pa force. The extent of the expansion in the glass may be exacerbated as a consequence of the constrained bottom surface. Nonetheless it gives rise to the question of when during the thermal cycle should the bond surfaces be brought into contact as the expansion could introduce misalignment of the bump and pads. One suggestion, based on this result, could be to perform alignment at the bonding temperature after which the chip and submount can be brought into contact with no further temperature ramp required. However this is likely to lead to a prolonged period at the elevated temperatures and the implications of such a step needs to be fully considered.
The behaviour of a bump has also been studied in isolation. Figure 6.13 shows an image of the bump model before and after the thermal load cycle described above. The images are displayed with an exaggeration factor applied to provide some visual clarity. At first glance the figure would suggest a bump deformation of around 3μm. However, this image includes the total system deformation from the original conditions thus the contribution from thermal expansion must be considered before the bonding load is applied. Taking this into account the model suggests an actual bump deformation of around 1μm.

While the level of deformation can be shown at peak applied loads, giving an insight into the stresses applied, the current licence agreement is restricted and prevents modelling beyond this. Modifications to the surface profiles, materials, alignment and other parameters can all be made to achieve a more realistic interaction and bond interface. Further analysis on the effects of the induced stresses, such as plastic
deformation or fracture, can not be fully analysed until the license is upgraded to accommodate these features.

Figure 6.13 – Images of the modelled bump deformation (a) before and (b) after the applied thermal and force cycles

6.4 Au to Au thermocompression bonding

6.4.1 Bump shear strength

The gold to gold flip-chip bonding experiments were performed using bumps formed by the electroplating procedures discussed in Chapter 5. Both the rough and smooth surface finish bumps have been studied. However before commencing with the bonding procedures the individual bumps formed on the glass were shear tested following the EIA/JEDEC standard 22-B116 [6.4].

Figure 6.14 – Individual bump on glass shear measurements for short (20µm) and tall (24µm) bumps.
The results in figure 6.14 show the average shear strengths for the 20 and 24\(\mu\)m thick bumps as 83gF and 127gF, respectively. While demonstrating higher shear strength the taller bumps also show a wider variation in shear strength over individual sample measurements. The trend lines that represent the JEDEC standard recommended minimum individual and sample average shear values are also plotted for a bump diameter of 60\(\mu\)m. As can be seen the electroplated gold bumps achieve far greater values suggesting good mechanical strength. However it is worth bearing in mind that this standard is based on gold wire stud bumps on aluminium alloy bonding surfaces and to the best of the author’s knowledge a dedicated electroplated bump equivalent standard has yet to be published.

6.4.2 Flip-chip bonding

Figure 6.15 provides a generic example of a bonding profile and the input parameters for the Karl Suss FC6 flip-chip bonder. A vast number of possible parameter combinations and bond profiles are available that will influence the performance and quality of the bond. Unfortunately the bonding experimentation is restricted by the limited number of available encoder chips and therefore a dedicated design of experiments was not possible.

Figure 6.15 – Generic example of the temperature and force parameter controls as a function of time, as input into the Karl Suss FC6 flip-chip bonder
The results discussed in this section concern the bonding of twelve optical encoder chips. Six chips in batch 1 have a rough bump surface finish as in figure 5.21 while the second six samples in batch 2 have a smoother bump surface finish similar to those of figure 5.22. The complete bonding profile parameters for each bonded chip can be viewed in Appendix D while tables 6.2 and 6.3 provide a summary of applied bond loads and temperatures as well as connectivity and shear test results.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Applied bonding force (g/bump)</th>
<th>Initial surface contact temperature (°C)</th>
<th>Bond temperature (°C)</th>
<th>Total cycle time (s)</th>
<th>Max load time (s)</th>
<th>Connectivity percentage (%)</th>
<th>Shear strength (g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>40</td>
<td>300</td>
<td>300</td>
<td>145</td>
<td>20</td>
<td>42%</td>
<td>757.1</td>
</tr>
<tr>
<td>2</td>
<td>40</td>
<td>300</td>
<td>300</td>
<td>71</td>
<td>10</td>
<td>34%</td>
<td>443.6</td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td>300</td>
<td>300</td>
<td>71</td>
<td>10</td>
<td>23%</td>
<td>273</td>
</tr>
<tr>
<td>4</td>
<td>50</td>
<td>300</td>
<td>300</td>
<td>145</td>
<td>20</td>
<td>43%</td>
<td>646</td>
</tr>
<tr>
<td>5</td>
<td>35</td>
<td>100</td>
<td>300</td>
<td>82</td>
<td>30</td>
<td>52%</td>
<td>764.9</td>
</tr>
<tr>
<td>6</td>
<td>35</td>
<td>100</td>
<td>300</td>
<td>250</td>
<td>30</td>
<td>25%</td>
<td>594.5</td>
</tr>
</tbody>
</table>

Table 6.2: Bond parameters and results summary

<table>
<thead>
<tr>
<th>Sample</th>
<th>Applied bonding force (g/bump)</th>
<th>Initial surface contact temperature (°C)</th>
<th>Bond temperature (°C)</th>
<th>Total cycle time (s)</th>
<th>Max load time (s)</th>
<th>Connectivity percentage (%)</th>
<th>Shear strength (g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>35</td>
<td>220</td>
<td>325</td>
<td>67</td>
<td>30</td>
<td>60%</td>
<td>932</td>
</tr>
<tr>
<td>2</td>
<td>35</td>
<td>220</td>
<td>300</td>
<td>57</td>
<td>10</td>
<td>50%</td>
<td>276.4</td>
</tr>
<tr>
<td>3</td>
<td>35</td>
<td>220</td>
<td>300</td>
<td>67</td>
<td>30</td>
<td>90%</td>
<td>957</td>
</tr>
<tr>
<td>4</td>
<td>35</td>
<td>220</td>
<td>300</td>
<td>67</td>
<td>30</td>
<td>100%</td>
<td>1127.1</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>220</td>
<td>300</td>
<td>67</td>
<td>30</td>
<td>90%</td>
<td>801.2</td>
</tr>
<tr>
<td>6</td>
<td>45</td>
<td>220</td>
<td>300</td>
<td>67</td>
<td>30</td>
<td>100%</td>
<td>1219</td>
</tr>
</tbody>
</table>

Table 6.3: Bond parameters and results summary

The chip shear strength increases with applied bonding load. However, as further emphasised in figure 6.16, the duration of the applied load may also influence the joint strength. Figure 6.16 shows that larger bonding strengths are achieved at loads between 35 and 45g/bump when the load is applied for a period of at least 30 seconds. Additionally, when bonding from hot i.e. no temperature ramp, higher bond strength was achieved than when starting the cycle from cold (50°C-100°C). However, in this instance, one must consider the additional load ramping profiles employed within the thermal cycle as more time is required to allow heating of the samples when starting from cold. However, in batch 2 the initial bond contact was made at 220°C for every sample with no apparent adverse effect on the achievable bond strength.
The level of connectivity between bond pads for the batch 1 flip chip bonded devices was very poor, achieving 50% at best. Moreover, the distribution of good connections across all chips bears no resemblance to the applied bonding forces. However, it is noted that for chips bonded with the same bonding force (samples 1-2 and 3-4) the chips with higher shear strength also demonstrated a higher level of connectivity suggesting, as would be expected, that the more good bonds that are made the higher the bond strength will be.

Figure 6.17 provides two images of the optical encoder chip after shear testing. In the first example, figure 6.17(a), a clear indication of bump to pad connection is observed with strong indentation marks on the lower half of the chip. Little or no evidence that the bumps and pads came into contact is present on the upper section. Similarly, in the second example, evidence that contact has been made is visible on most pads, perhaps less towards the left upper pads. These observations are indicative of a non-planar bond interface when the bumps and pads are brought in to contact.
For the batch 2 chips poor connectivity was once again observed. Subsequent inspection of the bonding equipment led to the discovery of further faults with the new pick-up tool. The front surface of the pick up tool was found to be non-planar with respect to the back surface. Consequently a non-planar distribution of the load occurred and reflected the angle of non-planarity between front and back surfaces of the work piece. After corrections were made the connectivity yield increases from less than 60% to 90%-100%.
Inspection after shear provides insight into the joint quality achieved for different bonding forces. Figure 6.18 shows images of the bond pads after shear test of a chip bonded before modification of the pick-up tool and then for chips, bonded after the pick-up tool modifications were made, with applied forces of 25g/bump, 35g/bump and 45g/bump. The level of interconnection at the bump to pad interface has increased accordingly until the shearing of the bond has been observed at the submount to bump interface where the local shear strength has been shown to be of the order of 100g/bump.

The doughnut shape showing an unbonded area in the centre as shown in figure 6.18(a) and 6.18(b) is also reported in the work [6.1, 6.2]. This shape stems from the lack of bump deformation as the bonding is initiated at the periphery and extends radially towards the centre as the interfacial barrier is overcome. Furthermore, additional inspection of the of pre-bumping bump profiles from batch 2 samples reveals a basin like topography, as shown by the Zygo white light interferometry measurements (figure 6.19).

![Figure 6.19 – Zygo measurements showing (a) 3D interpretation and (b)surface profile plots](image)

The trough in the centre of the bump surface is approximately 1.5μm below the peripheral edge. Therefore it is this edge of the bump that makes initial contact with the pads and sufficient deformation of the structure is required before the centre of the bump come into contact. This combination of factors must be overcome if bonding strengths are to be improved.
The bump shape could be as a result of the current crowding effects as illustrated in figure 6.20. One approach to resolve this would be to employ a continuous pulse electroplating process rather than DC plating. The pulse plating process has been shown to perform an in-situ electromechanical polish of the bump to produce higher quality finish flatter bumps [6.6-6.8]. A second consideration would be to improve the agitation method to improve the flow of the plating surface. This could potentially be achieved by replacing the stirring mechanism with a megasonic agitation induced by an ultrasonic transducer [6.9-6.10].

Despite these apparent shortcomings the signs for a successful development of a flip-chip bond solution are encouraging. InP encoder chips have been bonded to glass submounts and displayed 100% connectivity, as shown in figure 6.21, so demonstrator modules should be available in the imminent future.

The strongest bonded die exhibits die shear strength of 1220gF which corresponds to an average bump strength of 29g/bump, some way below that of the individual bump strength measurements discussed in section 6.4.1. In terms of standard specifications the MIL standard 883E for die shear strength [6.11] suggests minimum die shear strength of 2500gF for a chip of similar dimensions as the encoder. However this standard implies the attachment for the entire chip surface area and it is unclear how this translates to a bumped die with no underfill. If only the actual bonded surface area was to be considered then the equivalent minimum shear strength required for the optical encoder chip would only be 100gF, which would compare favourably with the 1220gF achieved. However, one must now consider the potential requirement of underfill to
improve the mechanical stability of the joints and meet the standard for the entire chip surface area. Additionally this underfill must not impose any adverse effects on the optical performance of the device. The bump cross section shown in figure 6.21(b) shows a good bond with a bump height and thus gap of 20μm being maintained so application of underfill is applicable.

6.21 – Image showing (a) a flip-chip bonded encoder chip on glass submount and (b) bump cross section

6.5 Summary

The findings of a preliminary gold to gold solid state thermocompression flip-chip bonding investigation have been presented. InP is a more brittle material than Si or GaAs particularly for the relatively large chip dimensions considered. The compatibility of InP with the bonding forces used during thermocompression bonding was therefore carefully considered. Among other challenges and obstacles the development of a pick-up tool design coupled with gentle load ramps and extended
bond times have yielded fully bonded optical encoder chips, which demonstrate 100% connectivity and have reached shear strengths of 1220gF. However limited availability of encoder chips has placed heavy restrictions on the investigation and prevented the production of a flip-chip bonded module for full functional experimentation against a scale grating.

The lack of more detailed results limits the available analysis and results and observations obtained thus far must be substantiated with further experimentation. Nonetheless, given the resources and time frame available for this section of the project, the progress made is extremely encouraging and the expectations that a successful FCB module be produced are high. Furthermore the early stages of a FEA bonding model have been implemented and discussed. Despite the infancy of this design early results have indicated that in order to maintain the temperature of the bump at the desired bonding temperature it may be necessary to provide heat to the system through the chip, rather than through the submount as in traditional thermocompression assembly. This result requires verification through practical experimentation. Much development work is still required in this field and it is hoped that the model will not only aid in the optimisation of the bond parameters (temperature and force) but also in the optimisation of the bump dimensions and location to help elevate stressing in the chip. This may prove particularly useful in future chip design iterations where the bond may not necessarily be required at the periphery of the chip.

6.6 References


Chapter 7

Wire-bonded surface emitting LED optical encoder chip

The InP chips are relatively large (compared to traditional InP components, such as LEDs) and the material brittle in comparison to Si or GaAs. Thus significant development work was required to determine a suitable thermocompression flip-chip bonding solution. In the meantime verification of the operation of the optical encoder chip as an integrated device presented itself as a matter of the utmost urgency. To satisfy this need a wire bonded solution was implemented, allowing the LED, photo detectors and the performance of the incremental function of the encoder to be assessed. Subsequently, reference marker schemes are introduced onto the chip design, discussed in chapter 8. In this chapter are presented the results of tests and characterisation performed on the wire bonded optical encoder chips.

![Image](a) ![Image](b)

Figure 7.1 – Images showing a surface emitting optical encoder chip wire bonded to (a) glass carrier having two layer metallisation and (b) to a Si carrier having only a single layer of metal

In the first instance optical encoder chips were wire bonded on a glass carrier as shown in figure 7.1. The glass carrier is fabricated with future Flip-Chip on Glass (FCoG) packaging technology options in mind. A dual-level metallisation process is used to create tracks that were large enough to accommodate a wire bond. The configuration of the tracks is such that photo diodes signals of appropriate phase shift are connected in common on a track via the wire bonds. The tracks connect to larger pads used for probing or soldering to a secondary test board. Such an approach maximises the use of the glass real estate while either removing the need for multilevel (up to 4 levels or more) interconnects in future LTCC or PCB packaging designs or reducing the real
estate used in less intricate designs. In future chip designs multi level metallisation can be applied to the chip design. This is avoided in these early prototype designs as it removes any risk associated with the process, such as on chip short circuits, and allows for individual probing of each of the photo diode and LED interconnections to assess chip prior to bonding.

The double metal layer fabrication process carried out in collaboration with CST Ltd, is shown in figure 7.3. First attempts used a 500nm layer of Si₃N₄ as the dielectric...
isolation layer. However, delamination and warping problems, such as shown in figure 7.2, resulted in a change to SiO$_2$. This change coupled with an increase in thickness and a move from a dry etch to a dry/wet etch combination largely resolved these issues, although subsequent testing of bonded chips were prone to short circuits. Vias were opened in the dielectric layer, using standard lithographic techniques, where connections between lower and upper level metals are required. The second level metallisation is sputtered, which adequately fills the vias to provide electrical connection to the lower level metal. Resist is then applied and patterned for the selective etching of the upper metal layer. However this step differs depending on the assembly approach. For wire bonding the upper level metal is etched down to the dielectric layer, the resist is stripped and the wafer is ready for dicing. For flip-chip bonding a seed layer is required to electroplate the bumps, therefore no etching is necessary at this stage, although etching of the Au layer maybe performed.

As a consequence of the multi-level metal delamination and shorting problems a silicon carrier, having only one layer of metallisation, was also investigated. Failures in this case can be attributed to problems with the chip, chip fabrication or wire bond assembly. The results obtained with both glass and Si carriers are now discussed.

7.1 Chip on glass carrier

7.1.1 LED response

Four LED configurations are considered: they are either lozenge (loz) or circular (circ) mesa geometries with bridged or non-bridged connections as shown in figure 7.4. The outputs of each configuration are measured to determine which one provides the largest optical power. As seen from figure 7.5, the larger lozenge shaped LED gives a larger output while the continuous contact is slightly more efficient than a bridged wire bond contact.
The absorbing region located directly underneath the LED is connected to act as a monitor of the LED output, as discussed in chapter 4. Figure 7.5(b) suggests that output powers of 50mW should be achievable. However this was not found to be the case and outputs were considerably lower. Further investigation reveals some transistor like characteristics between the LED and the monitor diodes. Referring back to chapter 4 and the cross-section schematic of the chip, figure 4.2, a PNP transistor was inadvertently formed in the monitor/ LED stack, still capable of drawing a significant proportion of the current.
Figure 7.5 – Plots for (a) L-I curves for lozenge and circle LEDs and (b) I-V curves of LED output measured by monitor diode where the circled area suggests large photo currents of 50mA implying a 50mW LED output

A schematic representation of the ideal circuit versus the measured circuit is presented in figure 7.6. If this non-ideal transistor was to behave as a switch and the poor isolation between the emitting and photo detector circuit was acting as a load on the monitor node then the load is switched on when the common node is low. As this node is connected to zero volt, it is assumed that the load is pulling current through the monitor diode. For 100mA input a near 60:40 split is observed at the transistor
junction, which corresponds with the measurements presented in figure 7.5. The detector load circuit is also connected to zero volts so a further assumption is that some of this current forward biases the photo detectors, thus swamping the signals. Unfortunately on-chip measurements cannot be made to verify if current is entering the detectors. Nonetheless a substantial increase of noise is to be expected.

Figure 7.6 – Schematic representation of (a) the ideal optical encoder chip and (b) highlighting observed transistor like effects and poor isolation between the LED and detector sections

In such a case the photodiodes will not be allowed to operate in true photovoltaic mode with zero biasing conditions. Therefore to combat this problem a resistor is placed in series with each of the semiconductor diodes, which are reverse biased, so that only a very small current is present. As a result the photo diodes are more sensitive to light than in the photovoltaic mode but these circuits are also more susceptible to electrical noise. The voltage drop can then be measured when the change in electrical conductivity varies the current flowing in the photo conductive circuit. Furthermore the poor isolation is bypassed by making a physical connection from the LED/monitor n-type node to the photo detector common p-type node, as shown in figure 7.7.
7.1.2 Photo detector response

Before a full functional test is carried out with the optical encoder chip set-up against a scale grating, the response of the photo detectors was investigated. With the photo detectors illuminated by an incandescent lamp at 150mm, the output from the photo detectors is measured directly with no buffer or amplification circuitry. Figure 7.8 shows a typical response from a single photo detector channel; outputs of over 200mV are observed.

![Graph showing the response of a single photo detector channel to an incandescent lamp. The x-axis represents time (s) and the y-axis represents photodiode output (mV). The graph shows the output of a channel when the lamp is off, on, and then off again.]

Figure 7.8 – Single channel photo detector response when illuminated with an incandescent light source at 150mm

The effect on the output signal as a result of stray light from the LED is investigated by switching on the LED and observing the response. With the LED drive conditions of 5V and 200mA the response from a single detector channel is shown in figure 7.9.
Once again no buffer or amplification circuits are present. There is an initial jump in the output when the LED is first switched on followed by a lengthy settling time.

Figure 7.9 – Measured response from a single photo detector channel after the integrated LED is switched on with 5V, 200mA drive conditions

Combining the two experiments the detector diodes were periodically illuminated using the incandescent lamp while the LED is repeatedly switched on then off. The results are presented in figure 7.10. As the LED is switched on, the response from the photo detector decays and then recovers when the LED is turned off. This result suggests a
possible thermal instability, but electrical effects such as charging or leakage through the poorly isolated submount to the monitor could also be occurring.

7.1.3 Thermal probe

The surface temperature of the chip and various locations on the glass carrier tile are recorded against increasing LED drive currents as shown in figure 7.11. The increase in LED drive current is followed by an increase in temperature reaching a high of 111°C at location 1, directly on the optical chip surface. It is not possible to probe directly on top of the LED without damaging either the LED structure or the index grating or both, but it is reasonable to assume that this location would generate the most heat. Such thermal changes can have a negative impact on the performance of the optical encoder and highlights the need for adequate heat sinking.

An additional benefit is therefore found in the use of silicon as the chip carrier as the thermal conductivity, 149 W.m⁻¹.K⁻¹, is substantially greater than that of glass, 1.1 W.m⁻¹.K⁻¹. The thermal performance when using silicon is discussed in section 7.2.1.
7.1.4 Functional test

7.1.4.1 Experimental set-up
The wire bonded chip on glass carrier has been mounted onto a PCB with connectors to a bank of $1\, \text{M}\Omega$ resistors, such that the detectors are operating in photo conductive mode as described above in figure 7.7. For the purpose of these experiments discrete bulk resistors were used. In the future, the pre-amplifier circuitry will be located on the PCB or ASIC and, with future advances in integration, some electronic functionalities may also be integrated on to the chip. The four incremental output signals are then passed to a simple differential amplification circuit to produce the phase and quadrature outputs with gain of 33, which are viewed on an oscilloscope.
The PCB that accommodates the optical encoder and electronic circuitry is mounted in a mechanical test rig that consists of linear and rotary translation stages, as shown in figure 7.12. Such a rig provides six degrees of freedom for simple fine alignment and a means to translate the optical encoder chip relative to the scale. A Ronchette scale grating is located above the PCB supported on four pillars.

Alignment is achieved with surprising ease. A very short set-up time of the order of a minute is all that is needed even for a manual set-up such as this. This highlights a particular advantage of the optical chip encoder as time required for the set up and alignment is considerably shorter than required for current readheads. Set-up times for the latter can be as long as 30 minutes.

7.1.4.2 Results
As discussed in chapter 3, the two oscillations can be used to produce a Lissajous curve. A ratio of phase and quadrature equal to 1 produces an ellipse and when the phase difference is 90° a circle is produced. Lissajous curves therefore enable an assessment of the signal outputs and alignment tolerances.

The first observed Lissajous, figure 7.13, had a maximum peak-to-peak voltage of 700mV at a nominal stand-off distance of approximately 1.25mm. This Lissajous curve is not a perfect circle as a result of a small imbalance between the phase and quadrature
outputs and a slight phase angle error, i.e. the ratio of phase to quadrature is not quite equal to 1 and the phase angle shift is not quite 90°.

![Figure 7.13 – Maximum voltage peak-to-peak Lissajous curve observed for circuit connected as described in figure 7.7 with 1MΩ resistor](image)

Further investigation found that shorting out the monitor diode under the LED, as indicated in figure 7.14, improved the performance with the maximum AC signal increasing to 1.5V peak-to-peak, figure 7.15. In this state the worst case noise observed was approximately 75mV giving a signal to noise ratio, SNR, of 20:1. The SNR is an important characteristic as it determines the level of interpolation that can be applied. A SNR of 50:1 would indicate a straightforward 200x interpolation achieving 20nm resolution.

![Figure 7.14 – Representation of the optical encoder chip connected in photoconductive mode with a reverse bias voltage of 2V applied to the photo-detectors, 1MΩ resistors and shorting of the monitor diode](image)
Finally the mechanical tolerances are recorded by measuring the full width half maximum, FWHM, values of the Lissajous curve. This measure is obtained by finding the maximum Lissajous curve and adjusting in turn each of the alignment mechanisms to rotate around the three axis, as shown in figure 7.16, until the observed Lissajous curve reaches half the maximum starting value. This measurement is also made for the stand-off distance. The FWHM values are used to determine the readhead installation tolerances and calibration at the system level.

![Figure 7.16 – rotational axis for mechanical alignment showing (a) yaw, (b) roll and (c) pitch](image)

The current tolerances for encoders having a parallel line detector arrangement and operating with a 20µm scale are known. These values can be scaled directly with grating pitch so that a comparison of current encoder tolerances can be made against those measured for the 4µm pitch optical encoder chips on glass submounts which were investigated. The results are presented in table 7.1. The yaw tolerances of the new design are comparable but a significant improvement has been achieved in roll, pitch...
and stand-off. As there is no protective glass window the negative stand-off tolerance has not been measured for fear of damaging the wire bonds should they come into contact with the scale.

<table>
<thead>
<tr>
<th>Alignment</th>
<th>Current 20µm tolerances</th>
<th>Current tolerances scaled to 4µm</th>
<th>Chip on glass</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yaw</td>
<td>±0.4°</td>
<td>±0.08°</td>
<td>±0.083°</td>
</tr>
<tr>
<td>Roll</td>
<td>±0.5°</td>
<td>±0.1°</td>
<td>±1.75°</td>
</tr>
<tr>
<td>Pitch</td>
<td>±0.5°</td>
<td>±0.1°</td>
<td>±0.5°</td>
</tr>
<tr>
<td>Stand-off</td>
<td>±0.08mm</td>
<td>±0.02mm</td>
<td>±0.65mm</td>
</tr>
</tbody>
</table>

Table 7.1: Full width half maximum tolerances

7.1.5 Summary

From the measurements and observations discussed it has been demonstrated that an optical encoder chip is feasible with a monolithically integrated light source, photo detectors and optical gratings. The monitor diode failed to operate as desired and after shorting it out a Lissajous curve with peak-to-peak voltage of 1.5V and SNR of 20:1 was demonstrated. For real applications a minimum of 1Vp-p and SNR of 50:1 would be required so some work necessary in order to reduce the noise. Furthermore the response of the photo detectors appears to be adversely affected by the poor heat sinking properties provided by the glass submount. The next section presents the results with a silicon submount which has a much larger thermal conductivity to improve the thermal stability of the chip and has only a single level metallisation to improve electrical integrity.

7.2 Silicon submount

The previous sections demonstrated the concept of integrating the components of an optical encoder onto a single chip. However, problems occurring with manufacturing the multi-level metal conductor layers on the glass which severely limited the number of sample chips available for test.

For this submount, the wire bonding of the chips was outsourced to the specialist packaging company OptoCap, of Livingston, which used a Polamar automated wire bonder, increasing the throughput while removing the risk of manual error. Coupled with the removal of the problematic multi-level process the yield for testable die
increased from around 10% to 75% at the cost of a 30% increase in carrier footprint. The majority of the remaining failures were found to be from short circuits on the chip, which can be corrected for with slight modifications to the chip manufacturing process at CST Ltd.

7.2.1 Thermal probe

Before testing the encoder with a scale grating, temperature and diode response measurements are carried out. Firstly the chip is probed using a thermocouple at the locations on the chip directly next to the LED, equivalent to the location of the chip on glass presented in figure 7.11a. The recorded temperature change with increasing LED drive current at each location is presented in figure 7.17.

The temperature reaches 45°C at 300mA drive current an increase of 22°C as compared with the 111°C increase observed on the glass submount. Therefore the heat spreading properties of the submount or package to which the die is to attached (wire bonded case) must be considered. Moreover for a flip-chip on glass assembly a specific thermal dissipation designs may be required.

![Figure 7.17 – Comparison of chip surface temperature for increasing LED drive current](image)

Figure 7.17 – Comparison of chip surface temperature for increasing LED drive current
7.2.2 Response

In section 7.1.2 the response of the photo detectors was influenced by the change in temperature when the LED is switched on. With the encoder set-up in photovoltaic mode the same procedure as before is repeated, that is, periodically illuminating the photo-detectors with an incandescent lamp while the integrated LED is switched on, with a drive current of 200mA, then off. Once again no buffer or amplification circuitry is connected. The output from a single detector channel is shown in figure 7.18. As with the chip on glass submount when the LED is off a signal change in excess of 200mV is observed.

![Figure 7.18 – Response from a single photo detector channel periodically illuminated by an incandescent lamp while the integrated LED is switched on then off](image)

While the lamp is off the LED is switched on and a shift in the base level signal is observed. With the influence of stray light and other electrical effects such a small shift can be expected. As shown in figure 7.18, the lengthy decay and recovery times observed with the chip on glass submount are no longer present. There is however a 20% reduction in the magnitude of signal changes between the LED on and off states. In comparison the reduction with the chip on glass submount is initially around 25% dropping to greater than 60% as the response decays. The above two experiments highlight the importance of heat sinking the optical chip to achieve greater stability and improve performance. Heat dissipation techniques will need to be considered for the final package and may even require the use of a silicon submount.
7.2.3 Functional test

Once again the chip on silicon carrier is mounted to a PCB. When testing the encoder against the scale grating the pre-amplification voltage conversion circuit is connected as shown in figure 7.7. Alignment is once again achieved with minimal time and effort. Where previously the maximum Lissajous curve was 1.5V peak-to-peak, here peak-to-peak voltages greater than 2.5V are achieved. However these signals are bandwidth limited and as the LED drive current increases so do the peak-to-peak voltages and even more so the noise until eventually saturation occurs, as shown in figure 7.19.

![Figure 7.19 - Encoder chip on silicon showing an example of a maximum peak-to-peak voltage Lissajous curve with bandwidth limited signals (a) before saturation and (b) at the onset of saturation](image)

To stabilise the output a dedicated electronics circuit was implemented. The bandwidth is increased to 2MHz, which is many orders of magnitude greater than previous designs, at the expense of reducing the maximum achievable peak-to-peak voltage outputs by reducing the resistor values in the pre-amp circuit down to 100kΩ. The biasing on the photo detectors is also connected to be more representative of the perceived final electronics by placing a reverse bias on the photo diode anode, as illustrated in figure 7.20.
Figure 7.20 – Representation of the optical encoder chip connected in photoconductive mode with a reverse bias voltage of -2V applied to the photo-detectors, 100kΩ resistors and shorting of the monitor diode

In this configuration ratios of the amplified output signals quadrature:phase approximately equal 1 and phase angle shifts of 90° are achieved, producing circular Lissajous curves. As shown in figure 7.21, peak-to-peak signals of 1.5V and signal-to-noise ratios of greater than 40:1 have been observed.

Figure 7.21 – 1.5Vpp signal Lissajous curve observed when the encoder chip is connected as shown in figure 7.20

Manipulation of the translation stages allows measurement of the FWHM mechanical tolerance values. With a nominal stand-off of 1.5mm the results are displayed in table 7.2 alongside the scaled down 20μm encoder values and the chip on glass measurements found previously.
<table>
<thead>
<tr>
<th>Alignment</th>
<th>20µm tolerances scaled to 4µm</th>
<th>Chip on glass</th>
<th>Chip on Silicon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yaw</td>
<td>±0.08°</td>
<td>±0.08°</td>
<td>±0.08°</td>
</tr>
<tr>
<td>Roll</td>
<td>±0.1°</td>
<td>±1.75°</td>
<td>±0.9°</td>
</tr>
<tr>
<td>Pitch</td>
<td>±0.1°</td>
<td>±0.5°</td>
<td>±0.8°</td>
</tr>
<tr>
<td>Stand-off</td>
<td>±0.02mm</td>
<td>+0.65mm</td>
<td>+0.92mm</td>
</tr>
</tbody>
</table>

Table 7.2: Full width half maximum mechanical tolerances

The four chip on Si devices tested were found to be consistent to within 5% of each other, for all tolerance measurements. The yaw value is again found to be comparable with the current encoder tolerance with an improvement in roll, pitch and stand-off. However there is discrepancy between the chip on glass and chip on Si results. Given that these are all chevron design the tolerances would be expected to be similar regardless of the submount selected. It is possible that an error has been made when measuring the chip on glass values but unfortunately this can not be confirmed as the remaining chip on glass devices were damaged during retesting.

The signal to noise ratios for the four LED configurations are mapped against the LED drive current and photo detector bias voltages to investigate potential operating conditions, the results are shown in figure 7.22.

The maximum SNR ranges vary for each device from 25-30 to 45-50. Possible reasons for this could include factors such as slight variations in index/analyser grating marked space ratios, variation in the output intensity from the LED, brightness and temperature, pick-up from additional equipment running in the lab or circuit design. The also oscilloscope contributes to the noise level worsening as the measuring scale is increased.

Achieving larger SNR allows for higher levels of interpolation to be applied and thus maximum resolution systems. Therefore operating conditions could be based on achieving the largest possible SNR. Averaging over the four chevron type encoder chips investigated here would suggest a drive current of 185mA with a -1V bias applied to the photo detectors, which is considerably higher than the 20-50mA required for current encoders. However due to the unique aspects of this development no dedicated operating specification has yet been set.
Figure 7.22 – Signal-to-Noise Ratio contour plots for (a) lozenge shaped LED with non-bridged contacts, (b) lozenge shaped LED with bridged contacts, (c) circular LED with non-bridged contacts and (d) circular LED with bridged contacts

The chip on Si carrier was also tested with dedicated pre-amp circuitry lifted from current optical encoder technology. However large AC noise values and low signal outputs were observed. Investigation suggests that noise in the pre-amp circuit can be minimised by compensating for the junction capacitance. Therefore the junction capacitance is calculated by measuring the terminal capacitance at the individual photo diodes and summing over the area contributing to a detector channel. With an assumed operating bias voltage of -1V, the junction capacitance, per channel, is found to be approximately 220pF. This value is some twenty times larger than that of the discrete photo detectors used with current encoders under similar operating conditions. This could in part be attributed to the grating metallisation being deposited directly onto the detectors rather than on a separate piece of glass. Consequently current pre-amplification electronics circuitry is not fully compatible with this device and modification is required. This work is ongoing and results presented in the proceeding chapters incorporate some changes; however full optimisation of the electronics is beyond the scope of this thesis.
7.2.3.1 Introduction of phase scale grating

In a final experiment the device was tested against a phase scale. Rather than the transmissive and reflective line segments of the Ronche grating, a phase scale has only reflective segments which differ in height by a quarter of the operating wavelength such that a phase delay of 180° is introduced, as illustrated in figure 7.23.

![Illustration of phase grating principle](Image)

Although inherently more difficult and expensive to manufacture, this design feature can be advantageous as the zero order beam is suppressed and the energy redistributed into the higher order beams, as demonstrated by equations (7.1) and (7.2) where $I_{0}^{th}$ and $I_{m}$ are the intensity for the zero and m\textsuperscript{th} diffraction orders, $s$ is the mark space ratio (the ratio of the length, $r_1$ and $r_2$ over the pitch $P_s$, such that if $r_1=r_2$ then $s=0.5$) and $\Phi$ is the associated phase change induced by the height $d$ in figure 7.23. Thus, as we are interested in the first order beam, the signal from the photo diodes is boosted. In addition more energy is gained because, with no transmissive segments, the number of reflections increases.

\[ I_{0}^{th} = 2s^2 - 2s + 1 + 2(s - s^2)\cos \Phi \quad (7.1) \]
\[ I_{m} = \frac{1}{\pi^2}(1 - \cos(2\pi sm))(1 - \cos\Phi) \quad (7.2) \]

![Poorly formed phase scale with damaged working area highlighted](Image)
The phase scale is fabricated by performing an argon plasma etch of the features in the glass wafer before applying a reflective coating metallic layer. As a first attempt the trial sample is, for convenience, e-beam evaporated with titanium to act as the reflective medium. Unfortunately the incorrect portion of the grating sample was copied across the glass wafer, leaving only a small portion of poorly metallised scale with the correct pitch as indicated in figure 7.24. Furthermore the accuracy of the glass etch was of low tolerance, ±8%. Figure 7.25 gives an indication of the error associated to deviation from 50:50 mark space ratio and phase angle errors resulting in etch depth and intensity errors. Assuming all else to be perfect and 8% error in etch depth corresponds to a drop in first order intensity from 41% to 31%.

Nevertheless, figure 7.26 shows the observed Lissajous curve, being circular in shape, with peak-to-peak voltage of 4.4V more than double the value achieved from the amplitude scale.

![Intensity Vs Grating depth for change in r and angle of incidence](image)

**Figure 7.25 – Intensity vs effective grating depth resulting from changes in mark space ratio and angle of incidence**

Further performance enhancement is therefore expected with better quality phase scale grating, having more accurate etch depth control and higher reflectivity metallisation. This also highlights a need for good control of the marked space ratio across all gratings. The benefits of phase grating against amplitude grating are investigated further in Chapter 8.
7.2.4 Chip on silicon carrier summary

The Si carrier trials provide further evidence to validate the integrated optical encoder chip concept. As with the glass carrier good circular Lissajous curves are obtained and mechanical tolerances are shown to better current Renishaw designs.

As expected, the increased heat capacity of Si results in lower surface temperature measurements, while a more stable signal response is observed from the photo detectors. Potential drive conditions are identified while additional tweaking of the pre-amplification circuitry and a change of scale from Ronche grating to phase grating serves to improve performance, particularly in relation to achievable signal strength and SNR.

The results and observations from these first test samples provide valuable information to take into second and third iteration chips where reference marks and some layout design alterations are considered.
Chapter 8
Design and Characterisation of the Reference Marker

The viability of an integrated optical encoder chip was demonstrated in the previous chapters. However, the usefulness of an incremental optical encoder having no reference marker is limited to only a few specific applications. It is therefore vitally important to integrate a reference mark scheme into the encoder chip design. Thus, in this chapter reference mark schemes including variations of auto-correlation, cross-correlation and zone plate designs are introduced and evaluated. Furthermore, the incremental operation of the encoder demonstrated in chapter 7 is for a 4μm grating pitch device, representing the high end of the optical encoder market. With the majority of optical encoder devices sold on a 20μm pitch this as well as an intermediate 8μm pitch device are also included in order to compare the encoders performance and demonstrate its flexibility.

8.1 Comparison of the incremental encoder chips

8.1.1 Set-up

The electronics have been upgraded from bulk components to more dedicated circuitry housed on a single PCB including preconditioning amplifier IC chips and differential amplifiers for signal processing for both incremental and reference mark signals. Additional modifications include on-chip shorting of the “monitor” diode and compensation for the increased junction capacitance. The chips are packaged into a Leadless Chip Carrier (LCC) package, as shown in figure 8.1, and mounted to a separate PCB having connectors for easy interchange of chip designs with the processing electronics board.

The chips are mounted into a Leadless Chip Carrier (LCC) package as either bare die or on a silicon interposer similar to the silicon tile discussed in chapter 7. The use of a silicon interposer serves two purposes here: firstly, it reduces the number of pin outs required from the LCC packaging resulting in a smaller PCB and secondly it raises the chip by approximately 0.5mm providing a reduced minimal stand off distance from the scale. However fault finding with chip on Si interposer packages proved to be
problematic as a fault from even a single diode failure propagates to all contacts on that particular channel, where as all contacts for the bare die in LCC package can be interrogated individually. Subsequently bare die packages are favoured for the validation of the new chip iterations with a view to using an interposer design as a potential first package solution.

The LCC package has a large metallised area providing a substantial heat sink onto which the chip or chip on Si is die bonded. The surface temperature of the chip is monitored as the drive current to the LED increases by fixing a thermocouple to the centre of the chip. Figure 8.2 shows a comparison of the bare chip and chip on a Si interposer. Little difference between the two regimes is observed. With potential operating conditions expected to be below 200mA this means the maximum surface temperature of the chip is less than 40°C and should therefore be thermally stable. In actual fact the maximum applied current is limited by the capacity of the 25μm Au wire connecting the LED, which is rated at 1A for a 2mm length in air. This value corresponds with the findings here.

Figure 8.1 – Leadless Chip Carrier packages for (a) bare die mounted optical encoder chip and (b) encoder chip on Si interposer
8.1.2 Grating pitch variation

The three grating pitches investigated are chosen to coincide with current market placement in mind. The 20μm pitch is the most commonly sold grating today and an 8μm pitch is at the high end of Renishaw’s competitor product offerings. A 4μm pitch, as demonstrated in chapter 7, is again considered as an encoder at this resolution would give Renishaw an advantage at the high value end of market as indicated in table 8.1.

Figure 8.3 shows the Lissajous curves obtained from the three grating pitches against both Ronche (amplitude) and phase scale gratings. The Ronche scale is once again obtained by sputtering Cr on glass. The phase scale now has an Au finish for the reflective coating with a more accurately controlled glass etch, thus providing higher reflectivity and increased zero order cancellation compared to the Ti on glass trial sample previously used.
<table>
<thead>
<tr>
<th>Manufacturer</th>
<th>Product</th>
<th>Scale Pitch (µm)</th>
<th>Maximum Move Speed (m/s)</th>
<th>Maximum Analogue Signal Frequency (kHz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Renishaw</td>
<td>RELM</td>
<td>20</td>
<td>10.00</td>
<td>1000</td>
</tr>
<tr>
<td></td>
<td>RGH-22</td>
<td>20</td>
<td>5.00</td>
<td>500</td>
</tr>
<tr>
<td></td>
<td>RGH-40</td>
<td>40</td>
<td>12.50</td>
<td>625</td>
</tr>
<tr>
<td>Heidenhain</td>
<td>LIP</td>
<td>4</td>
<td>0.50</td>
<td>250</td>
</tr>
<tr>
<td></td>
<td>LIF</td>
<td>8</td>
<td>2.00</td>
<td>500</td>
</tr>
<tr>
<td></td>
<td>LIDA</td>
<td>40</td>
<td>8.00</td>
<td>400</td>
</tr>
<tr>
<td>MicroE</td>
<td>Mercury</td>
<td>20</td>
<td>7.20</td>
<td>720</td>
</tr>
<tr>
<td></td>
<td>Mercury II</td>
<td>20</td>
<td>5.00</td>
<td>500</td>
</tr>
<tr>
<td></td>
<td>ChipEncoder</td>
<td>40</td>
<td>14.00</td>
<td>700</td>
</tr>
<tr>
<td>Numerik Jena</td>
<td>LIA21</td>
<td>20</td>
<td>10.00</td>
<td>1000</td>
</tr>
<tr>
<td></td>
<td>LIK21</td>
<td>20</td>
<td>10.00</td>
<td>1000</td>
</tr>
<tr>
<td>Canon</td>
<td>ML-16</td>
<td>3.2</td>
<td>0.32</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>ML-08</td>
<td>1.6</td>
<td>0.16</td>
<td>200</td>
</tr>
<tr>
<td>Renishaw</td>
<td>HiOCTANE</td>
<td>4</td>
<td>3.00</td>
<td>1500</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>6.00</td>
<td>1500</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>15.00</td>
<td>1500</td>
</tr>
</tbody>
</table>

Table 8.1 – Grating pitch comparison of Renishaw and competitors encoder products

As can be seen from the results, the use of a phase scale grating offers a number of performance advantages over the amplitude grating, such as signal strength, clarity, SNR and AC:DC. Indeed SNR greater than 50:1 have been achieved which means that simple 200 times interpolation can be implemented.

The AC:DC ratio appears to improve as the pitch reduces. When the current increases the AC:DC ratio improves with a phase grating scale but deteriorates with an amplitude grating scale. Furthermore the SNR increases with increasing current, in accordance with an increase in signal strength. Although ultimately a number of factors affect the signal strength and SNR, the achievable signal amplitude has a strong relationship to the voltage supply rail of the electronics circuit, ±5V in this instance. Peak to peak voltages of over 8.5V have been recorded before saturation begins to occur. A drastic decrease in SNR is however witnessed at the on set of saturation.
Figure 8.3 – Lissajous curves obtained from (a) the 4\(\mu\)m amplitude grating, (b) the 4\(\mu\)m phase grating, (c) the 8\(\mu\)m amplitude grating, (d) the 8\(\mu\)m phase grating, (e) the 20\(\mu\)m amplitude grating and (f) the 20\(\mu\)m phase grating

By contrast real world optical encoder products tend to be operated using a Lissajous curves of only 1Vp-p and have ASIC’s and other more sophisticated and complex electronics to further process and manipulate the signals. Lissajous curves of 1Vp-p are easily achieved, particularly against a phase scale grating such that it is anticipated that a supply rail of 0V to 5V will be used for the encoder chip. The influence on the signal
strength and SNR by the available supply voltage will require further investigation in order to optimise the electronics design and overall performance.

<table>
<thead>
<tr>
<th>Alignment</th>
<th>Current 20µm</th>
<th>Chevron 20µm</th>
<th>Current scaled to 8µm</th>
<th>Chevron 8µm</th>
<th>Current scaled to 4µm</th>
<th>Chevron 4µm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yaw (XY)</td>
<td>± 0.4°</td>
<td>± 0.55°</td>
<td>± 0.16°</td>
<td>± 0.25°</td>
<td>± 0.08°</td>
<td>± 0.08°</td>
</tr>
<tr>
<td>Roll (XZ)</td>
<td>± 0.5°</td>
<td>± 4°</td>
<td>± 0.2°</td>
<td>± 1.75°</td>
<td>± 0.1°</td>
<td>± 0.9°</td>
</tr>
<tr>
<td>Pitch (YZ)</td>
<td>± 0.5°</td>
<td>± 2.75°</td>
<td>± 0.2°</td>
<td>± 1.35°</td>
<td>± 0.1°</td>
<td>± 0.8°</td>
</tr>
<tr>
<td>Stand-off</td>
<td>± 0.08mm</td>
<td>+1mm</td>
<td>± 0.03mm</td>
<td>+1mm</td>
<td>± 0.02mm</td>
<td>+1mm</td>
</tr>
</tbody>
</table>

Table 8.2 – Comparison of full width half maximum mechanical tolerances for the current parallel detector arrangement against the new chevron detector arrangement with 20, 8 and 4µm scale grating pitches

The FWHM mechanical tolerances measured against phase scale for each grating pitch are presented in table 8.2. The results again highlight the improvements achieved in using the new chevron layout. In particular the current 20µm encoder tolerances can be directly compared against the new arrangements and the benefits for each alignment are clear to see. The tolerances roughly scale down with pitch size at which point the slight advantage in yaw tolerance becomes less obvious while pitch and roll tolerances remain extremely good.

8.2 Investigation of the reference mark schemes

This section discusses the design and performance of the reference mark features, essential to the successful operation of the optical encoder. There are many implementations of the reference marks including precision switching, magnetic and Hall Effect sensors and optical patterns either separate or within the scale and variations thereof. Table 8.3 provides some advantages and drawbacks of the different techniques.

The target market for this optical encoder is intended for applications requiring high accuracy systems, so it is essential to have a bi-directional reference marker with as high a resolution as possible. The miniaturisation of the encoder also creates a level of compactness that must be maintained at the system level so the most compact scale possible is also desirable. For these reasons the reference scheme is also optical and the optical pattern has to be of similar form to the scale. The most compact solution would be to embed the reference pattern into the incremental scale. This feature is required to
produce a distinct and repeatable signal while, at the same time, it must not prevent the incremental element of the scale from functioning.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Precision switch</th>
<th>Magnetic and hall effect sensors</th>
<th>Optical: Separate pattern</th>
<th>Optical: In scale pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advantages</td>
<td>• robust</td>
<td>• robust</td>
<td>• EMI immune</td>
<td>• EMI immune</td>
</tr>
<tr>
<td></td>
<td>• EMI immune</td>
<td>• fairly EMI immune</td>
<td>• Bi-directional</td>
<td>• Bi-directional</td>
</tr>
<tr>
<td></td>
<td>• simple to</td>
<td>• inexpensive</td>
<td>• Unit of resolution</td>
<td>• Unit of resolution</td>
</tr>
<tr>
<td></td>
<td>implement</td>
<td></td>
<td>accuracy</td>
<td>accuracy</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drawbacks</td>
<td>• Relatively</td>
<td>• Requires setup</td>
<td>• Generally requires</td>
<td>• Generally requires</td>
</tr>
<tr>
<td></td>
<td>expensive</td>
<td>• Not bi-directional</td>
<td>made to order scale</td>
<td>made to order scale</td>
</tr>
<tr>
<td></td>
<td>• Not bi-</td>
<td>• Can suffer from Abbe error</td>
<td>• Can suffer from</td>
<td>• Can suffer from</td>
</tr>
<tr>
<td></td>
<td>directional</td>
<td></td>
<td>Abbe error</td>
<td>Abbe error</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Adds to scale</td>
<td>• Adds to scale</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>dimension</td>
<td>dimension</td>
</tr>
<tr>
<td>Other</td>
<td>• accurate to</td>
<td>• accurate to</td>
<td></td>
<td></td>
</tr>
<tr>
<td>comments</td>
<td>~0.5μm</td>
<td>~0.05μm</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 8.3 – Comparison of different reference mark implementations

The optical reference schemes investigated in this work include variations of auto-correlation, cross-correlation and zone plate configurations. The reference features on the scale are achieved by interrupting the incremental features as explained from the description of the schemes below. Although discussed together under the reference regime context, the actual designs presented evolve from numerous iterations some designs taking full benefit from the previous results. However, for clarity, the discussions focus on the performance of reference marks under the categories defined below.
8.2.1 Chip reference marker configurations

8.2.1.1 Z-type chip

Figure 8.4 illustrates a Z-type chip. This chip comprises banks of photo detectors located to the left and right of a central incremental LED in relation to the x-axis translation. The index and analyser gratings are then re-orientated by 90° in order for the incremental features to match the direction of travel. A second LED for reference detector illumination of the zone plate reference detectors are presented in this design. Each of the three grating pitches has been included.

![Figure 8.4 – Image of a zone plate reference mark chip](image)

The zone plate reference marker on the scale, illustrated in figure 8.5, is patterned to act as a lens focusing the reference LED image onto the split zone plate reference photo detectors sequentially as the chip moves relative to the scale as illustrated in figure 8.6.

![Figure 8.5 – Illustration of a zone plate scale reference mark feature](image)
8.2.1.2 A-type chip

Figure 8.7 provides examples of two A-type chips: 1) one orientated as before with the incremental test chips having two detector banks, one located above and one located below central incremental LED, reference LED and reference detectors and 2) a second configuration whereby the gratings are re-orientated as with the Z-type chip. Both operate as illustrated in figure 8.8.

For each orientation, the reference schemes are for a 4-bit and 7-bit divergent beam auto-correlation patterns. The reference detectors are appropriately patterned and offset against each other such that, as the chip passes, the scale reference mark, see figure 8.9, the two offset signals are generated. The sum of these signals is used to gate a zero crossing that occurs on the difference of the signals, which acts as a trigger to set a known position. Ideally the signals should provide a narrow gate with a steep slew rate (quick transition) between peak and trough of the difference signal.
Figure 8.7 – Image of A-type optical encoder chip examples with (a) a 4-bit reference pattern and (b) a 7-bit rotated reference pattern

Figure 8.8 – Illustration of auto-correlation reference scheme operation
Figure 8.9 – Illustration of A-type scale grating reference mark features for (a) 4-bit and (b) 7-bit patterns

Figure 8.10 illustrates simulations of the ideal signals generated from the 4-bit and 7-bit auto-correlation markers. While the 7-bit scheme has a nice sum peak and short slew rate, discrimination of the features on the scale could be a problem. On the other hand the 4-bit scheme gives better discrimination between features but the shorter pattern results in much of the power going into the side lobes and the sum gate being quite broad, which could lead to false triggers.

Figure 8.10 – Simulated ideal sum and reference signals from given scale features for (a) 4-bit and (b) 7-bit patterns

Restrictions imposed by the wafer real estate mean that for each scenario, 4-bit, 7-bit, horizontal orientation and vertical orientation an 8μm pitch device is available. One 20μm pitch device with a 7-bit pattern is also made while there are no 4μm pitch devices to study.
8.2.1.3 A2-type chip

The A2-type chips, shown in figure 8.11, are also auto-correlation patterns consisting of 4-bit, 7-bit and 13-bit patterns. However, in this case, only rotated chips with 8μm pitch are investigated. The principle of the auto-correlation scheme is as discussed for the A-type chip but here a novel approach for the illumination of the scale considers the LED to serve as a source for both the reference scheme, in a convergent mode, and the incremental function. The chip dimensions are also slightly increased to accommodate larger reference photo detector banks located outwith the confines of the central reservation between the incremental detector banks. The subsequent scale patterns used are illustrated in figure 8.12.

Figure 8.11 – Image of A2-type optical encoder chip with 7-bit pattern

Figure 8.12 – Illustration of A2-type scale reference mark features for (a) 4-bit, (b) 7-bit and (c) 13-bit patterns
8.2.1.4 B-type chip

The B-type chips, shown in figure 8.13, use a cross-correlation scheme, see figure 8.14, whereby the code or part of the code is removed from the scale and formed by specifically patterning the LED source. The generated sum and difference signals follow the same principles as those discussed for the A-type schemes. For the B-type chips 7-bit and 13-bit patterns are generated when used against the two scale patterns illustrated in figure 8.15. Only the top/bottom incremental detector bank orientation are used on an 8μm pitch.
8.2.1.5 C-type chip

This type can be described as a modified cross-correlation scheme but a binary type code is not employed. Instead it is a variation on the Young’s effect principles whereby the light diffracted at two slits or reflecting segments on the scale, see figure 8.17, are detected at three reference detector locations shown as x, y, and z in figure 8.16. When fully aligned the central, y, detectors will have twice the intensity of the left and right detectors, see figure 8.18 for an illustrated example.
The central detector, \( y \), would actually be a split detector from which the difference pulse would be generated, while the sum signal would be generated from the outer \( x \) and \( z \) detectors. From figure 8.16 the detector banks located above and below the central LED section are now separated into four sections, to allow the inclusion of the reference detectors while maintaining the compact design. Once again only 8\( \mu \)m pitch gratings are studied.

8.2.1.6 D-type chip

Type D chips are essentially a variation of type C, the difference being that the incremental LED sections also serve to provide the reference signals and the reference detectors are positioned wider apart, as shown in figure 8.19, resulting in a slight increase in chip dimensions. Similarly the scale reference mark requires adjustment, with wider spacing between the features. Also the features are retained between the reference marks, as shown in figure 8.20, Once again only 8\( \mu \)m pitch versions are studied here.
8.3 Chip reference marker results

For clarity the performance results of the reference mark designs discussed follow the order given in the classifications above rather than chronologically. Some chip results have been used in the creation of other types of chips. Consequently some design concepts have therefore received more attention as they have been iterated more than once.

8.3.1 Z-type chip

Figure 8.21 shows the recorded sum and difference signals for the zone plate reference mark scheme on a 4μm, 8μm, and 20μm pitch grating, respectively. In each example a reasonably well defined difference signals is achieved, whereas the sum signals is not so clear. The gate period as highlighted by the parallel lines and used to determine the
slew rate feature of the reference pulse. The number of oscillations of the incremental signal can be counted as the difference signal traverses from peak to trough or vice versa through the zero crossing giving an indication of the length of travel required for the active reference signal. For a 20μm pitch scheme, a good count would be only 10 oscillations, whereas 15 oscillations are considered as the upper limit of acceptable tolerance. The 8μm and 4μm pitch versions fare poorly with an oscillations count increasing to 30 and 45 as the pitch decreases.

Additional observations worth noting are a slight decrease of the incremental signal strength, which can be explained by the loss of incremental features on the scale to the reference pattern. While observing the Lissajous curve as the reference mark is passed this reduction is also observed in the size of the circle. However, it was also observed that the curve moves about the x-y axis indicating undesirable changes in the DC levels. Offset correction techniques can compensate to a degree but the level of movement observed here is deemed unacceptable. If the DC voltage across each channel was identical then the effect would cancel in the electronics and a steady state Lissajous would remain. However, what is observed is an indiscriminate fluctuation as the reference mark is traversed resulting in an unstable Lissajous position. Consequently, should correction mechanisms prove inadequate, interpolation would not be possible. For example, if the reading was expecting a negative cosine value the addition of the DC voltage fluctuation could cause an offset resulting in a positive cosine value thus inflicting positional measurement and accuracy errors.
8.3.2 A-type chip
As with the signals from the zone plate the sum signals of the A-type chip variants are weak and unclear. Although a difference signal can be observed for each of the five A-type cases these signals bear little resemblance to the expected signals shown in figure 8.22. In addition the slew rates are long covering 60 oscillations for each of the 8μm examples, dropping to 25 oscillations for the 7-bit on 20μm pitch example.

Figure 8.21 – Recorded Incremental, sum and difference signals on (a) 4μm, (b) 8μm and (c) 20μm pitch scale gratings for the zone plate reference mark scheme
Figure 8.22 – Incremental sum and difference signals from (a) 4-bit pattern on a 8\(\mu\)m pitch, (b) 4-bit pattern on a 8\(\mu\)m pitch with rotated orientation, (c), 7-bit pattern on a 8\(\mu\)m pitch, (d), 7-bit pattern on a 8\(\mu\)m pitch with rotated orientation and (e) 7-bit pattern on a 20\(\mu\)m pitch for the A-type auto-correlation reference mark schemes
Further investigation reveals that the root cause of the poor reference signal performance lies with the transmission source. These schemes rely on the divergent beam profile originating from the reference LED. However this LED is relatively small in comparison to the main incremental LED and fails to produce enough power such that the measured signal is dominated by light from the incremental source. Subsequently the layout of the reference detectors in relation to the incremental source means that the incremental LED signal is too extended thus failing to provide a fast slew rate and the difference signal.

Consequently divergent schemes such as this and the zone plate, designed with LED and reference markers positioned centrally along side the incremental LED are considered unsuitable for meeting the goals set out for this task. The divergent scheme could possibly be made to work by placing the reference LEDs and reference detectors outside the perimeter of the incremental detector arrangement. However this would require larger chips and additional out of track reference features on the scale compromising the compact design targets for each.

8.3.3 A2-type chip
The observed signals from the A2-type chips immediately reveal frailties with the capability of the scale patterns that can be implemented. In figure 8.23(a) and (b) the incremental signals drop away to zero as the reference marks are passed rendering these designs unusable. However the incremental signal, shown in figure 8.23(c), still shows a signal albeit by an unacceptable 45% drop in strength. This effect is attributed to the scale patterns as presented in figure 8.12. The 13-bit and 4-bit patterns interrupt the incremental features with fully reflective and non-reflecting segments where as the 7-bit pattern, 8.12(b), merely interrupts the incremental signal periodically with reflecting or indeed non-reflecting segments as appropriate, such that the percentage of incremental signal is proportionate to the loss of incremental features. It therefore stands that subsequent designs employing such correlation patterns require careful design to ensure that losses to the incremental signal are minimised by removing as little of the scale grating incremental features as practically possible. As loss of the incremental signal appears inevitable an acceptable level that does not compromise the accuracy and performance of the system needs to be found.
Despite the unacceptable loss in incremental signal strength, strong reference signals are recorded. The sum and difference signals are compared against expected signals in figure 8.24. Notwithstanding some undesirable peaks and troughs these reference signals are reasonably well matched with the sum signal gating well over the zero crossing of the difference pulse, giving a workable solution. However the slew rate covers approximately 19 oscillations and would require reduction. Furthermore some significant fluctuation of the D.C. voltage can be observed from the incremental signal, the level of which may be too quick to correct rendering this scheme potentially useless.

Figure 8.23 – Single channel incremental and difference signals observed from (a) 13-bit, (b) 4-bit and (c) 7-bit patterns for the A2-type chip auto-correlation reference mark scheme
The ripples in the simulated difference signal obtained in figure 8.24 can be attributed to the width of the photo detector aperture. The narrower the detectors the narrower the gate and steeper the slew/trigger with better side band suppression at the cost of some ringing.

8.3.4 B-type chip
The difference signal observed in figure 8.25(a) for the B-type against the bi-reference mark scale is significantly different from an expected 7-bit correlation pattern, illustrated in figure 8.10. The error is traced back to a mistake at the mask level design stage, where the overlay of the copied grating structure inadvertently deleted part of the desired design such that the LED pattern shown in figure 8.26(a) was fabricated rather than the desired design shown in figure 8.26(b)

Figure 8.24 – Comparison of simulated and recorded (a) sum signals and (b) difference signals for the A2-type 7-bit pattern

Figure 8.25 – Single channel, incremental and difference signals from (a) reference mark scheme 1 and (b) reference mark scheme 2 for the B-type cross-correlation patterns
Recalibrating for this error showed a reasonable match between measured and expected results however these values have no relevance towards a working reference mark solution. On a positive note the signals are well defined and there is very little fluctuation of the incremental signal due to the single narrow reference feature present on the scale grating. Consequently, at this stage, the validity of the scheme remains undetermined and requires further investigation.

Similarly the LED design is erroneous for use with the second B-type, as the bi-scale reference pattern and the observed signals, shown in 8.25(b) are of no real value. Moreover, as this scale pattern contains some of the reference code it is wider, and removes all incremental features over the period. This feature, as before, causes a drop of the incremental signal by up to 80% and would occur regardless of corrections to the LED design. Therefore, while the B-type design with bi-scale may merit further investigation the second example is considered to be unusable for the intended application.
8.3.5 C-type chip

Unfortunately a design error has meant that a single photo detector has been fabricated rather than a split photo detector. Signals are observed from each of the reference channels as shown in figure 8.27. Using the signal data from the y detector two signals can be produced, by copying and shifting appropriately, as shown in figure 8.28, to create a y- and y+ from which the difference pulse is generated, while the real data from x and z detector channels provides the sum signals.

![Image](IncrementalSingle_detector_channels.png)

**Figure 8.27 – Signals observed from C-type cross-correlation reference mark scheme**

![Image](y-y+_response.png)

**Figure 8.28 – Manufactured signal response for a split y photo detector based on the measured y reference detector signal**

Figure 8.29 plots these reference pulses together and a zero crossing with sum gate having reasonable alignment can be observed as indicated. However by the nature of the design described in section 8.2.1.5 one would expect three zero crossing transitions with the largest peaking in the centre rather than at a side lobe as seen here. The most likely cause for the discrepancy comes from maintaining the tight separation of the reference detectors and alignment against the reference source. A higher level of separation, as applied to the D-type design, has more relaxed tolerances and should result in improved performance.
8.3.6 D-type chip

As with the C-type chip an oversight has meant that the central reference photo detector, y on figure 8.19, has been manufactured as a single diode rather than a split detector. Additionally these D-type chips have suffered from a shorted connection at this same diode so no signal response could be measured.

In order to gain an appreciation of the reference schemes potential it can reasonably be assumed that a signal similar to those, shown in figure 8.30, observed on the x and z detectors would also be present on y. Therefore the data obtained from the x and y responses can be used to mimic the y response, the y detector being a split detector with the signals located between the x and z signals. Using the available data the split signal is generated by moving the signal first to the centre location. From here two identical copies are moved by the appropriate distance to create y- and y+ signals as was performed for the C-type chip and shown in figure 8.31.
The difference between these two signals creates the zero crossing reference pulse while the sum of the x and z are used as a gating mechanism with an option to use the two outer lobes as potential early warning triggers. This would result in the sum and difference signals presented in figure 8.32.

As can be seen the sum is gated well over the zero crossing of the reference mark and a slew rate covering 15 oscillations is achieved. The outer lobes and zero crossing also have potential. Furthermore comparison with ideal signals is provided in figure 8.33 and shows good correlation. As can be seen in figure 8.30 there is some significant wobble on the incremental signal that would need to be compensated for. Should this be achieved and should corrections to the y-node photo detector prove effective, the D-type reference scheme would provide a viable solution.
8.4 Additional observations

During these measurements, the Lissajous curves expanded and collapsed several times as the device is moved through the yaw angle of rotation. Figure 8.34 demonstrates the occurrence of Lissajous curves with rotation angle and their relative peak-to-peak strength. It is by the very nature of the design i.e. two chevron detector array banks located about a centrally located LED, that this effect is observed. As the encoder is rotated through the X-Y axis the fringes move in and out of phase. The central maximum peak indicates perfect alignment while the secondary peaks indicate the position where the rotation is such that diagonally opposite quarters of detector banks are either in or out of phase. The example shown here is for the 8μm pitch case and the secondary peaks occur within a tight range, less than ±0.5°, and reach 65% of the central maximum. The size of these secondary peaks could prove problematic as considerable care would be required to ensure that alignment takes place at the correct position.
Particularly for the arrangements where the photo detector banks are located above (north) and below (south) of the centrally located LED this problem can be alleviated by processing and re-phasing of the north and south arrays separately. This technique broadens the central peak while suppressing the secondary peak as demonstrated in figure 8.35. As well as removing the possibility of alignment on the wrong peak, the yaw tolerance is considerably improved. This is further highlighted in table 8.4 where the separately processed detectors full width half maximum mechanical tolerances are compared against the previous findings. However these benefits come at the cost of increasing the capacity and complexity of the processing electronics to perform the re-phasing; automatic phase correction will also be required.

![Figure 8.35 – Magnitude of Lissajous curves as the encoder is rotated through the X-Y axis for separately processed detector banks](image)

<table>
<thead>
<tr>
<th>Alignment</th>
<th>Previous 8μm</th>
<th>B-type</th>
<th>C-type</th>
<th>D-type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yaw (XY)</td>
<td>± 0.25°</td>
<td>± 0.55°</td>
<td>± 0.55°</td>
<td>± 0.55°</td>
</tr>
<tr>
<td>Roll (XZ)</td>
<td>± 1.75°</td>
<td>± 2°</td>
<td>± 2.25°</td>
<td>± 1°</td>
</tr>
<tr>
<td>Pitch (YZ)</td>
<td>± 1.35°</td>
<td>± 1.1°</td>
<td>± 1.45°</td>
<td>± 1°</td>
</tr>
<tr>
<td>Stand-off</td>
<td>+ 1mm</td>
<td>+ 1mm</td>
<td>+ 1mm</td>
<td>+ 1mm</td>
</tr>
</tbody>
</table>

Table 8.4 – Full width half maximum mechanical tolerances for separately processed detector banks

Table 8.5 provides a summary of some additional observations. Despite the differences in LED and detector configurations each design shows a near identical power requirements for driving the LED and achieves a 1Vp-p Lissajous curve. The
roundness of these curves, which give an indication of the balance between the phase and quadrature outputs, are found to be greater than 85%. In future developments of the processing electronics this figure should be improved upon. The signal-to-noise ratios are found to fall just short of the 50:1 required for implementation of 200X interpolation. Ideally this value will eventually top 100:1 so clearly much work is required to optimise the performance of both the chips detectors and the electronics for this to be achieved.

<table>
<thead>
<tr>
<th>Chip type</th>
<th>Power (mW)</th>
<th>Roundness</th>
<th>Worst case SNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>37.6</td>
<td>96%</td>
<td>45:1</td>
</tr>
<tr>
<td>B</td>
<td>40</td>
<td>85%</td>
<td>49:1</td>
</tr>
<tr>
<td>C</td>
<td>39.2</td>
<td>85%</td>
<td>40:1</td>
</tr>
<tr>
<td>D</td>
<td>38.8</td>
<td>91%</td>
<td>41:1</td>
</tr>
</tbody>
</table>

Table 8.5 – Summary of additional observations for 1V<sub>p-p</sub> Lissajous curves

8.5 Summary

The optical encoder chips have been successfully bonded into LCC packages for test. The packages allow ease of handling and provide good heat sinking capability. With the addition of a glass cover slide to help protect both the chip and wire bonds the LCC package makes a good early candidate as a final wire bond package solution.

The range of chips designs studied in this set of experiments increases, from those presented in chapter 7, to include grating pitches of 4μm, 8μm and 20μm. The incremental signals are shown to operate well particularly against a phase scale grating where signal strengths and SNR are dramatically improved in each case. The largest improvement is observed for the 20μm pitch grating where SNR reaches 65:1, an 8 times improvement factor. The FWHM mechanical tolerances also demonstrate the performance benefits achieved with the chevron detector design layout. A direct comparison of current 20μm parallel line technology against the chevron equivalent design shows significant improvements in each rotation. These measurements scale accordingly with pitch when all detector diodes are connected in common. Furthermore additional advantages have been identified whereby the detector banks on either side of the LED are initially processed independently with the common signals to be rephased.
electronically. This technique, demonstrated with 8\(\mu\)m pitch samples, shows that a significant gain can be made particularly for the yaw motion where the effective tolerance is doubled, surpassing what is achievable with the current 20\(\mu\)m encoder devices. However rephasing of the signals has yet to be performed and the electronics required need to be integrated into the current circuit design.

The main focus of this chapter has been on the development of a suitable reference mark design. Essentially six variations are presented that utilise four key design techniques of zone plate, auto-correlation, cross correlation and modified cross correlation configurations.

The reference marker on the scale is traditionally located outwith the body of the incremental scale pattern. However this would compromise the miniaturisation as larger chips would be required in order to operate against such a scale design. Therefore the development work also investigates the possibility of embedding the reference features within the incremental section at the cost of a drop of the incremental signal strength as the reference feature is passed. There is also some concerns over the levels of D.C. voltage drift introduced to the incremental signal that could compromise the use of interpolation techniques by causing erroneous position and movement measurements.

Potential methods to overcome these issues have been indicated. Where the reference marks on the scale retain incremental segments within the reference pattern and where narrower reference features are used, the level of signal drop and D.C. voltage movements can be reduced. In the light of these discoveries the B-type chip design is encouraging but unfortunately the error in the LED configuration prevents verification at this moment in time.

While the B-type chip appears, theoretically at least, to have potential, the results obtained from the A2 and D-type chips, having incremental segments within the reference marks, also show signs of promise. On the other hand the zone plate and A-type chips, which have individual reference LED’s, as well as the C-type chip have been shown to be unsuitable. The need to manufacture a patterned LED structure is highlighted by this result.
Presently some clear and distinct difference signals have been observed. However the sum signals do not have the same level of common mode noise reduction and, as such, appear nosier while also suffering from the phenomena of beats. In order for the signals to become more useable some filtering arrangement may be required. Also the signals slew rates are presently too wide, covering in excess of 15 oscillations, and must be reduced to maintain high accuracy.

Clearly no one reference design performs outstandingly above all others and a lot of further development work is therefore required. Additionally the influence of the reference mark on the achievable mechanical tolerances, yaw and stand-off degradation in particular, must also be considered. Development work to be performed in the near future on A2, B and D-type reference regimes is discussed in chapter 9.
Chapter 9

Conclusions and Future Work

This section provides a review of the previous chapters and discusses possible future work both to be carried out for this particular encoder and for the next generation of encoders.

9.1 Review summary

9.1.1 Integrated encoder chip – design and development

The development towards a novel optical encoder has been described and discussed in this thesis. The advantages of microsystems manufacturing techniques are utilised to monolithically integrate the encoder components: light source, photo detectors and optical gratings, onto a single semiconductor chip. Several LED/photo detector layouts have been considered but it is the chevron detector array layout, arranged about a centrally located LED that offers the best performance. The integrated design provides a number of advantages over its traditionally manufactured counterpart including, but not restricted to, smaller size and form factor, reduced material waste, wafer scale alignment of the optical grating to the photo detector array in a single process step and reduced costs.

Packaging options are discussed and, as such, the chip layouts are designed to be compatible with both wire and flip-chip bonding technologies. Flip-chip bonding offers a number of advantages over wire bonding both in terms of performance and flexibility for advanced packaging options and is the preferred method. However a suitable flip-chip bonding solution requires further development beyond the time frame of this project. In the short term a wire bonded solution offers a fast effective means to perform the characterisation and testing of the encoder chip design in order to facilitate the development process.

Prototype devices, manufactured in indium phosphide, having 4µm pitch gratings over both the LED and photo detectors, have been wire bonded to glass and silicon submount chip carriers for initial test, characterisation and verification. The submounts serve two purposes: 1) to allow the testing of the prototype device and 2) to assess the suitability
as a potential interposer in future packaging applications. Discrete component testing confirms the operation of the LED and photo detectors on the same substrate and also highlights the importance of heat dissipation to the signal stability. Consequential failings of the glass submount mean that silicon is considered to be the more suitable of the two candidate materials. This becomes further evident as the incremental operation of the encoder chips is verified. Although Lissajous curves are observed for each case and demonstrate improved mechanical tolerances larger signals are obtained from the chip on silicon configuration. After modification of the electronics circuitry and implementing reverse bias operation of the photo detectors larger, cleaner signals are obtained with increased bandwidth. The measured signals are still further improved upon with the introduction of a phase scale grating, with peak to peak voltage measurements more than doubled compared to those obtained with a Ronche scale.

After proving the feasibility of the optical encoder on a chip and verifying the incremental operation of the device, the reference mark schemes have been introduced. Such markers are vital to the successful operation of an optical encoder system as they provide known locations on the scale from which all positional measurements are acquired. At this stage, a ceramic leadless chip carrier has been identified as a potential packaging candidate should wire bonded devices be brought to market. Chips in LCC packages are firstly used to verify incremental operation of a wider range of grating pitches with 8μm and 20μm also introduced. Signal to noise ratios greater than 50:1 and AC:DC ratios of the order of 5:1 have been achieved.

Several reference mark designs based on auto-correlation, cross-correlation and split diode techniques have been discussed. These designs are then tested against phase scale grating that has the reference features embedded within the incremental pattern. While no one reference design performs above all others, some key design issues are identified. Firstly a divergent beam operated scheme proves inefficient with the slew rate of the difference signal being too extended and compromising the positional measurement. Secondly, blank spaces in the reference mark design should be filled with incremental features rather than left blank otherwise the strength of the incremental signal becomes compromised. The reference features should be kept as tight as practically possible to avoid any fluctuation of the D.C. voltage of the incremental signals. Finally the use of an independent reference light source is shown to be inefficient as the signal is swamped by the main incremental light source. This last
observation opens up the novel approach whereby the incremental LED is also patterned to form part of the reference design.

In addition to reference mark performance a method of increasing the yaw tolerance still further has been discovered. By processing the detector banks separately and re-phasing the signals later the bandwidth of the yaw angle has become twice the mechanical tolerance of current Renishaw encoder products. Of the designs discussed, three are identified as having the potential to succeed as a reference mark scheme and put forward for further development work. This includes an auto-correlation scheme and two cross-correlation approaches.

9.1.2 Flip-chip bonding process development

After establishing the feasibility of the monolithically integrated optical encoder, work has commenced on establishing a suitable flip-chip bonding process. Assessment of the technologies and available equipment identifies solid-state diffusion bonding of gold to gold interconnects by thermocompression as the most suitable candidate, whereby the gold bumps are formed using an adapted LIGA process utilising DC electrodeposition.

Modifications have been made to the equipment to improve the deposition process and the bath conditions. Resist mould bake-out time and bath temperature have been considered as the most critical process parameters. The full extent of the influence of applied current density has yet to be investigated. Bumps have been fabricated with a 4μm height difference on the same wafer with a smooth surface finish. However current crowding effect at the mould edge at the deposition interface has resulted in a basin like surface profile with the edge of the bump being approximately 1.5μm taller than at the centre. To improve the joint quality a flatter profile is required.

For the flip-chip bonding process, a redesign of the pick-up tool to include a perforated membrane vacuum slot has been carried out to remove detrimental stress effects induced on the chip with the application of bonding force. With limited sample availability a select few bonding profiles have been implemented, some of which are influenced by results obtained from an early stage finite element model. This suggests that the bump/bonding temperature is directly affected by the heat applied through the InP chip as apposed to the substrate, when the substrate is glass.
To date flip-chip bonded samples have demonstrated 100% connectivity and shear strength of 1220gF. Individual bump strengths in the order of 100gF, three times the JEDEC standard minimum, are also measured. However the quality of the bonds is compromised by a non-planarity flaw discovered in the pick-up tool manufacture. By resolving this issue and coupled with the fabrication of flatter bump profiles improvements to the joint quality are expected and the production of a flip-chip bonded demonstrator module should be imminent.

9.2 Future work

The original objectives set out at the beginning of this project have been partially fulfilled. Crucially the feasibility of the concept has been proven and the advantages of monolithically integrating the components of the optical encoder justified. However the work has just fallen short of producing a fully packaged optical encoder readhead. To this end a number of key issues are to be addressed including additional work on the reference mark design optimisation and flip-chip bonding process improvements.

Further characterisation will continue to be performed with wire bonded devices in LCC packages while the first generation flip-chip bonded on glass demonstrators will be mounted on similar PCB’s modified to accommodate the chip on glass module. This field of work will be completed in the near future and is discussed in further details below.

9.2.1 Reference mark and chip design development

The new chip layouts, shown in figure 9.1, for the three reference mark schemes identified for further development in chapter 8, have been completed and submitted for manufacture. For the A2-type chips, little in the way of optical layout has changed except to the scale grating, to include incremental features in the reference mark spaces. In addition more consideration needs to be given to the bonding regime whether it be wire bonding or flip-chip bonding that is used. The bond pads are more evenly distributed along each edge of the chip reducing congestion for wire bonding and helping to achieve a more uniform load distribution during flip-chip bonding.
Additionally dummy bond sites are introduced in more central locations to again alleviate bonding stresses.

![Diagram of reference mark chip designs](image)

**Figure 9.1 – Schematics illustrations showing the next generation reference mark chip designs for (a) A-type, (b) B-type and (c) D-type chips**

The optical layout of the B-type chip remains unchanged but the LED pattern error is now corrected. Similar considerations towards the bump pad positioning are made. For
the D-type chip the six reference detector diodes that split the incremental detector arrays are replaced with three detectors that are repositioned towards the centre of the chip. While the reference mark design concept remains the same this change should improve the yaw tolerance of the reference scheme while the increased number of incremental detector area will boost the incremental signal strength.

Once the reference mark design is confirmed, work can progress on establishing the maximum levels of interpolation that can be applied and thus the maximum resolution of the system. The encoder will also be operated against a significant length of scale such that speed, accuracy and reliability can all be determined.

Other significant design changes include the planarisation of the bond pads thus removing the need for a 4µm bump height difference. The implications to the bump fabrication process include reduced process times, reduced exposure to the plating bath conditions reducing the risk of photoresist mould damage, and could also lead to the negative resist mould solution being revisited.

There will also be a move from InP to a GaAs substrate. As discussed in chapter 4 the primary reason for choosing InP was to allow for the co-design of a substrate emitting version of the optical encoder. As this was ultimately unsuccessful the economies of scale associated with a move to GaAs make logical financial sense for the surface emitting device. In today’s market GaAs is not only cheaper than InP the processing capabilities are typically for 6-8 inch wafers or greater while InP is currently limited to an absolute maximum of 4 inch but more commonly 2-3 inch. The risk associated with a change lies with CST’s ability to transfer the manufacturing process from InP to GaAs. Although CST have experience in processing GaAs, it is recognised that the process is inherently more complex and will require a degree of development work before working die are available.

In addition to the reference mark development an optical design engineer at Renishaw has identified a design opportunity for the incremental grating, more specifically for the analyser grating features over the chevron detectors. By introducing a marginal degree of offset to the features the yaw angle tolerance can be greatly improved. However, as shown in figure 9.2, this comes at the expense of peak to peak signal strength. The figure shows a comparison of a stepped and smooth offset designs against the original
signal as well as the current de-phased set up, which improved the tolerance to $\pm 0.55^\circ$. As can be seen this new approach could potentially increase the yaw angle tolerance to greater than $\pm 2^\circ$.

Figure 9.2 – Simulated yaw angle tolerance improvements as a result of analyser grating modification

9.2.2 Flip-chip bonding development

While characterisation of the reference marks takes precedence the change to larger GaAs wafers will release significantly more samples for flip-chip bonding studies. As discussed above, a primary concern is with the levelling of the surface profile of the bumps to facilitate improved interfacial contact during bonding. For the DC set-up described in this thesis, manipulating the bump shape will be attempted by employing patterned electrode designs at the deposition pads. Presently deposition is made directly on to the gold seed layer pad. The pad will be patterned exposing some of the underlying seed layer, presently Ti, to investigate if the influence on the conductive path can be used to shape the bump. Possible patterns are illustrated by way of example in figure 9.3. Should results prove positive the addition of materials to the seed layer may also be investigated.
An additional experiment will investigate the effect of exposing the resist to further development towards the later stages of the plating cycle. This should slightly widen the mould cavity and ease the effect of current crowding at the edge.

An alternative to DC electrodeposition will also be considered. Pulse plating or more specifically pulse reverse plating will be investigated in collaboration with CST. Applying a reverse pulse performs an in-situ electrochemical polish by first depositing material in the forward part of the duty cycle and then removing some when the polarity is reversed. This may lead to slightly increased plating durations but higher quality bumps, both structurally and electrically, can be achieved. For pulse plating to be implemented some capital equipment needs to be purchased and a control interface is to be developed using LabView at CST.

After these changes are implemented and a robust flip-chip bonding process established attention turns to lifetime reliability study. This shall include tests such as MTTF, thermal cycling, humidity, vibration, shock and such like.

9.2.3 Advanced Packaging

While the wire bonded optical encoder chip in LCC package is useful for characterisation and perhaps even as an entry level product, more advanced packaging solutions are required to ensure that the dimensional gains achieved through monolithic integration are not undone by bulky packaging.

In chapter 4 idealistic packaging solutions utilising the advantages of LTCC technology are presented and the benefits of flip-chip bonding the optical encoder chip discussed.

Figure 9.3 – Examples of patterned electrode designs for electroplated bump shaping
The feasibility of employing LTCC to this application is unknown and, with this in mind, work has been initiated to develop an LTCC solution. The first prototype will be a simplified version of that presented in chapter 4, whereby only the encoder chip on glass will be packaged as illustrated in figure 9.4. Signals and power will be delivered using a flat flexi-cable such that a packaging profile of less than 4mm is maintained and features included allowing easy mounting to the sub-assembly unit.

![Figure 9.4 – Illustrated example of first prototype LTCC package](image)

Early progress has demonstrated the structural form to include multiple cavities and through holes can be achieved, but as yet no metallisation has been incorporated. Development work in this area will therefore include via filling, conductive path deposition including inks, pastes and metal deposition techniques. Furthermore investigation is required to establish the best course for connecting the FCoG component to the LTCC. One possibility is to perform a second thermocompression flip-chip bond between the glass and LTCC. Additionally development of the bonding model in ANSYS to aid the flip-chip bonding on glass assembly could be easily adapted to model a glass to LTCC bond process. Alternative assembly methods will also be researched.
9.2.4 Further development

There are numerous directions in which longer term development of the optical encoder device may go. These are most likely to be driven by targeted applications and customer demand. In this section some of the more imminent scenarios are discussed.

After the next phase of reference mark design is complete, optimisation will be undertaken. The reference mark layouts have received much attention through a number of iterations while the most appropriate scheme is determined. However, other than the identified opportunity to increase the yaw tolerance of chips by deployment of a squiggled analyser grating, the development of the incremental layout has stagnated at the current chevron design. Although it is clear that the chevron design offers advantages the actual layout and angle of the chevrons were brought about more through design convenience than by analytical study. To address this, future research will focus on optimising the chevron layout by investigating alternative configurations such as different or varying angles and geometries while sticking to the same principle.

Evolution of the chip is also envisaged with further integration likely. Of foremost importance would be the on-chip interconnection of the photo detector channels, which would significantly reduce the chip pin count. This would be closely followed by the reintroduction of an edge emitting optical communication laser scheme for transfer of the signals by optical means and achieving the benefits associated to optical fibre transmission over wire. Additionally, and in particular with the move to GaAs, it is possible that some of the processing electronics will begin to migrate onto the same optical encoder chip further increasing flexibility, functionality and performance.

From a packaging stand point an alternative to flip-chip on glass could see the introduction of through substrate vias for reversed interconnection. As the contacts would then be located on the backside away from the optical path this would allow for redistribution of the contacts over the entire back surface and reducing the length of subsequent interconnections. In addition, this possibility may also aid with the redistribution of the generated heat. Furthermore this opens the door to the possibility of applying chip stacking techniques whereby the optical encoder chips are directly bonded to an ASIC and/ or memory chips to create a compact system in package solution.
Wafer level testing techniques to identify known good die also need to be developed, which can then maximise the yield of packaged devices. Again a cost saving is realised as less packaging materials go to waste.

9.3 Concluding statements

The work presented in this thesis represents a significant advancement in optical encoder technology. By employing microsystems technologies to monolithically integrate the key encoder components onto a single semiconductor chip an optical encoder readhead of an order of magnitude smaller and cheaper will be produced. The application of novel layouts and reference marker configurations will ensure the miniaturisation is matched by improvements in performance. These benefits will mean that the optical encoders can become an important technology in previously unsuitable applications and markets, while creating a paradigm shift in the capabilities of existing applications and fields.
Appendices

Appendix A

Edge emitting communications laser

An advanced scheme proposed for signal transfer of the processed encoder signals involves the use of optical communications via an edge emitting laser that has also been fabricated on the optical encoder chip. Figure A1 provides an illustrated example of the scheme. In this instance, the signals are passed first to the processing electronic and then back to the optical encoder chip and the edge emitting laser for optical transmission, using time multiplexed methods, to a downstream evaluation unit via the optical fibre.

Figure A1 – Illustrated example of bonded encoder chip solution having an edge emitting optical communication laser

Figure A2 – Optical encoder chip with edge emitting laser
Figure A2 presents an example of a prototype optical encoder chip with an edge emitting laser fabricated on the same chip. Lasers with various cavity lengths are examined. Figure A3 shows the output power versus DC drive current for as-cleaved lasers with cavity lengths of 500μm, 1000μm and 1500μm. As expected, the threshold current increases and the slope efficiency decreases as the cavity length increases. The output spectrum in figure A4 shows a very typical narrow spectral linewidth of around 1.5nm with clear Fabry-Perot peaks resulting from the cavity resonance. These results indicate that the epi-material is of very good quality as the performance is amongst the best reported for InGaAsP based ridge waveguide laser devices operating around 1300nm [8.1-8.3].
The beam divergence for a laser with 500µm cavity length operating at 50mA in room temperature is shown in figure A5. A maximum divergence angle of 34.2° is measured in the vertical plane, which is close to the 30° expected.

Finally the laser is subjected to a continual life time test running at 100mA for over 500 hours operating in room temperature (20°C), as shown in figure A6. After a slight initial reduction in output power over the burn in phase, the results show good stability with steady outputs of around 13mW.

![Figure A5 – Beam divergence](image1)

![Figure A6 – Lifetime reliability](image2)
Appendix B
Substrate emitting optical encoder

InP has been chosen as the semiconductor substrate for the manufacturing of the optical encoder chips because this substrate is transparent at the operating wavelength of 1300nm. This property allows for an alternative design proposal whereby the LED emission is directed through the substrate material in a substrate emitting optical encoder regime as illustrated in figure B1.

As emission and detection take place through the backside of the chip, the optical gratings are fabricated on the opposite side to the electrical interconnects. This means that the contact pads can be distributed over the entire top surface and allows bonding to be performed without obscuring the optical path.

A reordering of the surface emitting epitaxy deposition, as detailed in table B1, onto the InP substrate relocates the position of the light emitting and photo-sensitive regions such that the LED now resides below the photo-sensitive elements. A more detailed illustration is provided in figure B2. The structure comprises two polarised n-type and p-type regions sandwiching laminated active areas to form a junction surface for light emission and photo-detection. The n-type and p-type layers are electronically connected by metal film electrodes formed from gold deposition. An isolation layer is deposited to insure appropriate contact to only the required locations on the n-type and p-type layers.

![Illustration of substrate emitting optical encoder](image-url)
The light emitting and photo sensitive regions of the optical chip are located at different heights within the epitaxy structure (approximately 4µm difference). However for effective flip chip bonding to the submount, a planarised set of bond pads are required. This is achieved by precision controlled electroforming of a gold contact post onto the p-type contact electrode on top of the light emitting regions. Contact posts formed on top of each of the light emitting elements also act as a heat sink to that component and provide a large contact surface area for flip chip bonding.

<table>
<thead>
<tr>
<th>Layer</th>
<th>t (nm)</th>
<th>Strain (ppm)</th>
<th>PL (nm)</th>
<th>Doping (cm⁻³)</th>
</tr>
</thead>
<tbody>
<tr>
<td>28 Detector n-contact InP</td>
<td>400</td>
<td>n</td>
<td>4.00E+18</td>
<td></td>
</tr>
<tr>
<td>27 Detector absorber InGaAs</td>
<td>2250</td>
<td>0</td>
<td>un</td>
<td></td>
</tr>
<tr>
<td>25 InGaAs</td>
<td>250</td>
<td>p</td>
<td>5.00E+18</td>
<td></td>
</tr>
<tr>
<td>24 InP</td>
<td>400</td>
<td>p</td>
<td>1.50E+18</td>
<td></td>
</tr>
<tr>
<td>23 LED/Detector p-contact</td>
<td>InGaAs</td>
<td>150</td>
<td>p</td>
<td>&gt;1e19</td>
</tr>
<tr>
<td>22 Q1.3</td>
<td>20</td>
<td>0</td>
<td>1.00E+18</td>
<td></td>
</tr>
<tr>
<td>21 p-cladding upper InP</td>
<td>750</td>
<td>p</td>
<td>7.00E+17</td>
<td></td>
</tr>
<tr>
<td>20 p-cladding lower InP</td>
<td>750</td>
<td>p</td>
<td>6.00E+17</td>
<td></td>
</tr>
<tr>
<td>19 Q1.2</td>
<td>20</td>
<td>0</td>
<td>6.00E+17</td>
<td></td>
</tr>
<tr>
<td>18 InP Spacer</td>
<td>150</td>
<td>p</td>
<td>6.00E+17</td>
<td></td>
</tr>
<tr>
<td>17 LED p-SCH Q1.06</td>
<td>30</td>
<td>0</td>
<td>6.00E+17</td>
<td></td>
</tr>
<tr>
<td>16 LED p-SCH Q1.06</td>
<td>30</td>
<td>0</td>
<td>0.00E+00</td>
<td></td>
</tr>
<tr>
<td>15 barrier InGaAsP</td>
<td>10</td>
<td>-1500</td>
<td>1060</td>
<td>---</td>
</tr>
<tr>
<td>14 QW6 InGaAsP</td>
<td>6</td>
<td>7800</td>
<td>---</td>
<td>0</td>
</tr>
<tr>
<td>13 barrier InGaAsP</td>
<td>10</td>
<td>-1500</td>
<td>1060</td>
<td>---</td>
</tr>
<tr>
<td>12 QW5 InGaAsP</td>
<td>6</td>
<td>7800</td>
<td>---</td>
<td>0</td>
</tr>
<tr>
<td>11 barrier InGaAsP</td>
<td>10</td>
<td>-1500</td>
<td>1060</td>
<td>---</td>
</tr>
<tr>
<td>10 QW4 InGaAsP</td>
<td>6</td>
<td>7800</td>
<td>---</td>
<td>0</td>
</tr>
<tr>
<td>9 barrier InGaAsP</td>
<td>10</td>
<td>-1500</td>
<td>1060</td>
<td>---</td>
</tr>
<tr>
<td>8 QW3 InGaAsP</td>
<td>6</td>
<td>7800</td>
<td>---</td>
<td>0</td>
</tr>
<tr>
<td>7 barrier InGaAsP</td>
<td>10</td>
<td>-1500</td>
<td>1060</td>
<td>---</td>
</tr>
<tr>
<td>6 QW2 InGaAsP</td>
<td>6</td>
<td>7800</td>
<td>---</td>
<td>0</td>
</tr>
<tr>
<td>5 barrier InGaAsP</td>
<td>10</td>
<td>-1500</td>
<td>1060</td>
<td>---</td>
</tr>
<tr>
<td>4 QW1 InGaAsP</td>
<td>6</td>
<td>7800</td>
<td>---</td>
<td>0</td>
</tr>
<tr>
<td>3 barrier InGaAsP</td>
<td>10</td>
<td>-1500</td>
<td>1060</td>
<td>---</td>
</tr>
<tr>
<td>2 LED n-SCH InGaAsP</td>
<td>60</td>
<td>0</td>
<td>1060</td>
<td>---</td>
</tr>
<tr>
<td>1 LED n-contact Layer InP</td>
<td>1000</td>
<td>---</td>
<td>---</td>
<td>n 4.00E+18</td>
</tr>
<tr>
<td>0 Substrate InP</td>
<td>---</td>
<td>---</td>
<td>Semi-insulating</td>
<td></td>
</tr>
</tbody>
</table>

| Total epi thickness          | 6566 nm|
| Target Pl Wavelength of QW emission: | 1295 nm |

Table B1 – Substrate emitting epitaxy order
Figure B2 – Detailed illustration of substrate emitting optical encoder chip construction
Appendix C

Effect of substrate thickness

The current encoder designs, such as those of Chapter 2.3.1, have a separation between the detectors and the analyser grating which depends on how well the glass graticule can be bonded to the detector chip. As these detectors are absorbing from the surface this separation can be very small and the coupling efficiency can be optimised, with the application of index matching gel. In the new integrated design the light must propagate through the substrate increasing thereby the distance between grating and detector. As there are many possible paths from the LED to each photodiode and because of the non-zero thickness of the substrate, two rays incident on the same point on the analyser grating may illuminate different points on the photodiode array, an effect known as crosstalk, as illustrated in figures C1 and C3. Therefore sufficient gaps defined by these dead zones need to be created between photodiodes to avoid optical crosstalk.

![Figure C1](image)

Figure C1 – Optical crosstalk due to many paths

Figure C2 presents the variation of the stand off distance between the detectors and the scale for various substrate thicknesses as a function of the length of the dead zone in order to minimise crosstalk. From this graph, thinner substrates and larger stand off distances reduce the required length of the dead zone areas. A thinner substrate is also demonstrated to allow for a smaller stand off distance, which is desirable.
Figure C2 – Difference in position on the photodiode array for the marginal rays from the LED illuminating the same point on the analyser Grating for several substrate thicknesses

This point is further emphasised by considering the diffraction at each aperture of the light incident on the analyser grating. Figure C3 again illustrates how a thicker substrate causes the light intended for photodiode A to spread over onto photodiode B. The distance between neighbouring photodiodes should also be calculated to minimise the crosstalk. A thinner substrate would narrow the spread of light and therefore minimise the distance between photodiodes as demonstrated in figure C4.
As a result an additional lapping and polishing step is introduced before deposition of the metal for index and analyser grating fabrication. This reduces the wafer substrate thickness to approximately 200µm - 250µm.

Figure C3 – Crosstalk originated from the diffraction of light at analyser grating

Figure C4 – Signal crosstalk as a function of wafer thickness and photodiode separation
Appendix D

Functional test of flip-chip bonded substrate emitting devices

The primary objective of the functional device test is to verify the operation of the monolithically integrated LED and PD, and assess the performance of the Au-Au flip-chip bonded encoder.

D1 Electrical measurements

First the LED output power is measured. This is performed by positioning a photo detector directly opposite the light source emission path. Figure D1 shows the measured output power for a drive current up to 75mA. A flaw in the substrate design limited the input stimulus as, at higher currents, electromigration effects caused an open circuit in the supply track.

![Figure D1 – Drive current versus output power for functional LED test](image)

Electrical testing of the photo diode arrays, A, B, C, and D, are tested individually. Using an external light source the detectors are illuminated directly and the voltage recorded after the pre-amp stage. Figure D2 shows an example from one chip where a change in voltage of around 20mV is observed as the light source detector stand-off distance is increased. There is a discrepancy between photo diode arrays PD B, PD D and arrays PD A and PD C. Under ideal conditions, the measured signals for each array would be identical. Possible reasons for the discrepancy include a misalignment of the independent light source against the detector array, failed photo detectors on the chip, or
failed flip-chip interconnects to the respective diodes. The result could be verified by inspection of a Lissajous curve to determine phase and amplitude of the coupled signals.

![Graph showing voltage change in photo diode array upon illumination from external light source at increasing light source detector stand-off distance.](image)

**Figure D2 – Voltage change in photo diode array upon illumination from external light source at increasing light source detector stand-off distance**

**D2 Optical measurements**

As both the LED and PD appear to respond well independently, a functional test of the optical encoder against a piece of reflective Cr on glass graduation scale is performed. For an LED drive current of 70mA a photocurrent of around 200nA would not be unreasonable. However figure D3 shows the results for a pixel detector arrangement and a parallel detector arrangement to be 100 times and 30 times greater than this expected value, respectively.

Saturation of the signal in such a manner could be attributed to an electrical fault on the chip design but, as the LED and PD function well independently, this is unlikely. A plausible explanation is an excessive quantity of light is reaching the photo detectors causing saturation. This is most likely to be as a result of larger than anticipated optical leakage within the InP substrate. Figure D4 illustrates some potential stray light reflection paths that would contribute to an increased number of incident photons on the detectors. Such effects had been considered at the concept stage but it was thought that the vast majority of light from the LED would be focused in the desired direction and thus would not be a problem.
The results indicate that, with the existing optical layout, the substrate emitting optical encoder cannot function properly. It was also not possible to prove the concept of the novel detector layouts by setting up the LED of one chip against the detectors of another, in optical transmission fashion, because of the impracticalities of aligning the 2µm features of the index, scale and analyser gratings.

Potential solutions to this problem such as light scattering techniques and optical barriers have been discussed but the increased risk and cost of such development result in a substrate emitting design being aborted in favour of perusing the substantially less risky surface emitting design.
Appendix E

Au capped Cu bumps

The Cu bumps are an attractive alternative to pure Au bumps because of the electrical and thermal properties of Cu and also by forming the core of the bump (approx. 75% Vol.) substantial material cost saving can be made. However these are slightly offset against the cost of running and maintaining three electroplating facilities as opposed to one.

Figure E1 – Cu core bump fabrication process

Figure E2 – Illustration of Cu core bump structure
The process flow shown in figure E1 for the fabrication of Cu core bumps is similar to that of the pure Au bumps process with the addition of further electrodeposition steps for the Au cap that acts as both prevention against oxidation and facilitates Au-Au bonding. In addition a thin Ni intermetallic layer is deposited between the Cu and Au to act as a diffusion barrier to prevent migration of Au into the Cu, which would cause reliability issues over time, such that a Cu/Ni/Au bump structure is formed as illustrated in figure E2. For the Cu core bumps a Cu underbump seed layer is also employed. The compositions and optimal parameters for the extra Cu and Ni plating baths are provided in tables E1 and E2.

<table>
<thead>
<tr>
<th>Description</th>
<th>Unit</th>
<th>Optimum</th>
<th>Vol.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copper sulfate</td>
<td>ml/l</td>
<td>100</td>
<td>2500</td>
</tr>
<tr>
<td>Sulfuric acid</td>
<td>g/l</td>
<td>40</td>
<td>1000</td>
</tr>
<tr>
<td>Electroposit 1200M</td>
<td>ml/l</td>
<td>40</td>
<td>1000</td>
</tr>
<tr>
<td>Electroposit 1300S</td>
<td>ml/l</td>
<td>1</td>
<td>25</td>
</tr>
<tr>
<td>Hydrochloric acid, HCl (S.G. 1.18, 37%)</td>
<td>ml/l</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Cupronal BP carrier</td>
<td>ml/l</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Upronal BP additive</td>
<td>ml/l</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Temperature</td>
<td>°C</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Current Density</td>
<td>A/dm²</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

**Table E1 – Cu electroplating bath composition**

<table>
<thead>
<tr>
<th>Electro-nic 10-03S (Sulphamate)</th>
<th>Unit</th>
<th>optimum</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ni sulphamate</td>
<td>g/l</td>
<td>320</td>
<td>256-385</td>
</tr>
<tr>
<td>Ni metal</td>
<td>g/l</td>
<td>75</td>
<td>60-90</td>
</tr>
<tr>
<td>Boric Acid</td>
<td>g/l</td>
<td>40</td>
<td>30-45</td>
</tr>
<tr>
<td>Lectro-nic Anode Activator</td>
<td>ml/l</td>
<td>65</td>
<td>50-100</td>
</tr>
<tr>
<td>Lectro-Nic Addition Agent</td>
<td>ml/l</td>
<td>12</td>
<td>10-20</td>
</tr>
<tr>
<td>Non-Pitter 62A</td>
<td>ml/l</td>
<td>2</td>
<td>1-2</td>
</tr>
<tr>
<td>Temperature</td>
<td>°C</td>
<td>57</td>
<td>52-60</td>
</tr>
<tr>
<td>PH (electrometric)</td>
<td></td>
<td>3.25</td>
<td>2.5-3.75</td>
</tr>
<tr>
<td>Current Density</td>
<td>A/dm²</td>
<td>4</td>
<td>1.5-8.0</td>
</tr>
<tr>
<td>Lectro-Nic Addition Agent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consumption</td>
<td>ml/ah</td>
<td>1</td>
<td>0.5-2.0</td>
</tr>
</tbody>
</table>

**Table E2 – Ni electroplating bath composition**

Initial plating results demonstrated poor adhesion failing a basic tacky tape test. Two failure modes were identified. Firstly at the Ti to Cu seed layer interface and secondly at the Cu seed layer to Cu deposition interface. Yields are improved to over 95% by first performing a shallow Ti etch to expose fresh material before deposition of the Cu seed layer. Secondly it is known that higher current densities produce smaller grain sizes but can lead to poor adhesion and if too high can cause electrical scorching of the seed.
layer. Lower current densities produce increased grain sizes that adhere well but produce uneven surface finishes and introduce more stress in the resist mould.

Therefore pulsed profiles are introduced at the beginning of the plating cycle to promote adhesion as demonstrated in figure E3. The pulsed section of the plating profile has incremental increases in applied current density of around 4%. The pulse durations are considered in the overall time required to reach the desired bump heights and the total pulse section lasts for around 7% of the overall plating cycle.
Appendix F

Flip-chip bonding of Cu core bumps

The flip-chip bonding of the optical encoder chips is already challenging because of the brittle nature of the InP. The difficulties are further compounded by the need to thin the substrate emitting chip device down to around 200µm-250µm.

To prevent damage to the chip prolonged bonding cycles are required with extended ramping profiles for applied pressure. Bonding durations at peak applied loads of between 40g/bump and 55g/bump were made between 30 seconds and 50 seconds.

Figure F1 shows an unsuccessfully bonded chip displaying a number of failure modes including failure at the Au-Au interface despite the clear indication that intimate contact has been made. Chip Au pad metallization lift-off and notably a piercing of the Au pad through to the underlying semiconductor layer can be seen.

Where chips did remain attached the bonds were very weak achieving an average shear force of only 6g/bump. The poor bonding result is attributed to a combination of the sputtered metal pads on the chip, which is hard by comparison to the electrodeposited gold cap, and also the fact that the Cu core will work harden under applied pressure allowing a ridged bump to pierce through the thin metal pad.
In order to improve the compliance at the bond interface a 1µm thick Au layer is electroplated over the sputtered Au surface metallization on the chip. The number of chips that then remained attached after bonding improved to around 60% but despite an increase, the average bond strength remained poor at 11g/bump. Figure F2 shows a SEM image of detached bumps after the chip is sheared and reveals two failure mechanisms, one at the Au-Au interface but also at the seed layer to glass interface. Significant deformation of the Au cap can also be observed suggesting a high level of contact between bump and pad. However the level of interconnectivity remained poor.

Significant improvements were achieved when bonding was performed using pure Au bumps, even at lower bonding pressures, where average shear strengths increased to 25g/bump and interconnectivity reached 100%. This allowed for the testing of flip-chip bonded substrate emitting devices as discussed before.
## Appendix G

Flip-chip bonding parameters as input into the Karl Suss FC6 flip-chip bonder.

### Table G1 – Karl Suss FC6 thermocompression flip-chip bonding parameters for Batch 1 bumps

<table>
<thead>
<tr>
<th>Sample</th>
<th>Batch 1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table G2 – Karl Suss FC6 thermocompression flip-chip bonding parameters for Batch 2 bumps

<table>
<thead>
<tr>
<th>Sample</th>
<th>Batch 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>