Design and Fabrication of Nanoscale Bismuth Hall Probes

by

Ricky Chu

B. Sc., The University of British Columbia, 2005

Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

in the Department of Physics Faculty of Science

© Ricky Chu 2016

SIMON FRASER UNIVERSITY

Fall 2016

Permission is granted to reproduce this work, in whole or in part, for purposes that are simultaneously educational and non-commercial. Use of original figures from this thesis in a graduate thesis or an academic journal article is allowed if this thesis is referenced and credited. All other rights reserved.
Approval

Name: Ricky Chu
Degree: Doctor of Philosophy (Physics)
Title: Design and Fabrication of Nanoscale Bismuth Hall Probes

Examining Committee: Chair: Eldon Emberly Professor

David M. Broun
Senior Supervisor
Associate Professor

Karen L. Kavanagh
Supervisor
Professor

Simon P. Watkins
Supervisor
Professor

Erol Girt
Internal Examiner
Associate Professor

Joshua A. Folk
External Examiner
Associate Professor
Department of Physics and Astronomy
The University of British Columbia

Date Defended: 9 December 2016
Abstract

Scanning Hall probe microscopy is a quantitative magnetic imaging technique with high magnetic flux sensitivity and high spatial resolution. Hall sensors have untapped potential to match the sensitivity of superconducting quantum interference devices (SQUIDs), which are well-known in magnetic microscopy for their flux sensitivity. Furthermore, Hall probes can do so with better spatial resolution. My thesis supports this conclusion with a theoretical calculation while comparing the Hall probe technique to other kinds of magnetic imaging.

I have explored further improvements in the overall design and materials of Hall probes. I have obtained and analyzed magnetotransport data for various concentrations of lead in bismuth films and Hall probes. Bismuth, a compensated metal, is a good alternative to semiconductor Hall probes. The presence of electron and hole carriers, though, reduces the Hall effect, and bismuth would be even better for Hall sensors if one kind of carrier were compensated. A doping between 0 and 0.1% lead in bismuth appears to be best for lead-doped bismuth Hall probe operation.

I have also made significant progress in the design and fabrication of a more durable Hall probe shape, inspired by hard drive read heads. The novel design should enable operation closer to the sample surface, improving spatial resolution and making it easier to detect flux.

**Keywords:** Hall effect; sensor; probe; microscopy; bismuth; doped
Acknowledgements

Behind most achievements, great and small (in my case, humble), there are support people who could be taking the lead in a project, but choose auxiliary roles instead. Modern science would not advance very quickly without support staff, but they rarely get their names on papers or posters, whether their contributions are direct or indirect, major or minor. This is for those who toil in the shadows so that scientists’ not-so-grand schemes can be seen by a small collection of esoterically-educated eyeballs around the world. This is for those who share their sunshine (or thunderclouds) and wit as they work just as hard on projects that may affect a few thousand people or just a few people. Thank you.

I have probably omitted a few people by accident since I have been here so long, so apologies if I have not mentioned you.

Thanks to the physics office staff: Rose Evans, Andrei Badelescu, Bianca Beaulieu, Jen Chang, Misty Jorgensen, Amy Lau, Dagni Lund, Candida Mazza, Sada Rangnekar, Susan Roy, Ayako Nagasawa, Sukhy Sidhu, and Margaret van Soest.

Thanks to the physics technical staff: Vic Allen, Bryan Gormann, Ken Myrtle, Mehrdad Rastan, David Lee, Jeff Rudd, and Laura Schmidt.

Thanks to the 4D LABS staff: Li Yang, Tom Cherng, Liang Hong, Grace Li, Nathanael Sieb, Chris Balicki, Yi Xu, and Mary Boysel.

Thanks to the Science Technical Centre staff: Pawel Kowalski, Jim Shoults, Ken Van Wieren, and Howard Proulx.

Thanks to Paul Mulyk from chemistry. Thanks to Finlay MacNab and Xin Zhang for their cleanroom advice, or, in the case of Finlay, valuable advice laced with acid wit. Thanks to Jamal Bahari for assistance with dicing.

Thanks to my parents, who have been patiently waiting for me to finish all this, and have been quietly supportive all this time.

Thanks to the Broun Lab for help, support, good humour, and countless stimulating conversations: Nigel David, Wendell Huttema, Patrick Turner, Adam Schneider,
Taras Chouinard, Xiaoqing Zhou, Andisheh Beiki, Sonia Milbradt, Colin Truncik, Natalie Murphy, Kevin Morse, Megan Boothby, Nicholas Lee-Hone, and Lauriane Gérémie. I learned a lot talking to smart people about so many random things: physics (of course), politics, investing, house prices, and general mathematical models of anything. I still miss the hallway soccer games.

Thanks to my thesis examining committee for their thoughtful contributions: Karen Kavanagh, Simon Watkins, Erol Girt, and Joshua Folk.

Of course, the greatest thanks go to my long-suffering senior supervisor, David Broun, who has put in so much time, effort, and careful thought into helping me learn and ultimately complete my PhD. What a world it would be if it were full of people as fiercely intelligent, thoughtful, hard-working, compassionate, understanding, and helpful as him. It amazes me how he stayed relentlessly positive and supportive in the face of some of the challenges I faced in my project.
Table of Contents

Approval ii
Abstract iii
Acknowledgements iv
Table of Contents vi
List of Tables xii
List of Figures xiii

1 Introduction to Hall Nanoprobe Microscopy 1
  1.1 Thesis Overview ........................................... 3
  1.2 Applications ............................................. 5
    1.2.1 Vortex Imaging and Type II Superconductivity .......... 6
    1.2.2 Time-Resolved Hall Probe Technique .................. 6
    1.2.3 Bioassays Using Magnetic Beads ......................... 7
    1.2.4 In-Situ Nanofabrication Using Photodoping ............ 8
    1.2.5 Magnetic Thin Films .................................... 8

2 Survey of Magnetic Imaging Techniques 10
  2.1 Desirable Traits in Magnetic Microscopy .................. 10
  2.2 Magnetic Flux Detection Mechanisms ....................... 11
  2.3 Magnetic Flux Resolution .................................. 12
  2.4 Spatial Resolution ....................................... 13
  2.5 Imaging Speed .......................................... 14
  2.6 Operational Temperature Range ........................... 15
  2.7 Non-Invasive Measurement ................................ 15
  2.8 High Sensor Durability ................................... 15
2.9 Easy Interpretability of Data ........................................ 16
2.10 A Summary of the Comparisons ................................. 16

3 Flux Sensitivity in Detail ........................................... 18
  3.1 The Generic Experiment ........................................ 19
  3.2 Noise .......................................................... 20
    3.2.1 Noise Distributions ................................. 20
    3.2.2 Thermal Electronic Noise ......................... 21
    3.2.3 Thermal Vibrational Noise ....................... 22
    3.2.4 Telegraph Noise ................................. 22
  3.3 A Fair Flux Sensitivity Metric .................................. 23
  3.4 Flux Sensitivity of SQUIDs .................................... 25
  3.5 Magnetic Force Microscopy Flux Sensitivity .................... 29
    3.5.1 Magnetic Interactions ........................... 29
    3.5.2 Fundamental Limits ........................... 30
  3.6 Hall Probe Flux Sensitivity: Basics ............................ 33
    3.6.1 The Hall Effect .................................. 33
    3.6.2 Noise in Hall Probes .......................... 35
    3.6.3 Length Scales .................................. 37
    3.6.4 Fundamental Limits .......................... 38
  3.7 Hall Probe Flux Sensitivity: Bismuth .......................... 41
    3.7.1 Why Bismuth? .................................. 42
    3.7.2 Heating .......................................... 43
    3.7.3 Metallic Limit ................................ 45
    3.7.4 Semimetal Probes ............................... 46
  3.8 A Fundamental Link? ........................................ 47

4 Scanning Probe Microscopy ....................................... 49
  4.1 Scanning Hall Probe Microscope ............................... 50
    4.1.1 Physical Design ................................ 50
    4.1.2 LabVIEW Software .............................. 54
  4.2 Spatial Resolution ........................................... 55
    4.2.1 Definition ....................................... 55
    4.2.2 Angle and Height Considerations ............... 58

5 Hall Probe Design ............................................. 61
  5.1 Conventional Probes ......................................... 62
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.2</td>
<td>The Novel Design</td>
<td>65</td>
</tr>
<tr>
<td>5.3</td>
<td>Simulations</td>
<td>68</td>
</tr>
<tr>
<td>5.3.1</td>
<td>Hall Probe Differential Equation</td>
<td>69</td>
</tr>
<tr>
<td>5.3.2</td>
<td>Simulation Details</td>
<td>70</td>
</tr>
<tr>
<td>5.3.3</td>
<td>Conventional Designs</td>
<td>72</td>
</tr>
<tr>
<td>5.3.4</td>
<td>Novel Designs</td>
<td>73</td>
</tr>
<tr>
<td>5.3.5</td>
<td>Novel Design with Fabrication Flaws</td>
<td>73</td>
</tr>
<tr>
<td>5.4</td>
<td>Thickness and Hall Probe Sensitivity</td>
<td>75</td>
</tr>
<tr>
<td>6</td>
<td>Hall Probe Fabrication</td>
<td>78</td>
</tr>
<tr>
<td>6.1</td>
<td>Conventional Probes</td>
<td>78</td>
</tr>
<tr>
<td>6.1.1</td>
<td>EBL Resolution Tests</td>
<td>80</td>
</tr>
<tr>
<td>6.1.2</td>
<td>EBL Writing Problems</td>
<td>80</td>
</tr>
<tr>
<td>6.2</td>
<td>Novel Probe Fabrication: A Blueprint</td>
<td>82</td>
</tr>
<tr>
<td>6.2.1</td>
<td>Bottom Gold Pads</td>
<td>83</td>
</tr>
<tr>
<td>6.2.2</td>
<td>Trilayer</td>
<td>86</td>
</tr>
<tr>
<td>6.2.3</td>
<td>Top Gold Pads and Dicing</td>
<td>87</td>
</tr>
<tr>
<td>6.2.4</td>
<td>Active Region</td>
<td>88</td>
</tr>
<tr>
<td>6.2.5</td>
<td>Finishing</td>
<td>90</td>
</tr>
<tr>
<td>6.3</td>
<td>Fabrication Results</td>
<td>90</td>
</tr>
<tr>
<td>6.3.1</td>
<td>Bottom Gold Pads</td>
<td>91</td>
</tr>
<tr>
<td>6.3.2</td>
<td>Trilayer</td>
<td>91</td>
</tr>
<tr>
<td>6.3.3</td>
<td>Top Gold Pads and Dicing</td>
<td>96</td>
</tr>
<tr>
<td>6.3.4</td>
<td>Active Region</td>
<td>98</td>
</tr>
<tr>
<td>6.4</td>
<td>Process Improvements</td>
<td>105</td>
</tr>
<tr>
<td>7</td>
<td>Doped Bismuth as a Hall Probe Material</td>
<td>108</td>
</tr>
<tr>
<td>7.1</td>
<td>Bismuth: Basic Properties</td>
<td>109</td>
</tr>
<tr>
<td>7.2</td>
<td>One Band Is Best</td>
<td>112</td>
</tr>
<tr>
<td>7.3</td>
<td>Optimal Doping and Bismuth Band Structure</td>
<td>114</td>
</tr>
<tr>
<td>7.4</td>
<td>A Three-Band Model of Bismuth</td>
<td>115</td>
</tr>
<tr>
<td>7.4.1</td>
<td>Crystallite Size and Orientation</td>
<td>117</td>
</tr>
<tr>
<td>7.4.2</td>
<td>Carrier Concentrations</td>
<td>119</td>
</tr>
<tr>
<td>7.4.3</td>
<td>Mobility and Scattering</td>
<td>123</td>
</tr>
<tr>
<td>7.5</td>
<td>Doping Methods</td>
<td>124</td>
</tr>
<tr>
<td>7.5.1</td>
<td>Melting in Sealed Glass Container</td>
<td>124</td>
</tr>
</tbody>
</table>
# Appendix B  Continuous Flow Cryostat  
B.1  Overview ................................................................. 177
  B.1.1  Physical Design ................................................. 178
  B.1.2  LabVIEW Software ............................................ 179
  B.1.3  Operational Notes ............................................. 180

# Appendix C  Metal Evaporation  
C.1  Evaporator System .................................................. 182
  C.2  Tooling Factor ..................................................... 183

# Appendix D  Hall Probe Microscope  

# Appendix E  Spatial Resolution Simulations  

# Appendix F  Various Points of Technique  
F.1  Profilometry .......................................................... 194
  F.2  The Delicate Art of Cleaving Silicon .............................. 194
    F.2.1  (100) Si ......................................................... 195
    F.2.2  Tools ........................................................... 195
    F.2.3  Scratch, Nick, and Snap .................................... 195
    F.2.4  Nick ............................................................ 196
    F.2.5  Score and Snap ............................................... 196
    F.2.6  Rough Cleaves ............................................... 197
  F.3  Piezo Cutting ....................................................... 197
  F.4  Sample Handling .................................................. 198
  F.5  Indium Contacts .................................................. 198
  F.6  Static Electricity .................................................. 199

# Appendix G  Hall Probe Production Recipes  
G.1  Conventional Hall Probe Fabrication .............................. 200
  G.1.1  Spin-Coating Resist onto Wafer ............................... 200
  G.1.2  Spin-Coating Resist onto Chip ............................... 201
  G.1.3  E-Beam Lithography ........................................... 202
  G.1.4  Bismuth and Germanium Evaporation .......................... 203
  G.1.5  Lift-Off ......................................................... 204
  G.1.6  Bismuth Doping ................................................. 205
  G.1.7  Atomic Absorption Spectroscopy ................................ 206
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>G.2</td>
<td>Novel Probe Production</td>
<td>206</td>
</tr>
<tr>
<td>G.2.1</td>
<td>Bottom Gold Pads</td>
<td>206</td>
</tr>
<tr>
<td>G.2.2</td>
<td>Trilayer Deposition</td>
<td>209</td>
</tr>
<tr>
<td>G.2.3</td>
<td>Trilayer Resist Mask</td>
<td>210</td>
</tr>
<tr>
<td>G.2.4</td>
<td>Trilayer Etching</td>
<td>210</td>
</tr>
<tr>
<td>G.2.5</td>
<td>Top Gold Pads</td>
<td>212</td>
</tr>
<tr>
<td>G.2.6</td>
<td>Dicing</td>
<td>214</td>
</tr>
<tr>
<td>G.2.7</td>
<td>Active Region Fabrication</td>
<td>214</td>
</tr>
<tr>
<td>G.2.8</td>
<td>Finishing</td>
<td>215</td>
</tr>
<tr>
<td>G.3</td>
<td>Carrier Chip Fabrication Recipe</td>
<td>216</td>
</tr>
<tr>
<td>H.1</td>
<td>Mac Installation</td>
<td>217</td>
</tr>
<tr>
<td>H.2</td>
<td>Program Modifications</td>
<td>218</td>
</tr>
<tr>
<td>I.1</td>
<td>Additional References</td>
<td>219</td>
</tr>
</tbody>
</table>
List of Tables

Table 2.1 Figures of merit comparison for magnetic imaging techniques. . 16
Table 5.1 Hardness of bismuth and germanium. . . . . . . . . . . . . . . . 65
Table 5.2 Table of parameters used in finite element simulations. . . . 71
Table 5.3 Hall voltage simulation results. Larger Hall voltage is better. . 72
Table 6.1 Sputtering settings for depositing the trilayer. . . . . . . . . . 92
Table 6.2 Si 2p electron binding energies. . . . . . . . . . . . . . . . . . . 92
Table 6.3 RIE etch test parameters. . . . . . . . . . . . . . . . . . . . . 95
Table 6.4 Expected vs. actual EDX signals for miling trilayer. . . . . . 101
Table 6.5 Milled trench sizes. . . . . . . . . . . . . . . . . . . . . . . . . 104
Table 7.1 Vapour pressure of pure bismuth and lead vs. temperature. . . 112
Table 7.2 X-ray diffraction line-broadening parameters. . . . . . . . . . 119
Table 7.3 Selected bismuth band and transport parameters. . . . . . . 121
Table 7.4 Bismuth carrier concentrations vs. band-specific Fermi temperature. 124
Table 7.5 Estimates of acceptor concentration in doped bismuth films. . 134
Table 7.6 Parameters used in analysis of doped film magnetotransport data. 146
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>$I, B, V_H$ vs. $t$ for time-resolved Hall sensor.</td>
<td>7</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>A generic experiment.</td>
<td>19</td>
</tr>
<tr>
<td>Figure 3.2</td>
<td>Telegraph noise.</td>
<td>23</td>
</tr>
<tr>
<td>Figure 3.3</td>
<td>DC SQUID schematic.</td>
<td>25</td>
</tr>
<tr>
<td>Figure 3.4</td>
<td>Periodicity of SQUID critical current with flux.</td>
<td>26</td>
</tr>
<tr>
<td>Figure 3.5</td>
<td>$V$ vs. $I$ curves for a SQUID.</td>
<td>28</td>
</tr>
<tr>
<td>Figure 3.6</td>
<td>$V$ vs. $\Phi$ curves for a SQUID.</td>
<td>28</td>
</tr>
<tr>
<td>Figure 3.7</td>
<td>A basic MFM experimental setup.</td>
<td>29</td>
</tr>
<tr>
<td>Figure 3.8</td>
<td>MFM tip dimensions.</td>
<td>29</td>
</tr>
<tr>
<td>Figure 3.9</td>
<td>A Hall probe with dimensions $s \times s \times d$.</td>
<td>35</td>
</tr>
<tr>
<td>Figure 3.10</td>
<td>Ballistic scattering at rounded corners is chaotic.</td>
<td>41</td>
</tr>
<tr>
<td>Figure 3.11</td>
<td>Heat flow for a Hall probe active region.</td>
<td>43</td>
</tr>
<tr>
<td>Figure 4.1</td>
<td>The Broun Lab’s scanning Hall probe microscope.</td>
<td>51</td>
</tr>
<tr>
<td>Figure 4.2</td>
<td>XY-motion piezo bender details.</td>
<td>52</td>
</tr>
<tr>
<td>Figure 4.3</td>
<td>Z-motion piezo bender details.</td>
<td>52</td>
</tr>
<tr>
<td>Figure 4.4</td>
<td>Photographs of the Broun Lab’s Hall probe microscope.</td>
<td>53</td>
</tr>
<tr>
<td>Figure 4.5</td>
<td>Hall probe carrier chip.</td>
<td>53</td>
</tr>
<tr>
<td>Figure 4.6</td>
<td>Model system for determining magnetic probe spatial resolution.</td>
<td>56</td>
</tr>
<tr>
<td>Figure 4.7</td>
<td>Hall probe flux detection simulation.</td>
<td>57</td>
</tr>
<tr>
<td>Figure 4.8</td>
<td>Definition of spatial resolution criterion.</td>
<td>57</td>
</tr>
<tr>
<td>Figure 4.9</td>
<td>Hall probe resolution with $\phi/\sqrt{2}$ criterion.</td>
<td>59</td>
</tr>
<tr>
<td>Figure 4.10</td>
<td>Hall probe resolution with $\phi/\sqrt{1.3}$ criterion.</td>
<td>59</td>
</tr>
<tr>
<td>Figure 4.11</td>
<td>Illustration of probe tilt geometry.</td>
<td>60</td>
</tr>
<tr>
<td>Figure 4.12</td>
<td>Graph of Hall probe tilt effects on spatial resolution.</td>
<td>60</td>
</tr>
<tr>
<td>Figure 5.1</td>
<td>Damage to germanium-coated Hall probe after scanning.</td>
<td>62</td>
</tr>
<tr>
<td>Figure 5.2</td>
<td>Hall probe design.</td>
<td>63</td>
</tr>
<tr>
<td>Figure 5.3</td>
<td>Hall sensor range of motion.</td>
<td>64</td>
</tr>
</tbody>
</table>
Figure 7.6  SEM images of crystallites in bismuth thin films. . . . . . . 120
Figure 7.7  Numbering of contacts for van der Pauw measurements. . . 129
Figure 7.8  Bismuth film resistivities and Hall coefficients at various dopings. . . . . . . . 131
Figure 7.9  Bismuth film conductivities at various dopings. . . . . . . . . 133
Figure 7.10 Conductivity data for 4% nominally-doped films. . . . . . . 137
Figure 7.11 Mobilities from 4% data inversion. . . . . . . . . . . . . . . . 137
Figure 7.12 Conductivity data for 1% nominally-doped films. . . . . . . 139
Figure 7.13 Mobility curves from 1% data inversion. . . . . . . . . . . . . 139
Figure 7.14 $\sigma_L$ and $\sigma_H$ graphs for undoped bismuth films. . . . . . . . . . . . . . 141
Figure 7.15 One-band $p_{\text{eff}}$ graphs for bismuth films. . . . . . . . . . . . . . 142
Figure 7.16 $\sigma_L$ and $\sigma_H$ graphs for 0.1% nominally-doped bismuth films. . . . . 143
Figure 7.17 $\sigma_L$ and $\sigma_H$ graphs for 0.5% nominally-doped bismuth films. . . 145

Figure 8.1  SEM micrographs of Hall probes. . . . . . . . . . . . . . . . 149
Figure 8.2  Magnetotransport graphs: undoped film and probe. . . . . . 152
Figure 8.3  Magnetotransport graphs: 0.1% doped film and probe. . . . . 153
Figure 8.4  Magnetotransport graphs: undoped films with, without Ge. . 155
Figure 8.5  Magnetotransport graphs: 0.1% doped films with, without Ge. . 156
Figure 8.6  Magnetotransport graphs: undoped probes with, without Ge. . 158
Figure 8.7  Magnetotransport graphs: 0.1% doped probes with, without Ge.159
Figure 8.8  Hall probe scan of worm domains in iron garnet film. . . . . . . 160
Figure 8.9  MOKE image of iron garnet film’s worm domains. . . . . . . 160
Chapter 1

Introduction to Hall Nanoprobe Microscopy

“I would like to describe a field in which little has been done, but in which an enormous amount can be done in principle. This field is not quite the same as the others in that it will not tell us much of fundamental physics (in the sense of, “What are the strange particles?”) but it is more like solid-state physics in the sense that it might tell us much of great interest about the strange phenomena that occur in complex situations. Furthermore, a point that is most important is that it would have an enormous number of technical applications.

What I want to talk about is the problem of manipulating and controlling things on a small scale.”

— Richard Philips Feynman, There’s Plenty of Room at the Bottom [1].

We have reached what can be called a “Nano Age,” when everyday items rely on nanoscale features and a good understanding of physics at that length scale. It has so permeated the public consciousness that “nano” is now the marketing buzzword of the day, pushing it towards becoming meaningless hype. In a more measured and scientific sense, nanotechnology has promised to miniaturize things while improving the versatility and effectiveness of human inventions. That promise is coming true.

This is nowhere more apparent than in the computer industry, where the number of transistors in an integrated circuit has doubled every two years since the 1960s. However, this breakneck pace is slowing. In late 2014, Intel released processors fabricated using a 14 nm process [2]. This is the last process node to obey Moore’s law,
which was finally broken in 2016 [3]. There will be more miniaturization, to be sure, but at a slower pace. The industry is running into fundamental physical problems at the nanoscale. The feature sizes of the transistors and hard drive bits are becoming so small that quantum effects hamper their further miniaturization. Transistors recently got a temporary reprieve from electron tunnelling problems with the development of new high-\(\kappa\) dielectrics, but one wonders how many more proverbial rabbits researchers can pull out of the hat. Hard disk bits will eventually become so small that they will be vulnerable to random spin-flipping, with no commercially viable solution in sight.

As the computer industry wrestles with fundamental problems in its devices, businesses are finding other uses for nanotechnology. Cosmetic companies are using nanoparticles in their creams to penetrate the skin and deliver nutrients and drugs. Sunscreen manufacturers are using nanoparticles of titanium dioxide in sunscreens to give the creams a white colour that goes away when rubbed into the skin, unlike regular sunscreen, which leaves a white tinge. Sports equipment makers claim to improve the performance of baseball bats and hockey sticks with carbon nanotubes [4–6].

With these myriad applications both established and establishing, academic researchers can help explore the limits of these technologies, develop new ideas, and investigate the health implications of this explosion of nanoparticle-containing products. The nano region has not been easily accessible to many scientists due to the expense of the apparatus to make and probe objects of this size. For instance, photolithography is diffraction-limited to features on the order of the wavelength of light used. While e-beam lithography has been used since the 1960s, early e-beam systems consisted of modified electron microscopes with software written for high accuracy and repeatability, but less so for ease of use. The arrival of the Nano Age has brought a focus of Canadian federal and provincial government funding for nanoscience projects, including the opening of SFU’s $35 million 4D LABS in 2007 and the University of Alberta’s $120 million National Institute for Nanotechnology in 2006.

These facilities are critical to physicists, chemists, engineers, and materials scientists exploring mesoscopic physics, probing the boundary between the macroscopic and quantum scales. There are now readily accessible, user-friendly tools gathered in cleanroom facilities for research on a diverse range of topics: photonics, metamaterials, nanoelectronics, nanosensors, nanoelectromechanical systems, and more. Who knows what weird and wondrous inventions will arise from such work? Mesoscopic science may help us directly observe how classical behaviour emerges from myriad
quantum interactions and, if we’re lucky, settle the fundamental philosophical question of how we should interpret the mathematical formalism of quantum mechanics.

A new generation of Canadian scientists has now been set loose in the nanoworld. This is just the beginning. As Feynman said, there’s plenty of room at the bottom.

1.1 Thesis Overview

While tools of nanophysics have not advanced in the way Feynman imagined — small machines making even smaller machines all the way down to the nanoscale — our top-down fabrication tools have evolved significantly since he gave that talk. We now have much finer resolution in circuits and devices with e-beam lithography widely available, but some fields of study, such as magnetic probe microscopy, have not kept pace. There continues to be plenty of parameter space at the bottom to make improvements.

This chapter offers an extended overview of the thesis content and a look at interesting applications of scanning Hall probe microscopy (SHPM). The main theme of this thesis is improvements to SHPM, a technique for creating a map of the magnetic flux present at the surface of a sample. This work has four basic parts: a review of the field of magnetic imaging techniques comparing them to SHPM, a brief introduction to a scanning Hall probe microscope and theoretical aspects of Hall probes, a more in-depth theory- and data-driven discussion of improved Hall probes made of doped bismuth, and a study of a novel Hall sensor design that should be significantly more robust than typical Hall devices.

The first two chapters are introductory, with an overview of the thesis content and a survey of magnetic imaging techniques, in order to put SHPM in the context of the field. I will be discussing SHPM, superconducting quantum interference device (SQUID) microscopy, magnetic force microscopy (MFM), and magneto-optical (MO) imaging. In addition to explaining how these techniques work, I will identify the main obstacles to them being more sensitive to magnetic flux and having better spatial resolution.

According to Bending [7], Hall probes and MFM tips have flux sensitivities of $10^{-6}$ and $10^{-4} \Phi_0$, respectively. Calculations of maximum flux sensitivity for Hall probes and MFM tips indicate they can be as sensitive as SQUIDs, which set the benchmark for flux sensitivity to $10^{-7} \Phi_0/\sqrt{\text{Hz}}$. This hints at a universal limit for flux sensitivity,
which may have a fundamental physical basis. At least it is a good figure of merit for evaluating the performance of magnetic sensors.

Spatial resolution is strongly dependent on sensor size. SQUIDs are limited to pickup loops of a few hundred nanometres while Hall probes and MFM tips can be made on the order of tens of nanometres.

Chapter 3 deals with the basic theory of how MFM tips and Hall probes work, with a discussion of noise and fundamental limits of flux sensitivity. If intrinsic thermal noise and electrical heating are the only constraints to improving flux sensitivity, Hall probes have a figure of merit close to published values for SQUIDs.

Also important is that calculation shows we can make Hall probes much smaller than SQUIDs while maintaining the same high flux sensitivity, assuming the Hall effect is not quenched by quantum chaos in the ballistic limit.

The fourth chapter offers a discussion of the Broun Lab’s scanning Hall probe microscope design, with wider ramifications for scanning probe magnetic microscopy in general. The basic design follows the common approach of scanning probe microscopes, using piezoelectric benders providing fine translational motion and a stick-slip positioner system for coarse motion.

The Broun Lab’s design departs from this approach in a few important ways. Of particular interest is the innovative double gimbal rotational stage with piezoelectric stick-slip positioners that allow for a future automated, adaptive rotation of a scanning probe around contours of a sample surface. This is important for avoiding sensor damage and keeping the active region of the probe close to the surface. I will discuss the importance of this and probe alignment for spatial resolution of a scan.

The fifth chapter delves into the Hall probe design work I did, with a focus on a novel design conceived by the Broun Lab. Conventional probes are prone to destruction due to contact with the sample. Avoiding contact typically means staying far above the sample surface, which reduces the spatial resolution of a Hall probe scan. The novel design improves sensor robustness, which helps the sensor survive high resolution scans close to the sample surface.

Instead of an in-plane conventional design, the active region of the novel sensor is changed to an out-of-plane orientation and is made much thicker. These changes make the Hall sensor more robust by allowing us to use the substrate as a protective shield and by making the active region more tolerant to damage. I will present finite element simulations that I performed to test the viability of the design.

Chapter 6 discusses Hall probe fabrication work that I did. The focus here is on the plan I had to make the novel probe which, as a 3-layer device, is somewhat
complicated to make. The processes I used include photolithography, reactive ion etching, and ion beam milling. While I did not make a complete probe, I made significant progress in the fabrication.

Chapter 7 discusses the use of bismuth, a compensated semimetal, as a Hall probe material. Bismuth has been proposed as a replacement for semiconductor Hall probe materials since bismuth may be less prone to excess noise problems. Compared to semiconductors, bismuth has a more complicated three-band structure, so I will review the details of bismuth’s Fermi surface and Hall probe theory for three bands of conducting charge.

Doping of bismuth is necessary to optimize flux sensitivity because the holes and electrons in undoped material have opposing influences on the Hall voltage, creating a zero crossing for the Hall signal at low temperature. I will explain the effects of doping bismuth with lead and discuss magnetotransport data taken on undoped and doped bismuth films.

In Chapter 8, I will briefly discuss magnetotransport data from patterned bismuth Hall probes, with comparisons to unpatterned films and probes with germanium capping layers.

In the final chapter, I conclude by describing some future work that would naturally follow this thesis. For the Hall probe microscope section, that would include software implementation of an algorithm to adaptively adjust the angle of probe to the sample surface. For the bismuth Hall probe work, further sensor miniaturization and optimization of doping would be among the improvements. For the novel sensor work, a completion of the probe fabrication and characterization of the resulting devices would be very interesting.

1.2 Applications

Most of the thesis is devoted to making an improved Hall sensor, but to what end? Before we move on to the fine details of SHPM and magnetic microscopy, let us examine what interesting applications Hall probes have.

Hall nanosensors occupy a unique niche in the field of magnetic imaging, combining high flux sensitivity with high spatial resolution. As such, SHPM is well-positioned to be applied to some long-standing fundamental research problems that have, until now, not been studied in detail because the tools have not been sufficiently refined.
Here, we will look at what further refinements we can make to SHPM and what magnetic systems we can study with Hall probes.

1.2.1 Vortex Imaging and Type II Superconductivity

Unconventional superconductivity is one of the biggest puzzles in physics today, holding out the promise of a drastically different world if the phenomenon can be replicated at room temperature. Superconductors have no DC resistance and expel magnetic field from their interiors using screening supercurrents. A very prominent feature of type II superconductors is the formation of vortices with flux quantum $\frac{h}{2e}$ at high temperature and high magnetic field. The vortices consist of a non-superconducting core of size $\xi$ and a magnetic field penetrating region where the field falls off exponentially with length scale $\lambda$; screening superconducting currents suppress the magnetic field as one moves away from the non-superconducting core. The creation and dynamics of these structures play a central role in determining the properties of superconductors. SHPM has been used to study vortices (see [8], for instance), but with further improvements, SHPM is well-positioned to study vortices in greater detail than ever before.

Hall probes can be used to test models of high temperature superconductivity, such as Ioffe and Larkin’s variant [9] of Anderson’s resonating valence bond model [10]. Using their model, Ioffe and Larkin predicted that spin-charge separation would result in underdoped and highly-doped cuprates, producing an uncharged spin-1/2 fermion and a spin-0 boson with charge $e$. One of the signatures of this fractionalization is a vortex memory effect involving gauge flux trapped in superconducting annuli. A search for this effect by Bonn et al. did not find this effect, though Senthil and Lee came up with a new test [11] involving two concentric superconducting rings embedded in a non-superconducting substrate.

1.2.2 Time-Resolved Hall Probe Technique

At present, magnetic imaging high spatial resolution, high flux sensitivity, high temporal resolution magnetic imaging is unavailable. While the existing SHPM technique has high spatial resolution and flux sensitivity, it does not normally have a high temporal resolution. Usually, a Hall probe is biased with an audio-frequency sinusoidal current $I(t)$, and one measures the component of the voltage response $V_H(t)$ oscillating at the same frequency, with $V_H$ proportional to both the field and current.
Instead of using an audio-frequency bias, the Broun Lab has conceived of an approach inspired by the time-resolved magneto-optical technique: applying short, stroboscopic current pulses that briefly activate the Hall probe. Readout is still done at audio frequencies. The lower bound on the pulse time is the charge carrier relaxation time of the sensor material. Usually, this is in the picosecond range; for bismuth, it is a few hundreds of picoseconds [12].

![Figure 1.1: A sketched comparison of current (I), magnetic field (B), and Hall voltage vs. time (t) for a magnetic field scan using a time-resolved Hall sensor.](image)

With this higher speed, the time-resolved technique could be used to image rapidly-evolving magnetic systems. Time-resolved scanning would allow movies of such systems to be created, enabling the study of magnetodynamics over a micron-sized area while still retaining the core advantages of the basic SHPM technique: high spatial resolution combined with high flux sensitivity.

While time-resolved SHPM would not be able to simultaneously monitor all points of a large area this way, it could easily study repeatable time-varying phenomena, such as vortex dynamics discussed in the previous section.

### 1.2.3 Bioassays Using Magnetic Beads

There is a great deal of excitement around small “lab-on-a-chip” (LOC) systems that can use Hall probes and magnetic nanobeads [13] to analyze samples for chemical, biological, or radioactive agents. One could design a system that detected, say, cholera-causing bacteria in drinking water by introducing chemically-modified mag-
netic beads that clump together in the presence of the bacteria. This could then be detected by a small magnetic sensor, like a Hall probe.

LOCs have many advantages, including reduced waste of materials, large scale parallelizability, and lower cost of operation. A small, sensitive Hall probe or perhaps even a Hall probe array would be a valuable addition to an LOC system.

1.2.4 In-Situ Nanofabrication Using Photodoping

Another interesting project would be to integrate a near-field scanning optical microscope into the Hall probe microscope in order to allow in-situ magnetic scanning and photodoping of cuprate superconductor thin films, allowing for rapid prototyping and testing of nanoscale high temperature superconductor devices. Such a system would essentially be a nano “Etch-a-Sketch” for superconductors, giving us the ability to write wires, loops, capacitors, Josephson junctions, or any other elements and construct a circuit from them. Decca and co-workers have already demonstrated fabrication of ∼150 nm-wide wires and Josephson junctions with gaps of ∼100 nm [14].

Some of the proposed experiments mentioned in the previous sections can be done with just such a setup. With a laser integrated into the apparatus, we would be able to introduce defects into a sample and do the vortex dynamics experiments with good precision and control.

1.2.5 Magnetic Thin Films

With hard disk bit size and read head size decreasing exponentially for the past several decades, a great deal of effort has been put into magnetic thin films in order to keep up with the demanding pace of computer miniaturization. The last couple of decades of progress have largely been driven by improvements in the read heads, using the Nobel Prize-winning discovery of giant magnetoresistance. Feature sizes for ferromagnetic films in both the read heads and hard drive platters have dipped into the nanoscale, making these components more unstable and resistant to simple analysis.

In addition to shrinking devices, there is also much interest in making them perform faster. Hard disks could use a speed boost and so could magnetic random access memory, a non-volatile form of storage that could replace the transistor-based memory common in computer memory today. Studies of spin dynamics and relaxation
aided by Hall probes could yield those speed improvements along with decreased size and improved stability.
Chapter 2
Survey of Magnetic Imaging Techniques

“I see miracles all around me.
Stop and look around, it’s all astounding.
Water, fire, air, and dirt.
Fucking magnets, how do they work?”
— Insane Clown Posse, “Miracles,” from their album Bang! Pow! Boom!

This thesis is about Hall probes, but why use those instead of other sensors, such as SQUIDs or MFM tips? I have not yet given a detailed account of the alternatives. In this chapter, I will explain the magnetic flux detection mechanisms and obstacles to improved flux sensitivity and spatial resolution for the following techniques: scanning Hall probe microscopy, SQUID microscopy, magnetic force microscopy, and magneto-optics. For a more detailed comparison, the reader may want to study Chapter 3, where a more quantitative discussion of MFM, SQUID microscopy, and SHPM is presented, but the qualitative discussion of SHPM in this chapter should be sufficient background to see where Hall probes fit into the field of magnetic microscopy.

2.1 Desirable Traits in Magnetic Microscopy

Let us take a step back from the details and think abstractly about what makes a good technique in magnetic microscopy. Desirable traits include magnetic flux sensitivity, spatial resolution, imaging speed, a range of operating temperatures, non-invasive measurement, sensor durability, and easy interpretability of data.
The virtues of these characteristics should be quite obvious except, perhaps, for flux sensitivity. Flux is the product of the magnetic field and the area the field penetrates. A large sensor area usually reduces the spatial resolution, so putting a minimum on flux sensitivity is a good way of capturing the trade-off between field sensitivity and spatial resolution. We will see later that a detailed semiclassical treatment of the Hall sensor shows that flux sensitivity is indeed a fundamental figure of merit, with the scale of flux noise set by the quantum of magnetic flux, $\Phi_0 = \hbar/2e$.

In the sections that follow, we will evaluate magnetic microscopy techniques with these criteria in mind. Bending [7] gives a fairly comprehensive review of the state-of-the-art in magnetic imaging. We will examine many of the same techniques that he looks at, but with an emphasis on methods that provide quantitative magnetic flux measurements.

### 2.2 Magnetic Flux Detection Mechanisms

Here, we will briefly look at how several quantitative magnetic imaging techniques work. A discussion of their relative merits follows this section.

As mentioned before, SHPM is a way to map magnetic flux on a surface. The structure of the Hall sensor is essentially a pair of crossed wires, one for running current, the other for measuring a Hall voltage proportional to the magnetic field in the junction, also known as the active region. The principle of operation is the classical Hall effect, in which a magnetic field perpendicular to a current creates a transverse potential difference proportional to the field strength. The probe, typically on a small semiconductor chip, is rastered over the sample surface to acquire a map of the local magnetic field. Existing Hall probes are notable for their high flux sensitivity and spatial resolution, currently $10^{-4}\Phi_0$ for 50 nm bismuth probes [15].

SQUIDs can also be used as scanning probes when structured as a superconducting ring connected to a pick-up coil for magnetic flux collection. The SQUID can be thought of as an electron interferometer, with electrons passing through two halves of the ring like photons through a double-slit optical interferometer. Magnetic flux changes the Aharonov-Bohm phase of the electrons, causing constructive and destructive interference on opposite sides of the ring. Although under certain conditions the SQUID is in a resistance-free state, it must be operated in a dissipative regime in order to extract a voltage signal that can be used to measure the flux through the pick-up coil. SQUIDs are particularly important to magnetic microscopy because
they are the benchmark for magnetic flux sensitivity. Surprisingly, we will see that properly optimized Hall probes have a flux sensitivity similar to that of SQUIDs.

MFM is yet another scanning probe technique, notable for its very high spatial resolution. MFM involves rastering a magnet stuck to an atomic force microscopy tip. The tip is attracted or repelled by magnetic features on the surface scanned. This causes deflections of the cantilever holding the tip, which can be detected and translated into measurements of force using Hooke’s law. Interestingly, calculations indicate that an MFM tip can also have a flux sensitivity close to that of a SQUID.

Magneto-optical Kerr effect microscopy, in contrast to the other techniques I am discussing, uses light instead of a matter probe to measure magnetic flux. When polarized light travels through an optically active material, the polarization of the reflected light is rotated by an angle proportional to the magnetic field in the material. The reflected light is then viewed through a polarizer to acquire magnetic field information.

2.3 Magnetic Flux Resolution

In this section, I will qualitatively discuss the magnetic flux resolution limits for SQUIDs, Hall probes, and MFM tips. As previously noted, a detailed quantitative discussion is deferred until chapter 3. Note that all flux resolutions stated here are for systems at a standard temperature of 4 K.

Let us start with SQUIDs. Naively, one would expect a superconducting interferometric sensor to be a noiseless and perfect sensor. This is not true. In fact, as alluded to earlier, SQUIDs must be operated in a dissipative flux-flow regime in order to acquire a measurable signal and are clearly not noiseless in such a state. In addition, Josephson junctions in a SQUID must be modified by the inclusion of shunt resistors in order to tame hysteretic behaviour; these shunt resistors are a source of thermal noise. Despite these issues, SQUIDs are widely regarded as the benchmark sensors for flux sensitivity, with flux sensitivity in the white noise regime as low as $0.3 \frac{\mu\Phi_0}{\sqrt{\text{Hz}}}$ [16], where $\Phi_0 = h/2e$. Furthermore, $1/f$ noise in SQUIDs is particularly low.

Now consider Hall probes. The Hall voltage is caused by the deflection of and accumulation of moving charge carriers to the sides of a conductor in the presence of a magnetic field. For a given Hall probe material, we would want to maximize the bias current passing through the probe because, in the linear response regime, the resulting Hall voltage is proportional to current. The amount of current that the
probe can tolerate is constrained by ohmic heating of the sensor and the amount of heat that can be conducted away from the probe’s active region. Heat is typically conducted away through the leads and the substrate on which the sensor sits. When these factors are taken into account, it turns out that the flux sensitivity of Hall probes can also be as low as $0.3 \, \mu\Phi_0/\sqrt{\text{Hz}}$, as I will show in a later calculation.

Let us also consider the limits of flux resolution for MFM tips. Recall that we want to measure the force on a cantilever-mounted magnetic tip. The main fundamental obstacle to the detection of magnet-induced deflections is thermal vibrations of the cantilever. To improve flux sensitivity, we can either increase the deflection response to a given magnetic flux or decrease the RMS amplitude of thermal vibrations. Increasing the deflection response involves increasing the magnetization of the MFM tip, though this is constrained by the degree to which the MFM tip material will permanently magnetize. In my later calculation of flux noise for the static mode MFM, we will see that flux noise also increases with cantilever damping. A decrease in damping would be beneficial to static mode flux detection. The same calculation shows that when using a particular commercially available MFM tip, the flux sensitivity is also $0.3 \, \mu\Phi_0/\sqrt{\text{Hz}}$. This could simply be an interesting coincidence, or perhaps this is indicative of a common underlying physical mechanism, like the Aharonov-Bohm phase. There will be more on that later.

## 2.4 Spatial Resolution

The scanning probe techniques, SHPM, MFM, and SQUID microscopy, generally have good spatial resolution due to the ability to make the scanning probes very small.

Hall probes can be made arbitrarily small as long as the charge carriers do not enter the ballistic regime, where the mean free path of electrons and holes exceeds the size of the active region of the probe. When this happens, the semiclassical trajectories of the charge carriers are randomized from chaotic collisions off rounded corners in the probe. The Hall effect relies on a separation of charge, so randomizing the locations of the charge carriers will quench the Hall effect \cite{17, 18}. So far, Hall probes have been made with active regions as small as 50 nm \cite{15}.

MFM’s spatial resolution is limited by the size of the magnetic tip used. As long as a tip can be made small enough and the sample’s fields are not sufficient to change the magnetization of the tip, then MFM can be used with very fine resolution. A commercial system claims a spatial resolution of 10 nm \cite{19}. As MFM tips are made
smaller, we may reach the super-paramagnetic limit, where the tips’ magnetization can flip spontaneously, either as a result of thermal fluctuations or in response to spatially-varying fields above the sample surface.

SQUIDs are limited in spatial resolution by material properties. SQUIDs work best when the magnetic field does not penetrate deep into the superconducting ring. This condition is met as long as the diameter of the wire for the ring is a few times larger than the magnetic penetration depth. For niobium, a popular SQUID material, the penetration depth is $\lambda = 90$ nm, which limits device size to a few hundred nanometres in diameter or larger.

MO imaging’s spatial resolution is diffraction-limited due to the use of light. The limit of resolution is on the order of hundreds of nanometres, depending on the specific wavelength of light used.

### 2.5 Imaging Speed

When it comes to imaging speed, magneto-optic microscopy is the big winner here. The imaging speed advantage manifests itself in two ways: optical techniques that allow a whole field of view to be imaged in parallel at video rates, and ultra-fast laser techniques that allow MO measurements at a single spot on sub-picosecond time scales.

There is one important caveat. Magneto-optical imaging has an additional complication: unless the sample is magnetic, e.g. a ferromagnet, MO imaging requires an optically active indicator medium to be grown on one side of the film being imaged. There are two possible problems. The first is that MO imaging studies the indicator medium directly and the sample of interest indirectly. If the magnetic features on the indicator change more slowly than those of the sample of interest, then the indicator cannot provide trustworthy measurements of the sample. The second possible problem is that the indicator medium could perturb the physics of the sample of interest.

The imaging speeds of the scanning probe techniques are all limited by the speed at which a probe can be rastered over a magnetic surface. It is possible to improve scan rates of scanning probes by using arrays of sensors, but the support and measurement hardware required for use of an array is cumbersome, especially for the field of view and spatial resolution possible in MO imaging.
2.6 Operational Temperature Range

There are no major temperature constraints for most of the sensors except SQUIDs. SQUIDs must be superconducting to work, so the operating temperature range will be cryogenic until someone discovers room temperature superconductors.

SHPM and MFM are constrained by the melting temperatures of their probes, which are, of course, material dependent. For bismuth Hall probes, for example, the melting temperature is about 270 °C. MFM has the additional constraint that it must be used with the MFM tip kept below its Curie temperature.

MO microscopy does not have a clear temperature limit since the only temperature requirement is for the detector to be sufficiently protected from heat to work.

2.7 Non-Invasive Measurement

Sensors that create strong magnetic fields are invasive and may significantly change the magnetic features being imaged. Of the techniques discussed, MFM is most invasive. This is desirable in some cases. For example, a magnetic tip has been used to manipulate vortices in high temperature superconductors [20].

For non-invasive techniques, we should turn to SQUID microscopy, SHPM, and MO imaging. SQUIDs and Hall probes operate using very small electrical currents, which do not generate significant magnetic fields. Magnetic fields generated from MO imaging are similarly small.

2.8 High Sensor Durability

The scanning probe techniques tend to be more hazardous to the integrity of the sensor since they involve rapidly rastering a sensor very close to the sample surface. This is especially true for high spatial resolution sensors, which tend to be small and delicate. Although I will not discuss hard drive read heads in detail here, they are notably well-protected and quite durable. Later, we will look at a novel design for Hall probes that takes some inspiration from hard drives to improve durability.

MO imaging, being a non-contact technique, is best for keeping the sensing hardware safe from most damage.
2.9 Easy Interpretability of Data

SHPM is easy to interpret at low magnetic fields, with the Hall voltage scaling linearly with magnetic field. Even at moderate to large fields, the magnetic field can be calculated or the Hall probe can be engineered to remove non-linearities.

SQUIDs measure magnetic flux going through the SQUID’s pickup coil, which is also straightforward to interpret.

MFM can be difficult to interpret because the sample’s magnetic fields interact with the MFM’s magnetic film in a non-trivial geometry. MFM users can also run into problems if the invasive magnetic tip changes the magnetization of the sample or vice versa.

MO imaging, with a signal linear in magnetic field, is also easily interpretable.

2.10 A Summary of the Comparisons

Table 2.1 sums up the above survey of quantitative magnetic imaging techniques. Hall probes compare favourably in most of the figures of merit, particularly in the vital areas of spatial resolution and flux sensitivity, where it rivals the best known magnetic imaging methods.

<table>
<thead>
<tr>
<th></th>
<th>MFM</th>
<th>SQUID</th>
<th>MO</th>
<th>SHPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial resolution (nm)</td>
<td>30</td>
<td>200</td>
<td>300</td>
<td>70</td>
</tr>
<tr>
<td>Flux sensitivity ((\mu \Phi_0/\sqrt{\text{Hz}}))</td>
<td>0.3</td>
<td>0.3</td>
<td>10^5</td>
<td>0.3</td>
</tr>
<tr>
<td>Imaging speed</td>
<td>slow</td>
<td>slow</td>
<td>fast</td>
<td>slow</td>
</tr>
<tr>
<td>Operating temperature range</td>
<td>wide</td>
<td>cryogenic</td>
<td>wide</td>
<td>wide</td>
</tr>
<tr>
<td>Non-invasive?</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Easily interpretable?</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Robust to abrasion?</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
</tbody>
</table>

Table 2.1: A comparison of estimated limits for capabilities of various magnetic imaging techniques used at 4 K. \(\Phi_0\) is the two-electron flux quantum. A “wide” operating temperature range means 0–300 K or higher. I have used characteristics of two different sensors for the SQUID flux sensitivity [16] and spatial resolution [21]. The MO flux sensitivity comes from Bending [7].

It is interesting to note that MFM, SQUID microscopy, and SHPM are all of similar flux sensitivity, though the comparison is done with the assumption that the sensors quickly eliminate transient oscillations that occur when the flux through
these sensors changes. SQUIDs and Hall probes are better in this regard since their
electron-based systems relax more quickly than the vibrating cantilever in an MFM
tip of the same flux sensitivity. Because signal-obscuring transients disappear faster
in SQUIDs and Hall probes, I would rank those devices above MFM tips in flux
sensitivity.

Perhaps it is not so surprising that SQUIDs and Hall probes are similarly sensitive
when we consider the fundamental physics of quantum interference that these probes
share. All of these magnetic imaging methods are essentially measurements of the
probabilities that electrons take different paths to get from point A to point B, and
may be a manifestation of the Ehrenberg-Siday-Aharonov-Bohm (ESAB) effect. The
SQUID makes almost ideal measurements of magnetic field; the main sources of
noise in SQUID magnetometry are thermal fluctuations from shunt resistors and
flux-flow resistivity. The flux sensitivity of a SQUID is probably quite close to the
best possible. Hall probes can be seen in a similar fundamental light, with the Hall
voltage determined by relative probabilities of electrons ending up in one voltage
lead or the other. These probabilities are modified based on the ESAB phase of the
electron wavefunction, making Hall probes plausibly quantum in nature, particularly
at small scales.

\footnote{Better-known as the Aharonov-Bohm effect, but properly known as in this text.}
Chapter 3

Flux Sensitivity in Detail

“A disk or strip of metal, forming part of an electric circuit, was placed between the poles of an electro-magnet, the disk cutting across the lines of force. The two poles of a sensitive galvanometer were then placed in connection with different parts of the disk, through which an electric current was passing, until two nearly equipotential points were found. The magnet current was then turned on and the galvanometer was observed, in order to detect any indication of a change in the relative potential of the two poles.

Owing probably to the fact that the metal disk had considerable thickness, the experiment at that time failed to give any positive result. Prof. Rowland now advised me, in repeating this experiment, to use gold leaf mounted on a plate of glass as my metal strip. I did so, and, experimenting as indicated above, succeeded on the 28th of October in obtaining, as the effect of the magnet’s action, a decided deflection of the galvanometer needle.”

— Edwin Herbert Hall, On a New Action of the Magnet on Electric Currents [22].

In the preceding survey chapter, I explained where Hall probes fit into the overall field of magnetic imaging. Hall probe flux sensitivity compares well with the best known sensor for flux sensitivity, the SQUID, while having a better spatial resolution. It is not the technique with the highest spatial or temporal resolution, but there are strong prospects for improving both of those attributes with miniaturization, the novel Hall probe design, and time-resolved Hall probe microscopy.
I did not present in detail how we can determine the flux sensitivities of the techniques we considered. Indeed, I kept quoting results from calculations that I said would be shown later. Later is now.

I will give a detailed account of the theoretical operation of SQUIDs, Hall probes, MFM tips, and magneto-optics, with a greater emphasis on mathematics and calculations. Of particular interest are detailed estimates of flux sensitivity for Hall probes and MFM tips. I go into considerable detail for Hall probes.

### 3.1 The Generic Experiment

When measuring magnetic flux, what is the best we can do? How accurate and precise can we possibly be? We will be answering this question in detail across three different techniques, so a look at the workings of a generic experiment may be useful for illuminating what to compare across all three techniques.

A schematic of a generic experiment is laid out in figure 3.1. It shows the series of operations involved in converting a physical variable, such as magnetic flux, into a numerical variable that corresponds to the amount of flux. The signal-to-noise ratio of the experiment will be highest at the source, at the physical system. All other steps to transform the signal, amplify it, and record the result can only degrade the signal-to-noise ratio by introducing noise. The best we can do is to minimize the degradation while still acquiring useful information.

![Figure 3.1: A generic experiment, using a transducer to convert a physical variable into an electronic variable that can be measured and digitized. Noise and errors have the greatest effect on the experiment prior to signal amplification.](image)

For magnetic microscopy, the physical system to be studied is something with spatially-varying magnetic features and the physical variable to be studied is magnetic...
flux. While physical noise is possible, a magnetic system is usually quite stable below the Curie temperature, so physical noise due to spin flips is usually minimal. We will also assume there is no systematic error in the measurement method.

The noise that limits the flux sensitivity in magnetic microscopy largely comes from the transducers. In MFM, there are thermal vibrations of the cantilever. For SQUID microscopy and SHPM, electronic thermal noise arises in resistive parts of the probes. For easy comparison between all these methods, the noise should be quantified as an equivalent flux noise. When reading such figures, it is important to keep in mind that for our magnetic systems, the variations in the numerical variable we record do not correspond to actual fluctuations in the physical variable. On the contrary, transducer noise causes the fluctuations in the measurements. The equivalent flux noise is the threshold of physical system flux changes that cannot be observed because of transducer noise (and, in more complex models, RF interference and amplifier noise).

3.2 Noise

Now that we have examined how noise enters experiments in general, let us build a mathematical description of noise. We will mainly be dealing with noise in the frequency domain as a spectral density, with specific examples of noise that limit the flux sensitivity of SQUIDs, MFM tips, and Hall probes.

3.2.1 Noise Distributions

Let us first discuss how we quantify and describe noise. Often, we want to consider noise in the frequency domain, where we have a suite of powerful mathematical and experimental tools, which includes Fourier transforms, power spectral densities, band-pass filters, and lock-in amplifiers. Lock-in techniques are particularly helpful in filtering out noise at non-signal frequencies. It would be very helpful to have some way of quantifying noise in the frequency domain and linking that to the statistical variance of the noise.

Enter the spectral density. It is proportional to the noise power absorbed by a system per unit of measurement bandwidth. Integrating it over the bandwidth gives the variance in the noise. Mathematically, we can represent it this way:

$$\delta y^2 = \int_{f_0}^{f_0+\Delta f} G(f)df,$$

(3.1)
where \( \delta y^2 \) is the variance in some variable \( y \), \( f \) is frequency, \( G(f) \) is the spectral density corresponding to given measurement variable \( y \), and \( \Delta f \) is the measurement bandwidth. \( G(f) \) has units of \([y^2]\) per unit frequency. When a noise spectrum varies smoothly in the frequency domain, as it often does with natural sources of noise, spectral density describes the effects of noise well.

The true power of the frequency domain framing of the problem becomes apparent when we use lock-in techniques. Noise is quite difficult to remove in the time domain, where simple time-averaging is the usual technique. It is relatively straightforward to filter out noise in the frequency domain, particularly when the noise spectrum is broadly distributed, as it often is. Lock-in amplifiers are able to capture an oscillating signal over a narrow measurement bandwidth and filter out all the broad spectrum noise. Judicious use of the lock-in technique would focus measurements on a part of frequency domain where the noise power spectral density is low, further minimizing the measured noise. For instance, when the noise distribution goes like \( 1/f \), measurements should be done at higher frequencies.

With that in mind, let us survey the varieties of noise we will encounter in our examination of magnetic probe techniques. We will look at three varieties: thermal electronic, thermal vibration, and telegraph noise.

### 3.2.2 Thermal Electronic Noise

Thermal electronic noise, also known as Johnson-Nyquist noise, is a good subject with which to start because it is quite simple and illustrative of how we can use spectral density. For the magnetic imaging techniques we are covering in detail, thermal electronic noise is a problem in Hall probes and SQUIDs. Interestingly, the noise problem can be framed as one of blackbody radiation from a resistor in one dimension, and the spectral density is flat in the frequency domain when \( hf \ll k_B T \).

In that limit, the spectral density for the resulting voltage noise is

\[
G(f) = 4k_B T R,
\]

where \( k_B \) is the Boltzmann constant, \( T \) is the temperature, and \( R \) is the resistance.

We can then calculate the RMS fluctuation in voltage using equation (3.1):

\[
\delta V = \sqrt{\int_{f_0}^{f_0+\Delta f} G(f) df} = \sqrt{\int_{f_0}^{f_0+\Delta f} 4k_B T R df} = \sqrt{4k_B T R \Delta f}.
\]

(3.3)
Note that by using a narrow measurement bandwidth $\Delta f$, we can screen out all of the broad spectrum noise outside of that frequency range.

3.2.3 Thermal Vibrational Noise

The physics of thermal vibrational noise on a cantilever in MFM are quite similar to those of thermal electronic voltage noise. Indeed, they have similar governing differential equations and can both be derived from the fluctuation-dissipation theorem, which links stronger dissipative forces to stronger random fluctuations at equilibrium. The spectral density for thermal noise force on a cantilever strongly resembles that of thermal electronic voltage noise in resistors:

$$G(f) = 4k_B T\gamma,$$  \hspace{1cm} (3.4)

where the only change has been to replace the resistance $R$ with the cantilever damping constant\(^1\) $\gamma$. The RMS cantilever force $\delta F$ is then

$$\delta F = \sqrt{4k_B T\gamma \Delta f}.$$  \hspace{1cm} (3.5)

As in the case of thermal electronic noise, these formulas apply in a low frequency limit, which for the cantilevers is $f \ll f_0$, where $f_0$ is the resonant frequency of the cantilever.

3.2.4 Telegraph Noise

Telegraph noise is a form of excess noise seen in Hall probes. It appears as changes in the Hall coefficient, which cause rapid shifts between discrete voltage levels, as shown in figure 3.2. Telegraph noise is probably caused by charges moving onto and off of dopants or defects in a sample with low carrier density or number of carriers [23, 24], though there is some dispute [25]. When the number of dopants or defects is large, the noise spectrum tends to go like $1/f$, as a detailed treatment in Appendix A.1 shows. There is no actual singularity in the spectral density at zero frequency; it levels off close to DC. However, it is still troublingly large, and low-frequency noise in multiplicative factors, such as the Hall coefficient, is not screened out well by either

---

\(^1\)The damping constant is defined such that equation of motion for the cantilever is $F(t) = m_{\text{eff}}\ddot{z} + \gamma \dot{z} + kz$, where $F$ is the thermal force and $k$ is the spring constant.
lock-in techniques or time-domain averaging. We will have to avoid telegraph noise some other way.

Telegraph noise is most prominently seen in small semiconductor Hall probes because carrier density is very low in these probes; small fluctuations in carrier number cause large fluctuations in carrier density and Hall coefficient.

![Graph of telegraph noise in a GaAs/Al$_x$Ga$_{1-x}$As Hall probe](image)

Figure 3.2: A graph of telegraph noise in a GaAs/Al$_x$Ga$_{1-x}$As Hall probe, from work by Hicks et al. Reproduced from [25], with the permission of AIP Publishing.

## 3.3 A Fair Flux Sensitivity Metric

We now know a few things about noise and how it makes accurate measurements more difficult, but we have not yet discussed how we evaluate the quality of such measurements. We will do so now. Of particular importance here is the measurement bandwidth, which we have largely glossed over so far. Bandwidth is strongly linked to measurement uncertainty and measurement time.

Before we dive into the detailed physics of these magnetic imaging techniques, let us first consider more carefully how we evaluate the flux sensitivity of a particular technique. Suppose technique #1 can resolve Φ$_0$ of flux with 1 ms of averaging and technique #2 can detect the same flux with 3 ms of averaging. To say that the two techniques have the same flux sensitivity would be unfair. Technique #1 has a better flux sensitivity because it requires less time to reach the same conclusion.

The averaging time is not always a trivial matter of a few milliseconds. One could tolerate an experiment that reaches a certain conclusion in several years, which is
compatible with a graduate degree, but it becomes very unpalatable as the necessary
averaging time extends to the duration of a career. An example of an experiment
in which these considerations are highly relevant is BICEP2, a telescope experiment
for observing the effects of gravitational waves on the cosmic microwave background.
About 590 days of data needed to be analyzed for the BICEP2 Collaboration to claim
the discovery of gravitational waves from the Big Bang [26]. Even that effort was
not sufficient as the claim was subsequently retracted [27]. A slightly more absurd
example is the 2005 Ig Nobel Prize-winning pitch drop experiment, in which pitch is
allowed to flow through a funnel in order to estimate its viscosity [28]. A single drop
takes 7–12 years to drip down. Clearly, the flux sensitivity metric must have some
consideration of averaging time.

To reiterate, in evaluating the flux sensitivity of a technique, we must consider
not only the absolute amount of flux that can be detected, but also the time taken
to average the flux measurements. One way to build averaging time into our flux
sensitivity metric is to incorporate the bandwidth of the measurement, or the spread
of Fourier space frequencies used to average the measurement. A low frequency or
small bandwidth corresponds to a longer averaging time.

For a concrete example, consider thermal noise in a resistor. In a resistor at finite
temperature, voltage fluctuations are observed, arising from blackbody radiation. The
RMS voltage from these fluctuations is given by the following:

\[ \delta V = \sqrt{4k_BTR\Delta f}. \]

We can decrease the bandwidth by increasing averaging time, thus decreasing the
noise voltage. The convention is to divide the voltage by \( \sqrt{\Delta f} \) to account for the
influence of bandwidth. We call the resulting quantity the noise voltage density.

One more thing we should note is the temperature of each sensor in a comparison.
Thermal noise is an important factor — usually the dominant factor — in the most
sensitive probes of flux. Quoting flux sensitivities at different temperatures would
unfairly skew comparisons.

In short, we should express the sensitivity of a measurement of \( x \) in the units
of \( x \) per \( \sqrt{\text{Hz}} \), at some temperature \( T \). For the comparison of magnetic microscopy
techniques, we will use \( T = 4 \text{ K} \). This is a typical temperature for studying vortices in
superconductors, for which MFM tips, SQUIDs, and Hall probes have all been used.

With the flux sensitivity metric now established, let us look at SQUIDs, Hall
probes, and MFM tip flux sensitivity limits more quantitatively.
3.4 Flux Sensitivity of SQUIDs

Recall that SQUIDs are regarded as the most sensitive magnetic flux sensors known today, with a flux sensitivity of about $0.3 \mu \Phi_0 / \sqrt{\text{Hz}}$. One big advantage of SQUIDs is that their coherence extends over long distances, so large loops can be used to obtain very good field sensitivity. This is not the case for Hall probes. SQUIDs’ sensitivity enables them to be used in situations where the signal may be very weak, such as in brain imaging and natural resource exploration. Our exploration of how the DC SQUID works and why it is very sensitive has been superficial thus far, so let us now go into more technical detail.

The physics of the DC SQUID is similar to that of the Michelson-Morley interferometer, except using superconducting currents instead of coherent light. Figure 3.3 shows a schematic of the SQUID, with the interferometer shown on the right side; the interferometer is made of a superconducting material with a pair of weak electrical connections called Josephson junctions. These links may be weak due to constrictions in the wires or thin insulating layers. Each Josephson junction is only able to sustain a certain maximum superconducting current before becoming dissipative. Shunt resistors are also placed in parallel with each junction to prevent hysteresis. In addition, the junction-shunt system must obey the following condition to prevent non-hysteretic retrapping behaviour:

$$\beta_C = \frac{2\pi I_0 R^2 C}{\Phi_0} \lesssim 1,$$  (3.6)

where $C$ is the effective capacitance of the junction, $I_0$ is the critical current for the junction, and $R$ is the shunt resistance.

Figure 3.3: A schematic of a DC SQUID. A current bias $I$ is applied through the SQUID loop, but has two interfering paths through which to travel.
The pickup loop couples magnetic flux into the SQUID, changing the Ehrenberg-Siday-Aharonov-Bohm phase of the electrons. Superconductor phase coherence in the presence of flux leads to fluxoid quantization, defined as follows:

\[
\Phi - \oint \frac{m_e \mathbf{v}_s}{e} \cdot d\mathbf{s} = \frac{n\hbar}{2e}
\]

where \(\Phi\) is the flux, \(m_e\) is the electron mass, \(\mathbf{v}_s\) is the velocity of superconducting electrons, \(n\) is an integer, and the line integral is performed around the SQUID loop. The SQUID cannot support a non-quantized fluxoid and will circulate the minimum amount of current required to keep the fluxoid quantized. For the rest of the discussion, assume the SQUID wire is thick enough to screen out all magnetic fields, with the result that there exists a contour along which \(\mathbf{v}_s\) is zero and therefore, the magnetic flux itself is quantized.

Fluxoid quantization, along with limits to supercurrent flow through Josephson junctions, make the SQUID dissipative and create a voltage signal that can be measured to calculate the flux coupling to the SQUID. I will now illustrate how to interpret that voltage as a flux.

\[I_{\text{max}} = 2I_0 |\cos(\pi \Phi / \Phi_0)|\]  

We will concentrate our efforts on a small range of fluxes that capture the basic pattern of SQUID behaviour; in the case of our ideal SQUID, this range is 0–0.5 \(\Phi_0\). Figure 3.5 shows voltage as a function of current for that range of fluxes. If we
operate the SQUID at some constant current greater than the SQUID’s maximum critical current, we get a range of voltage responses shown by the arrows in figure 3.5. As the SQUID flux increases, the system traverses different $V-I$ curves to higher voltages.

Let us now switch to a voltage-flux perspective, shown in figure 3.6. This is where we can convert a voltage measurement into a flux measurement. If we set the bias current in the SQUID too low, we lose flux information because we get curves that are discontinuous, like the ones below the curve with the thick line. For the SQUID to work, it must always be in the dissipative regime with bias current greater than the device’s maximum critical current. When this condition is met, we get a sinusoid-like voltage periodic with \( \Phi_0 \):

$$V = \frac{R}{2} \sqrt{I^2 - \left(2I_0 \cos(\pi \Phi/\Phi_0)\right)^2} \quad (3.9)$$

This setup is often combined with lock-in techniques and a solenoid to create an active feedback system that keeps the SQUID in the part of parameter space that is most sensitive to changes in magnetic field, where the $V$ vs. $\Phi$ curve has the greatest slope. This is called a flux-locked loop.

As mentioned, the SQUID operates in a dissipative regime and the addition of shunt resistors introduce additional noise to the sensor. Despite this, both thermal noise and $1/f$ noise in the SQUID are both very low, awarding it the crown of “most sensitive magnetic flux probe.” Assuming only thermal noise, computer simulations by Clarke [29] indicate that the minimum flux noise in a SQUID is

$$\Phi_{\text{min}} = \sqrt{\frac{16k_B T L^2}{R \Delta f}}, \quad (3.10)$$

where $L$ is the self-inductance of the SQUID loop and $R$ is the resistance of one of the shunts, with the condition that $2LI_0/\Phi_0 \sim 1$. That additional condition means that if we try to improve the flux sensitivity by shrinking the SQUID loop area, we must increase the critical current by changing the Josephson junction geometry. We are constrained in doing so by equation (3.6) since a change in junction geometry, say by increasing the junction area, changes both the critical current and capacitance. At some point, we would be unable to both decrease the SQUID loop area and increase the critical current.

While constraints of SQUID flux sensitivity are only loosely determined here, we note that the best reported flux sensitivity is $0.3 \, \mu\Phi_0/\sqrt{\text{Hz}}$ at 4 K [16].
Figure 3.5: Voltage vs. current curves are plotted from left to right for 0.5, 0.4, 0.3, 0.2, and 0 $\Phi_0$ of flux in a SQUID. As the flux through the SQUID increases at constant current, the system traverses different curves going up in voltage, as the arrows show.

Figure 3.6: Voltage vs. flux curves for different amounts of current flowing through a SQUID. Like the current vs. flux curves, these ones are periodic in flux. The family of curves is generated for different amounts of constant current bias through the SQUID, with the higher current curves generating higher voltages. The thick-lined curve and higher ones have a bias current greater than the zero-flux critical current.
3.5 Magnetic Force Microscopy Flux Sensitivity

MFM is a magnetic microscopy technique that is well-known for its high spatial resolution. It is often used for characterizing nanomagnetic features, such as hard disk bits. We will see that MFM’s flux sensitivity is about as good as that of SQUIDs or Hall probes, but the strong magnetic field of the tip is quite invasive and not suitable for all applications.

3.5.1 Magnetic Interactions

MFM, as its name suggests, involves the measurement of the force exerted on a magnetic tip as it scans over a magnetic surface. The MFM tip is attached to a silicon cantilever, which bends as the tip is pushed by magnetic forces due to surface fields. The displacement of the tip can be measured by a laser reflecting off the top side of the tip and into a photodetector, as shown in figure 3.7.

For the purposes of this calculation, we are only considering static mode MFM. Oscillating mode MFM is able to measure magnetic field gradients, but we are interested in magnetic field and flux measurements, which are in the domain of static mode or DC MFM.

The force exerted due to the tip-surface interaction is given by

\[
F = \iiint_{\text{tip}} \nabla (\mathbf{M} \cdot \mathbf{B}) \, dV,
\]  

\[ (3.11) \]
where $M$ is magnetization of the tip. Note that accurate determination of magnetic field from a sample is calculation-intensive because the surface fields are convolved with the gradient of the tip magnetization. For our flux sensitivity calculation, an approximation will suffice. Although MFM tips are usually pyramids or cones of ferromagnetic material, we will approximate the probe as a magnetic monopole with charge

$$m = \mu_0 M \pi R^2$$

(3.12)

at the end of the tip, where $M$ is the tip magnetization and $R$ is the radius of the tip’s end. This is a standard assumption [30], valid when two conditions are met: the magnetization of the tip is along the axis of the cone or pyramid and the length of the tip $l \gg R$. With these simplifications, equation (3.11) becomes

$$F = \frac{mB}{\mu_0} = M\pi R^2 B.$$  

(3.13)

Most magnetic force microscopes are made to detect only vertical deflections, so if we define the vertical to be parallel to the $z$-axis, then we get

$$F_z = M\pi R^2 B_z = M\Phi_z,$$

(3.14)

where $\Phi_z$ is the magnetic flux through the tip in the $z$-direction. Observe that flux is the quantity that naturally fits in the equation here and neatly eliminates the geometric factor.

### 3.5.2 Fundamental Limits

Thermal motion of the cantilever sets a fundamental limit on the sensitivity. The motion can be modelled as a random thermal noise force with a white noise spectrum. For a forced damped harmonic oscillator model of the MFM cantilever, the RMS thermal noise force is given by [31]

$$\delta F_z = \sqrt{4k_B T \gamma \Delta f},$$

(3.15)

where $k_B$ is Boltzmann’s constant, $T$ is the cantilever temperature, $\gamma$ is the cantilever drag coefficient, and $\Delta f$ is the measurement bandwidth.
Assuming a signal-to-noise ratio of 1, we can define a minimum detectable magnetic field by using the RMS thermal noise force and equation (3.14):

$$\delta F_z = \sqrt{4k_B T \gamma \Delta f} = M \delta \Phi_z = M \pi R^2 \delta B_z. \quad (3.16)$$

To be clear, $\delta B_z$ is the change in magnetic field we cannot distinguish due to cantilever thermal vibration. We can rearrange that equation into the following:

$$\delta B_z = \frac{1}{\pi R^2} \sqrt{\frac{4k_B T \gamma \Delta f}{M^2}}, \quad (3.17)$$

$$\delta \Phi_z = \sqrt{\frac{4k_B T \gamma \Delta f}{M^2}}. \quad (3.18)$$

Equation (3.18) is an important general result. We see from the equation that noise increases with temperature and dissipation, but decreases with magnetization $M$. It is not clear to me what the limit would be for decreasing the damping of the MFM cantilever, though unusual techniques can and have been used to decrease damping, such as operation in a vacuum chamber to decrease air friction. The magnetization $M$ can be increased with judicious choice of materials, but a fundamental upper bound is placed by the total number of magnetic dipoles in the tip.

Let us look at commercially available MFM tips to get an estimate of the flux sensitivity possible today. We must first recast the variables in equations (3.17) and (3.18) in terms of standard manufacturer’s characteristics of cantilevers. For a very underdamped harmonic oscillator, we substitute for the damping term $\gamma \approx m_{\text{eff}} \omega_0 / Q$, where $\omega_0$ is the resonant frequency and $Q$ is the quality factor of the cantilever. We also substitute for the effective mass with $m_{\text{eff}} = k/\omega_0^2$. Thus, we get the following expressions for field and flux sensitivity:

$$\delta B_z = \frac{1}{\pi R^2} \sqrt{\frac{4k_B T k \Delta f}{\omega_0 Q M^2}}, \quad (3.19)$$

$$\delta \Phi_z = \sqrt{\frac{4k_B T k \Delta f}{\omega_0 Q M^2}}. \quad (3.20)$$

Now let us calculate the field and flux sensitivities for a commercially available, high spatial resolution MFM tip, like NanoAndMore’s PPP-MFMR [32]. A typical PPP-MFMR tip has the following characteristics:

- $M = 300 \text{ emu/cm}^3 = 3.00 \times 10^5 \text{ A/m},$
• $\omega_0 = 2\pi \times 75 \text{ kHz}$,

• $k = 2.8 \text{ N/m}$,

• $R = 50 \text{ nm}$.

$Q$ values of 35000 can be achieved under UHV conditions [33], so let us assume that and $T = 4 \text{ K}$ to get estimates of the spectral densities of the effective magnetic field and flux noise:

\[
\frac{\delta B_z}{\sqrt{\Delta f}} = 8.22 \times 10^{-8} \frac{T}{\sqrt{Hz}}, \tag{3.21}
\]

\[
\frac{\delta \Phi_z}{\sqrt{\Delta f}} = 3.12 \times 10^{-7} \frac{\Phi_0}{\sqrt{Hz}}. \tag{3.22}
\]

Interestingly, the MFM tip flux sensitivity is quite close to that of the most sensitive SQUID, $0.3 \mu \Phi_0/\sqrt{Hz}$.

However, this level of performance comes with the assumption that the MFM tip quickly damps non-thermal contributions to tip vibrations, relaxing to the thermal limit of flux detection. This assumption may break down if the tip quickly scans between two regions with large differences in magnetization. Such a scan would cause the tip to vibrate strongly, and in systems with large $Q$, the relaxation time can be large relative to scan rate. There will be trade-offs between three factors: scan rate, flux sensitivity, and level of sample magnetization.

Another caveat to the use of MFM is the possibility of the magnetic tip inducing undesired behaviour in the sample that is scanned. We are particularly interested in the possibilities of imaging vortices in type-II superconductors, but do not want the invasive field of an MFM tip to induce vortices in the sample. The first critical field of such substances can be as low as tens of gauss [34], so we would like the field from the MFM tip to be less than that. We will use the magnetic monopole model from before, which, by Gauss’s law, gives us the following expression for the magnetic field:

\[
B = \frac{m}{4\pi r^2} = \frac{\mu_0 R^2 M}{4r^2}, \tag{3.23}
\]

where $r$ is the distance from the field point to the tip. In scanning probe microscopy, we typically have $R \sim r$. Thus, for the invasive field to be less than 10 G, we need $M < 3.18 \times 10^3 \text{ A/m}$, resulting in a reduction of the flux sensitivity by about two orders of magnitude. The invasive field is occasionally useful, such as in measurement of vortex pinning forces and manipulation of vortices, but it is generally undesirable.
3.6 Hall Probe Flux Sensitivity: Basics

Finally, we shift our attention to Hall probes for a detailed accounting of the limits to their flux sensitivity. As we will see in the calculations that follow, in principle, Hall probes can be made with flux sensitivity close to that of SQUIDs and with better spatial resolution. The scalability of the Hall effect to smaller geometries, while maintaining high flux sensitivity, is what sets Hall probes apart from SQUIDs. This combination of flux sensitivity and size scalability make Hall probes potentially important for such applications as “lab-on-a-chip” magnetic bioassay systems [35], hard drive read heads, and, of course, magnetic microscopy.

For this section, we will focus on building a relatively simple, semiclassical, analytic model for Hall sensors made of materials with one band of charge carriers. The discussion here will be the foundation for further refinements of Hall probe theory, particularly for the section about using bismuth, a three-band material, for Hall sensors. But first, the one-band case.

The advantage of a simple model is that we quickly gain valuable insights into how to optimize Hall sensor performance. In our simple model, thermal noise and sensor self-heating are the major limiting factors to Hall probe flux sensitivity. Further insights will emerge from expressing the Hall probe’s sensitivity in terms of two fundamental properties of a conductor: the carrier mean free path $\Lambda$ and the Fermi wavelength $\lambda_F$, which acts as a useful measure of carrier concentration in a quantum system.

While N. David discussed many of the same basic ideas for his Master’s thesis [36], I have refined and extended those ideas, particularly for bismuth probes.

The detailed Hall probe discussion in this chapter is split into two parts. The first part touches on all aspects of Hall probe operation, but for clarity, it omits details of Hall probe heating calculations and is limited to metal Hall probes. Section 3.7 fills in those gaps and adds an analysis for semimetal Hall probes, using the example of bismuth to get a numerical result.

3.6.1 The Hall Effect

The principle of operation for the Hall probe is the Hall effect, which was discovered by Edwin Hall and described in the American Journal of Mathematics in 1879 [22].
By Hall effect, we refer to Edwin Hall’s classical discovery and not to the quantum\(^2\) or fractional quantum Hall effects that were discovered in the late 20\(^{th}\) century. It is characterized by a voltage difference that arises when a current runs perpendicular to a magnetic field.

The derivation of the classical one-band Hall effect for isotropic materials at low magnetic fields is a good introduction to how the classical Hall effect works, so let us go through it quickly. Consider the situation in figure 3.9, where a hole current flows in the \(x\)-direction with a magnetic field in the \(z\)-direction. The magnetic force on the holes causes them to move in the negative \(y\)-direction until there are enough charges to push back with an equal electric force:

\[
e v \times B = -eE,
\]

where \(e\) is the hole charge, \(v\) is its drift velocity, \(B\) is the magnetic field, and \(E\) is the electric field. At a given moment in time, the geometry of the charge arrangement looks similar to that of a parallel plate capacitor, only the holes are flowing through the wire. We can calculate the Hall voltage of this arrangement by integrating from the positive to the negative charges:

\[
V_H = -\int E \cdot dx = vBs,
\]

where \(s\) is the lateral dimension of the Hall probe. Usually, the Hall voltage is expressed in terms of experimental parameters that we can control directly, so we substitute for the drift velocity using an equation for current:

\[
I = pevsd,
\]

where \(p\) is the hole density and \(d\) is the Hall probe thickness. In this formulation, changing the current, the hole density, or the geometry are mainly methods of adjusting the drift velocity. Substitution for \(v\) gives us the familiar form of the one-band Hall voltage:

\[
V_H = \frac{IB}{p_ped}.
\]

\(^2\)We are assuming that the devices operate in the low-field regime, where the quantum Hall effect is not seen.
In materials with more than one band, the material-dependent parameters in the Hall voltage are usually combined into a Hall coefficient term, $R_H$:

$$V_H = \frac{IBR_H}{d}.$$  \hfill (3.28)

Thus, the one-band Hall coefficient is

$$R_H = \frac{1}{ep}.$$  \hfill (3.29)

Figure 3.9: A Hall probe with dimensions $s \times s \times d$.

Considering the two equations above, (3.28) and (3.29), one may na"ively think that smaller carrier concentrations would be beneficial for boosting Hall probe sensitivity. However, we could also enhance the Hall voltage by using a greater current in the probes, provided they had a greater carrier concentration. It is not immediately clear how to reconcile these opposing views, so let us now broaden our theoretical palate by introducing the ingredients of constraints as we cook up our Hall probe optimization problem. We will see in the next few sections how noise, Fermi wavelength, and mean free path figure into Hall probe performance.

### 3.6.2 Noise in Hall Probes

The tidy, semiclassical Hall voltage relation depicted in equation (3.28) is an ideal case. Noise limits the precision with which we can determine the magnetic field from

---

3Some workers in the field also include the remaining geometric factor $1/d$. 

35
the Hall voltage. Hall probes have two main forms of noise: thermal voltage noise and telegraph noise. There is not much further qualitative comment to be made about thermal voltage noise other than to say it is an intrinsic effect that cannot be eliminated, though it can be mitigated using lock-in techniques.

There is a good deal more to say about telegraph noise as it pertains to Hall probes, though. Telegraph noise is due to defects and impurities that cause fluctuations in carrier concentration. Telegraph noise can be worse in small semiconductor Hall probes since there is a lower absolute number of carriers in small probes with low carrier concentration; the fluctuation in carrier number makes a larger percentage change in the carrier concentration. In principle, telegraph noise can be removed with sufficiently fine control of sample growth conditions or by choosing a more conductive material. Lock-in amplifiers, as mentioned in section 3.2.4, are not too helpful in dealing with telegraph noise.

I am aware of one alternative scheme to deal with telegraph noise, but it has had mixed results. Hicks and co-workers attempted to suppress carrier concentration fluctuations in nanoscale semiconductor Hall probes by applying a gate voltage. The idea is to tune the energy of the fluctuating mode out of the thermal window so that an impurity’s charge state remains stable. While this strategy worked in getting rid of large jumps in Hall coefficient, there remained a lower level of Hall coefficient noise, even in the best cases [37]. This is likely still due to fluctuations in carrier density.

One way to remove telegraph noise and other Hall coefficient noise is by changing to a material with greater carrier density. Suppose we do that. Then we have only to worry about thermal noise and we get the following for the signal-to-noise ratio:

\[
\frac{S}{N} = \frac{V_H}{\delta V} = \frac{IBR_H/d}{\sqrt{4k_BTR\Delta f}},
\]

(3.30)

where \( R \) is the transverse resistance of the active region. For simplicity, define a minimum detectable field \( B_{\text{min}} \) assuming a signal-to-noise ratio of 1. \( B_{\text{min}} \) is the equivalent field we would detect if the RMS thermal voltage noise were actually due to field fluctuations instead of being independent voltage fluctuations. This gives us

\[
B_{\text{min}} = \frac{\sqrt{4k_BTR\Delta f}d}{IR_H}.
\]

(3.31)
Now define $P_{\text{bias}} = I^2R$ and $P_{\text{th}} = 4k_B T \Delta f$, then rewrite equation (3.30) as

$$B_{\text{min}} = \frac{Rd}{R_H} \sqrt{\frac{P_{\text{th}}}{P_{\text{bias}}}}.$$  \hfill (3.32)

So what happens to flux sensitivity if we move away from low carrier density materials, like semiconductors? We could try semimetals, which have moderate carrier densities, or even good metals with many carriers. To see what might happen in these other materials, we need to further develop our model. Increasing the carrier density would change many things at once — the electrical conductivity, thermal conductivity, and the Hall coefficient, for instance. The degree to which we can change parameters in the system will depend on the length scales of Hall probe physics, and that is our next topic of discussion.

### 3.6.3 Length Scales

The goal of this section is to rewrite the equation for magnetic field sensitivity in terms of the fundamental physical quantities for this situation, the mean free path $\Lambda$ and the Fermi wavelength $\lambda_F$. It turns out that the length scales $\Lambda$ and $\lambda_F$ can be used to parameterize all the important electrical properties in this problem, namely the resistivity and Hall coefficient. Writing it this way simplifies very nicely in the single-band case and provides some physical insight into what materials we can use to make Hall sensors.

Let us write expressions for the resistivity $\rho$, mobility $\mu$, effective mass $m^*$, and mean free path $\Lambda$. The expression for resistivity comes from Pouillet’s law and Drude theory for one band:

$$\rho = \frac{1}{e \mu} = \frac{RA}{\ell} = \frac{Rsd}{s} = Rd,$$  \hfill (3.33)

where we have assumed a Hall probe that is $d$ thick and has an active region of dimensions $s \times s$. $m^*$ is given by

$$m^* = \frac{p_F}{v_F} = \frac{h}{\lambda_F v_F},$$  \hfill (3.34)

where $p_F$ is the Fermi momentum and $v_F$ is the Fermi speed. The mean free path $\Lambda$ depends on the particle speed $v_F$ and mean time between collisions $\tau$:

$$\Lambda = v_F \tau.$$  \hfill (3.35)
Using the two equations immediately above, the Drude mobility can be rewritten thus:

\[ \mu = \frac{e\tau}{m^*} = \frac{e\Lambda \lambda_F}{h}, \quad (3.36) \]

which allows us to rewrite equation (3.33) in this way:

\[ \rho = \frac{h}{e^2 p \Lambda \lambda_F} = Rd. \quad (3.37) \]

We also need to use the one-band Hall coefficient:

\[ R_H = \frac{1}{ep}. \]

Substituting the two above relations into equation (3.32), our \( B_{\text{min}} \) expression, gives

\[ B_{\text{min}} = \frac{2\Phi_0}{\lambda_F \Lambda} \sqrt{\frac{P_{\text{th}}}{P_{\text{bias}}}}, \quad (3.38) \]

where \( \Phi_0 = h/2e \). There we have it: an expression\(^4\) for Hall probe sensitivity in terms of two fundamental material properties \( \Lambda \) and \( \lambda_F \). We still need to work out the power terms in the radical, but the basic form will not change. Neglecting the radical, we see that when there is one dominant carrier, we want to increase the mean free path and Fermi wavelength as much as possible to get the most sensitive Hall probes. With some assumptions, this will lead to a calculation of minimum detectable field and flux, and we will see that Hall probes are fundamentally flux sensors.

### 3.6.4 Fundamental Limits

We are finally ready for an examination of how far we can push the flux sensitivity of Hall probes. In the model we have developed so far, we have three main parameters that we can adjust more or less independently: the bias current (embedded in the \( P_{\text{bias}} \) term), the Fermi wavelength, and the charge carrier mean free path. In this section, we will answer the question of how far can we tune each of these parameters before we encounter fundamental limits.

\(^4\)The convention in the field is to express flux sensitivity in terms of two-electron flux quantum \( h/2e \), and this thesis adheres to that convention, but it is notable that the equation is slightly more elegant if we use the single electron flux quantum \( h/e \) to get rid of the extra factor of 2. Hall probes are expected to operate with effectively independent electrons and not with the electron pairing seen in SQUIDs.
Let us start with bias current. An obvious thing to do to improve the signal-to-noise ratio is to increase the bias current $I$, but we are limited by how much electrically-generated thermal power, $P_{\text{bias}}$, can be conducted away from the active region, where the most heat will be produced. There are two main channels for this to happen: conduction directly to the substrate and conduction through the leads.

The heat flow calculations are quite involved and will be examined in detail later in this chapter. For now, let the qualitative details above suffice. The reader anxious for details can read section 3.7.2 and then return here.

The degree to which we can increase the bias current depends not only on heat conduction, but also on our tolerance of the size of the temperature gradient between the active region and the substrate. Let us explore the two limits of hot and cold sensors relative to the surroundings.

**Hot Sensor Limit**

For the hot sensor limit, it is difficult to come to definite numerical estimates of flux sensitivity without knowing one’s specific experimental system, but we can at least establish the scaling of the flux sensitivity with temperature. Let us start by seeing how the temperature scales with bias current and power.

Section 3.7.2 establishes a 1D heat equation, already taking into account the probe geometry:

$$P_{\text{bias}} = 3(\kappa_e + \kappa_{\text{ph}})s\Delta T,$$

where $\kappa_e$ and $\kappa_{\text{ph}}$ are the electron and phonon thermal conductivities, respectively. $\Delta T$ is the temperature gradient between the probe and the substrate.

If we allow the probe to get hot, $\Delta T$ will not be small, so we must use the integral form:

$$P_{\text{bias}} = I^2R = 3s\int(\kappa_e + \kappa_{\text{ph}})dT.$$  \hspace{1cm} (3.39)

At sufficiently low temperature, $\kappa_e$ scales linearly with $T$, but $\kappa_{\text{ph}}$ goes like $T^3$. Thus, $P_{\text{bias}}$ goes like $T^2$ in the electron-dominated regime and $T^4$ in the phonon-dominated regime. Similarly, bias current scales like $T$ when electron conduction is large and like $T^2$ when the phonon term is large.

When we put this scaling back into equation (3.38), including the thermal voltage noise, we get

$$B_{\text{min}} \propto \frac{1}{\lambda F\Lambda \sqrt{T_i}} = \frac{1}{\lambda F\Lambda T^{(1-i)/2}}, \hspace{1cm} (3.40)$$
where $i$ is 2 or 4 depending on the conductivity regime. Assuming the Fermi wavelength and mean free path do not change too much with temperature, the minimum detectable field is clearly decreasing with temperature. There are some sensitivity gains to be made, depending on one’s tolerance for sensor heating, which may go as high as the sensor’s melting point.

**Cool Sensor Limit**

Now suppose we want to limit the sensor-substrate temperature gradient. Again, heat can be conducted with electrons and phonons, but an especially elegant result emerges in the metallic limit, where electron heat conduction dominates. In that limit, if we assume some reasonable experimental parameters, for instance a 1 K temperature difference out of the plane of the probe and a 1 Hz measurement bandwidth, we get

$$\sqrt{\frac{P_{th}}{P_{bias}}} = 1.6 \times 10^{-6} \frac{\lambda_F}{s}.$$  \hspace{1cm} (3.41)

where we have assumed a Hall probe active region size of $s \times s$.

It turns out that the $\lambda_F$ term cancels and we are left with the following for minimum detectable field:

$$B_{min} = \frac{2\Phi_0}{s\Lambda} \sqrt{\frac{P_{th}}{P_{bias}}} = 3.2 \times 10^{-6} \frac{\Phi_0}{s\Lambda},$$  \hspace{1cm} (3.42)

where we have assumed a Hall probe active region size of $s \times s$.

This is independent of Fermi wavelength and electron density, so we are free to choose any metal while engineering the probes to maximize the mean free path and minimize the smallest detectable field. For instance, we could dope or control crystallite size in the probes. A metal Hall probe would not have the telegraph noise a semiconductor probe might have since metals have so many more carriers. Using good metals could be a promising direction for Hall sensors.

The dimensions of the sensor fundamentally limit $\Lambda$. Hall sensors typically operate in the diffusive limit, where $\Lambda \ll s$. In this regime, an electron would collide several times inside the Hall sensor’s active region. After each collision, the electron would be nudged towards a particular lead of the sensor by the magnetic force; this is desirable.

In the ballistic limit, where $\Lambda \gg s$, an electron would be accelerated to high speed and collide with the boundaries of the active region. The edges of the Hall cross tend to be rounded. As shown in figure 3.10, collisions with these rounded edges
are chaotic, randomizing the semiclassical trajectories of the electrons in the cross, which randomizes the destination lead of each electron. The Hall effect relies on the electrons travelling into a preferential lead, though, so the Hall effect is quenched in the ballistic limit [17, 18]. For the geometry we are considering here, a Hall sensor must have $\Lambda \sim s$ to work optimally. Hall sensors have been specifically designed to work in the ballistic regime [38], but their study is beyond the scope of this thesis.

Figure 3.10: An illustration of the chaotic nature of ballistic scattering in a device with rounded corners.

Thus, we have a minimum detectable magnetic flux of

$$\Phi_{\text{min}} = B_{\text{min}} s^2 = 3.2 \times 10^{-6} \Phi_0.$$  \hspace{1cm} (3.43)

While this flux sensitivity is worse than what we saw for SQUID microscopy and MFM, we have brushed many details under the rug. When we consider those details, particularly phonon heat conduction, we will see in the next section that the Hall probe flux sensitivity matches that of the other techniques.

3.7 Hall Probe Flux Sensitivity: Bismuth

Having built the framework for understanding and calculating the fundamental limits of flux measurement in Hall probes, we can think some more about the materials we can use to construct Hall probes. The conventional approach has been to use Hall
sensors from semiconductors, but this avenue has been limited by telegraph noise and charge depletion regions, which tend to be more detrimental in small sensors [25].

In the preceding discussion, we saw that in the metallic limit, in which electrons dominate heat conduction, we can get a respectable flux sensitivity on the order of $10^{-6} \Phi_0/\sqrt{Hz}$.

Can we do better? Perhaps, if we go between the extremes of semiconductors and full-blown metals. There is plenty of room in the middle. In that middle we will go, studying the performance of Hall sensors made of bismuth, a semimetal.

As we strive to understand the physics of bismuth Hall sensors, we will retrace some of the steps we took in analyzing the metallic probe and take the opportunity to fill in some details that we glossed over in our effort to understand the basics. In particular, we will look more closely at how to estimate heat flow and Hall probe sensitivity. Ultimately, we will see that in our model, the flux sensitivity of bismuth Hall sensors matches that of the best SQUIDs.

### 3.7.1 Why Bismuth?

One reason for using bismuth instead of semiconductors in Hall probes is to reduce telegraph noise. It is not certain to work since an evaporated film will have more defects on which telegraph noise can develop, but the increased carrier density in bismuth compared to conventional semiconductor probes may help. Perhaps researchers will go in other interesting directions, such as heavily-doped semiconductors or metals\(^5\).

Another reason to try other materials is the difficulty in miniaturizing the semiconductor sensors. When etching the probe pattern in a semiconductor, a depletion region forms that does not conduct. As semiconductor probes get smaller, the depletion layer threatens to pinch off the carriers' path to the active region. Even if this does not occur, the resistance in the device increases, making thermal voltage noise stronger. This was the main justification posed by Sandhu \textit{et al.} [39] for trying Hall probes made of bismuth.

There are other good reasons to use bismuth as a Hall probe material. It has the highest Hall coefficient of the metallic elements. It is an elemental semimetal with a

\(^5\)There may have been some reticence to use materials with higher carrier densities in Hall probes since, if we use the same excitation current, the Hall voltage would drop drastically in metals compared to semiconductors. However, according to our single band analysis and assuming strong electronic heat conduction, low carrier concentration is not essential to flux sensitivity and metals should be a fine choice if we increase the current sufficiently.
low melting point, making it easy to evaporate pure thin films for nanofabrication. It is not expensive or toxic.

It also turns out that bismuth phonons conduct heat well at low temperature, which allows us to tolerate more electrical power dissipation in the active region of the sensor. Heating being one of the somewhat neglected topics in our first detailed look at Hall probe physics, let us return to that now.

### 3.7.2 Heating

In the basic Hall probe discussion of this chapter, we worked out an expression in section 3.6.4 for a Hall sensor’s minimum detectable magnetic field:

\[
B_{\text{min}} = \frac{R_d}{R_H} \sqrt{\frac{P_{\text{th}}}{P_{\text{bias}}}}.
\]

We subsequently expanded the \( R, R_H, P_{\text{th}}, \) and \( P_{\text{bias}} \) terms, massaging the results to get the magnetic flux sensitivity as a function of two concrete and intuitive material properties: Fermi wavelength and mean free path. In obtaining that result, we only addressed briefly the competing concerns of overheating the sensor and using more bias power to improve the signal-to-noise ratio. Let us now flesh out that discussion.

The basic heat flow problem is that the Hall probe excitation current generates thermal energy resistively in the current leads and the active region; this heat must go somewhere. It can flow directly to the cooled substrate or indirectly there by conduction through the leads first. This is depicted in figure 3.11.

![Figure 3.11: Heat flow for the active region (the centre cube) of a Hall probe connected to four leads (dashed lines). Heat can either be conducted through the leads (blue arrows) or go directly to the substrate (red arrow).](image)

The calculation I did to estimate the influences of these two channels is rather detailed, so the interested reader can consult Appendix A.2 for an in-depth discussion.

43
of the problem. The basic result of the calculation is that for Hall probes with thickness close to or greater than the in-plane active region dimensions, both channels are important. It turns out that the calculation mathematically resembles the case of direct conduction only, with a geometry-dependent prefactor. For a Hall probe active region shaped like a cube, dimensions $s \times s \times s$, that prefactor is about 3. For our discussion here, let us take the much simpler case of direct conduction only and insert the factor of 3 by hand.

At thermal equilibrium, the amount of power that can be conducted out of the active region is\(^6\)

$$P_{\text{bias}} = 3 \frac{\kappa A}{\ell} \Delta T,$$

with conductivity $\kappa$, area $A$, length $\ell$, and temperature difference $\Delta T$ along $\ell$. For our $s \times s \times s$ Hall probe, $A = s^2$ and $\ell = s$. Note the factor of 3 put in for conduction through the leads. We can split the conductivity into electron and phonon components:

$$P_{\text{bias}} = 3\left(\kappa_e + \kappa_{ph}\right)\frac{s^2}{s} \Delta T = 3(\kappa_e + \kappa_{ph})s \Delta T.$$

The phonon thermal conductivity can be modelled as that of a gas of phonon particles:

$$\kappa_{ph} = \frac{1}{3} \frac{n_m C_V}{V} v_s \Lambda_{ph},$$

where $n_m$ is the number of moles of material in the probe, $C_V$ is the specific heat per mole at constant volume, $V$ is the volume of the probe, $v_s$ is the speed of sound, and $\Lambda_{ph}$ is the mean free path of the phonons. The Debye model approximates $C_V$ at low temperature:

$$C_V = \frac{12\pi^4}{5} \frac{N k_B}{n_m} \left(\frac{T}{\theta}\right)^3,$$

$$\theta = \frac{\hbar v_s}{k_B} \left(\frac{6\pi^2 N V}{V}\right)^{1/3},$$

where $N$ is the number of bismuth atoms per primitive cell and $V$ is the volume of the Hall probe active region. Thus, the phonon conductivity is

$$\kappa_{ph} = \frac{2\pi^2 k_B^4}{15 \hbar^3 v_s^2 \Lambda_{ph} T^3}.$$

\(^6\)Note that we neglect convective effects. Also, the truly obsessive scientist can immerse the probe in liquid helium to add further cooling through the top of the probe. I will not do that calculation.
The electrical conductivity is related to the electron thermal conductivity by the Wiedemann-Franz law:

\[ \kappa_e = \frac{\pi^2 k_B^2 T}{3 e^2} \sigma = \frac{\pi^2 k_B^2 T}{3 e^2} \frac{1}{\rho} \]  \hspace{1cm} (3.50)

Let us take stock of the state of our model now that we have more detailed treatment of heating concerns. We started with the following expression for minimum detectable field:

\[ B_{\text{min}} = \frac{R_d}{R_H} \sqrt{\frac{P_{\text{th}}}{P_{\text{bias}}}}. \]

We can perform substitutions for \( R_d, R_H, P_{\text{bias}}, \) and \( P_{\text{th}} \) that come from equations (3.29), (3.37), and (3.45), as well as the definition \( P_{\text{th}} = 4k_B T \Delta f \), which gives us

\[ B_{\text{min}} = \frac{h e \Lambda \lambda}{e \Lambda \lambda} \sqrt{\frac{4k_B T \Delta f}{3(\kappa_e + \kappa_{ph}) s \Delta T}}. \] \hspace{1cm} (3.51)

From here, our discussion bifurcates based on whether heat conduction is dominated by electrons or phonons. Generally, metallic Hall probes will have electron-dominated heat conduction while semiconductors will be phonon-dominated. Semimetals would be somewhere in the middle.

### 3.7.3 Metallic Limit

Let us first work out the field sensitivity of good metals, which also corresponds to the case of heavily-doped bismuth. For metals, phonon conductivity is negligible compared to the electron heat conductivity. Then we only need to combine equations (3.37) and (3.50) to calculate the electron heat conductivity:

\[ \kappa_e = \frac{\pi^2 k_B^2 T}{3 e^2} \frac{1}{\rho} = \frac{\pi^2 k_B^2 T e^2 p \Lambda \lambda}{h} = \frac{\pi^2 k_B^2 T p \Lambda \lambda}{h}. \] \hspace{1cm} (3.52)

Recall that for a particle-in-a-box calculation for a Fermi gas, the density of charge carriers is about \( 8/\lambda_F^3 \), so the average separation between carriers is about \( \lambda_F/2 \). If we do a substitution for \( p \), we get

\[ \kappa_e = \frac{8\pi^2 k_B^2 T \Lambda}{3 h \lambda_F^2}. \] \hspace{1cm} (3.53)
Substituting this into equation (3.51) yields a minimum detectable field of

\[ B_{\text{min}} = \frac{2\Phi_0}{\lambda_F \Lambda} \sqrt{\frac{1}{2\pi^2} \frac{h\Delta f}{k_B \Delta T} \frac{s^2 \lambda_F^2}{\Lambda s^2}} = \frac{2\Phi_0}{s\Lambda} \sqrt{\frac{1}{2\pi^2} \frac{h\Delta f}{k_B \Delta T} \Lambda} \]  

(3.54)

We can put in some numbers for an estimate of the minimum detectable field. Suppose we allow a 1 K warming of the probe and measure the Hall voltage with a bandwidth of 1 Hz. A typical Hall probe made of a three-dimensional metal typically has its mean free path limited by the thickness of the probe, so \( s \sim \Lambda \). This gives us a minimum detectable field of

\[ B_{\text{min}} \approx 3.2 \times 10^{-6} \frac{\Phi_0}{\Lambda s}. \]  

(3.55)

From this, we get the flux sensitivity that was stated earlier without proof:

\[ \Phi_{\text{min}} \approx 3.2 \times 10^{-6} \Phi_0. \]  

(3.56)

### 3.7.4 Semimetal Probes

The second case we want to look at is semimetal probes, with a choice of bismuth for a concrete example. Instead of making the usual semiconductor Hall probes, Sandhu et al. used the semimetal bismuth as their Hall sensor material in hopes of avoiding surface charge depletion effects as they miniaturized their probes [39]. The main difference between the semimetal and metal case is that the electron carrier density is relatively low for semimetals, so phonon heat conduction can be important.

Note that semimetals contain both holes and electrons. We will see in Chapter 7 that anything more than one band is not optimal for flux sensitivity. For our estimate of the limit of flux sensitivity, let us assume the optimal one-band case, then.

Recall the electron and phonon heat conductivities are as follows:

\[ \kappa_e = \frac{8\pi^2}{3} \frac{k_B^2}{h\lambda_F^2} \Lambda T, \]

\[ \kappa_{\text{ph}} = \frac{2\pi^2}{15} \frac{k_B^4}{h^3 v_s^2} \Lambda_{\text{ph}} T^3. \]

Also recall that we are assuming \( T = 4 \) K for our comparisons of flux sensitivity. Now apply the assumption that bismuth is the Hall probe material. Approximating the Fermi surface of bismuth as a sphere gives us a Fermi wavelength of 40 nm, corresponding to a carrier density of \( 1.3 \times 10^{23} \) m\(^{-3} \). The speed of sound in bismuth is
1100 m/s. With these numbers, we find that \( \kappa_e = 1.89 \times 10^4 \text{ J K}^{-1} \text{ m}^{-2} \text{ s}^{-1} \Lambda \) and \( \kappa_{ph} = 2.15 \times 10^6 \text{ J K}^{-1} \text{ m}^{-2} \text{ s}^{-1} \Lambda_{ph} \). The phonon mean free path should be comparable to the electron mean free path in a nanoscale device, so the calculation clearly indicates that phonon heat conduction is the dominant mechanism here. Experimentally, this is supported by thermal conductivity measurements, albeit on samples with crystallites on the order of hundreds of microns in size [40]. Let us now neglect the electron term and consider only the phonon term. For the 50 nm Hall probes we want to make, we will be in the regime where the mean free paths for phonons, holes, and electrons are all constrained by the sensor size \( s \), so \( \Lambda_{ph} \approx \Lambda \sim s \). As with the metal Hall probes, assume that we can sustain a \( \Delta T = 1 \text{ K} \) temperature differential due to heating. Under all these conditions, substituting for the phonon heat conductivity into equation (3.51) gives

\[
B_{\text{min}} = \frac{2\Phi_0}{\lambda_F \Lambda} \sqrt{\frac{10h^3 v_s^2 \Delta f}{\pi^2 \Lambda_{ph} s k_B T^2 \Delta T}} = 2.92 \times 10^{-7} \frac{\Phi_0}{s^2}, \tag{3.57}
\]

or a flux sensitivity of

\[
\Phi_{\text{min}} = B_{\text{min}} s^2 = 2.92 \times 10^{-7} \Phi_0. \tag{3.58}
\]

This matches the best SQUIDs I am aware of, with flux sensitivity of \( 3 \times 10^{-7} \Phi_0 \) [16], where \( \Phi_0 \) is the two-electron flux quantum.

The above analysis comes with two caveats. First, phonons from the bismuth must couple efficiently with the substrate to get the maximum flux sensitivity calculated. Phonons from soft bismuth metal may not couple well to lattice vibrations in typical hard silicon substrates. Second, the calculation assumes one-band bismuth while using the Fermi wavelength for bismuth that has not been doped to one band. \( \lambda_F \) should be somewhat smaller because the other carriers need to be removed by doping to make one-band bismuth.

### 3.8 A Fundamental Link?

We have now seen that Hall-effect probes and MFM tips can be as good as SQUIDs for magnetic flux detection, all with possible flux sensitivities of \( 3 \times 10^{-7} \Phi_0 \) at 4 K, though the theoretical estimates should be regarded as accurate only within an order-of-magnitude. Still, the fact that they are of the same order of magnitude, combined
with the knowledge of the detailed physics of these sensors, leads me to reiterate a conjecture made in Chapter 2. A case could be made that optimized Hall probes and SQUIDs have the same mechanism of operation: flux detection due to electrons picking up an Ehrenberg-Siday-Aharonov-Bohm (ESAB) phase\(^7\).

In addition to the values of flux sensitivity being very close, we also see that flux detection quite naturally emerges from the detailed theoretical models for SQUIDs and Hall probes. The flux sensor interpretation is embedded in SQUID theory since it relies on ESAB phase shifts to generate the signal. The flux sensor interpretation of Hall probe theory was not immediately obvious, but when rewritten in terms of Fermi wavelength and mean free path, we saw an associated effective sensor area that allowed us to easily convert field sensitivity to flux sensitivity.

The flux sensitivity similarities between techniques may also just be coincidental, but it is at least interesting to wonder if there is a deeper link.

\(^7\)MFM also has a similar estimated flux sensitivity, but I do not see a mechanism in common with SQUIDs and Hall probes.
Chapter 4

Scanning Probe Microscopy

“The scanning tunneling microscope is completely new, and we have so far seen only the beginning of its development. It is, however, clear that entirely new fields are opening up for the study of the structure of matter. Binnig’s and Rohrer’s great achievement is that, starting from earlier work and ideas, they have succeeded in mastering the enormous experimental difficulties involved in building an instrument of the precision and stability required.”


We have spent most of the last few chapters discussing in detail the physics of magnetic flux sensors of various flavours. We now consider how to optimize the scanning microscopy part of the project. The material here should be applicable in general to the wider field of scanning microscopy.

Scanning Hall probe microscopy owes a great deal to the work of Gerd Binnig and Heinrich Rohrer, who invented the scanning tunneling microscope (STM). In fact, the invention of the STM was the birth of the whole field of scanning microscopy, which now consists of a diverse range of techniques, spanning mechanisms of operation such as electrostatic force, capacitance, spin polarization, heat energy, and near-field optics. MFM, scanning SQUID microscopy, and SHPM are all children of STM. Some Hall probe microscopes even use STM as a way to perform scans at constant height above the sample [42–44].

Scanning tunnelling microscopy and its technological offspring rely on a microscope to carefully raster the sensor over a surface to be imaged. For SHPM, Hall
voltage data are taken from the sensor as the scan occurs, allowing us to make a
detailed map of the magnetic fields on that surface.

In this chapter, I will introduce the scanning Hall probe microscope built and
enhanced for this project, with an emphasis on points that are of general interest to
the field of scanning probe microscopy. Then I will discuss the definition of spatial
resolution and how scan height affects the ability to resolve small features.

4.1 Scanning Hall Probe Microscope

We use a specially-built microscope to move the Hall probe and magnetic sample
translationally and rotationally in a carefully controlled way over a wide range of
temperatures, from room temperature to 4 K. Nigel David designed and constructed
the microscope for his Master’s thesis [36] with assistance from undergraduate stu-
dents Taras Chouinard and Adam Schneider, all members of the Broun Lab at SFU.

4.1.1 Physical Design

A model of SFU’s scanning Hall probe microscope is shown in figure 4.1. Its coarse
motion stage is composed of Attocube stick-slip positioners, which are capable of
displacing the sample stage in steps from the millimetre scale down to hundreds of
nanometres. Unlike low temperature piezo motion, stick-slip motion is a stochas-
tic process, so T. Chouinard designed and built capacitive sensors to measure the
displacements of the Attocubes.

The fine motion stage, shown in figures 4.2 and 4.3, uses piezoelectric strips for
translational motion, echoing a design created by Siegel et al. [45]. Siegel and co-
workers’ achievement was to create a scanner that has better range of motion than
conventional tube scanners [46] while preserving the tube scanner’s most desirable
traits of simple design and small size.

The piezoelectric strips can translate the Hall probe from hundreds of microns
down to the sub-nanometre scale. $xy$-motion is achieved using twin piezo strips along
each axis. Each pair of strips bends into two S-shapes to translate the Hall probe.
$z$-motion is achieved using a pair of piezo strips that bulge out under voltage stimulus
and can flex from 14 $\mu m$ down to the sub-nanometre scale.

There is a further rotational stage that with a set of double gimbals which can be
turned using attached piezoelectric stacks. The double gimbals allow for about $\pm 5^\circ$
of rotation in two orthogonal directions and have been upgraded by T. Chouinard for automated, piezo-mediated, *in situ* adjustments of the angle.

The sample stage, as its name suggests, carries the sample, but it also has a small resistive heating element, a temperature sensor, and a sensitive capacitive touch-down sensor. The microscope is usually cooled to the boiling point of a cryogenic liquid, and the heating element allows us to bring the temperature above that boiling point. The touch-down transducer is composed of two parallel plates that come closer together when the Hall probe touches down on the sample, which is mounted on one of the plates. We use an Andeen Hagerling 2550A capacitance bridge in order to measure plate deflections as small as 2 nm. The $z$-motion piezos are calibrated using the touch-down sensor.

As mentioned earlier, other groups use an STM tip adjacent to the Hall sensor to keep the sensor at a constant scan height, but this is only useful if the sample
Figure 4.2: Fine motion stage (left) shown with a view of how the piezo benders change shape in response to applied voltage (right). The piezoelectric S-benders have a large lateral range of motion, but also move the probe a small amount vertically when making lateral movements. The range of motion is shaped like a paraboloid. Figure courtesy of N. David [36].

Figure 4.3: z-benders used on the fine motion stage. The benders are segmented so that the middle piezo segments spread apart, like puckered lips. An “anchor chip” is attached to the probe mount using cyanoacrylate, or “super glue,” for a strong bond. The Hall probe is glued to the base chip with GE varnish so that the probe can be removed independently of the base chip using ethanol. Figure courtesy of N. David [36].
Figure 4.4: Views of the fine motion stage from the side (left image) and bottom (right image). A Hall probe with carrier chip can be seen in the middle of the right image. A nickel is used for scale.

Figure 4.5: Close-up of the Hall probe carrier chip visible in the previous figure. The carrier chip allows for handling and soldering without applying direct mechanical or thermal stress to the Hall probe or its electrical contacts.
being scanned is conductive. The capacitive touch-down sensor performs a similar
function for the SFU Hall probe microscope, telling the user what control voltages
can be applied to the fine motion stage to not touch down. While the capacitive
touch-down sensor does not give constant feedback on the sensor height, it is useful
in both conductive and non-conductive samples, unlike the STM tip.

A superconducting solenoid producing a field of 66 gauss per ampere surrounds the
coarse motion, fine motion, and sample stages. This is useful for producing vortices
for some Hall probe imaging experiments. The magnet is wrapped around a vacuum
can to allow one to seal the microscope and sample.

4.1.2 LabVIEW Software

In order to control this hardware, A. Schneider wrote LabVIEW software to control
all aspects of the microscope, from regulating electrical signals going into the piezos,
attocubes, Hall probe, and magnet to writing algorithms for performing a polynomial
fit to the magnetic surface and making a map of magnetic field for the sample. It is
very important for the program to do all this while not crashing or otherwise damaging
the Hall probe, which requires a great deal of time and resources to fabricate and
characterize. A good program errs on the side of caution with regards to Hall probe
motion.

The basic algorithm for the LabVIEW program is this:

1. Use Attocubes to coarsely move the sample up and use z piezo benders to finely
   move the Hall probe down, checking for gentle contact between the Hall sensor
   and sample by checking for increases in the capacitance of the capacitor under
   the sample.

2. Once the Hall sensor is close enough to touch down on the sample using only
   the piezoelectric benders, touch down on a grid of widely-dispersed points in the
   scan area and fit a 2D curve to the touchdown points, finding the approximate
   topology of the sample surface.

3. Use the 2D curve and fine motion piezos to raster the Hall probe at some user-
defined height above the sample surface while taking measurements of Hall
   voltage using a Stanford Research Systems SR810 Lock-In Amplifier.

4. Write data to a file.
T. Chouinard conceived of and programmed the z bender motion in step 1.

T. Chouinard and I improved the touchdown algorithm to make it less prone to positioning errors due to hysteresis. This was done by increasing the number of touchdown points and throwing away the first point in each row of touchdown points, which would be affected most by hysteresis.

I fixed a few software bugs, described below:

- **Random capacitance bridge readout failures.** On rare occasions, the Hall probe microscope software will crash due to a failure of the relevant LabVIEW program to retrieve capacitance measurements from the capacitance bridge. The crash problem was fixed by inserting a loop to repeatedly query the capacitance bridge until a measurement is properly collected.

- **Improper program abort.** When a scan is aborted using the built-in abort button, it was possible for the Hall probe to crash into the sample.

## 4.2 Spatial Resolution

When using the Hall probe microscope program to do a simple scan, it is useful to know the spatial resolution of the probe in order to avoid spacing the scan points so closely that they overlap in the effective area measured.

Instinctively, one would think that spatial resolution is the effective size of the probe. We would not be able to see anything smaller than what we use to look at a sample. Height above the sample is also a factor, though, since flux signals start to blend together and fade as height increases. Sufficiently far from two dipoles, it will look to the sensor like there is just one source of magnetic flux lines. For example, when we have a perfect 1 µm square Hall probe at a height of 1 µm scanning two magnetic dipoles separated by 1 µm, there is only one peak in a plot of flux as a function of horizontal probe position. There is no indication that there are two dipoles. It is necessary to refine our instinctive definition of spatial resolution.

### 4.2.1 Definition

I define spatial resolution in the context of a model system of two perfect dipoles at distance $a$ apart as shown in figure 4.6 scanned by a noiseless, square Hall probe of size $s \times s$ at some height $z$ above the dipoles. When the Hall probe is small enough and low enough, it will be able to resolve the two dipoles. While this is a rather
simple model, it should give a rough idea of how resolution scales with scan height and probe size. A more complete model would consider the effect of flux noise on resolution.

![Diagram of model system](image)

Figure 4.6: The model system for determining spatial resolution of two perfect dipoles separated by distance $a$ and both pointing in the $+z$ direction as a noiseless, square, and infinitesimally thin Hall probe with sides of length $s$ and with plane perpendicular to the $z$ direction scans at a height $z$ above the spins. The probe is scanning in the $+x$ direction with the active region perfectly centred over the dipoles in the $y$ direction.

We will only consider probes with $s < a$. If this condition is met, a scan will usually reveal two peaks that correspond to the two dipoles\(^1\). We see in figure 4.7 that the peaks meld into one as the scan height is increased and we lose the ability to resolve the two peaks. In contrast, if $s > a$, we may get spurious peaks, including a peak directly in the middle of the two dipoles. This is a spatial aliasing effect from parts of the probe being over both spins at the same time.

Let us define two dipoles with flux peaks of size $\varphi$ as resolved if the minimum of the flux at midway $\varphi_{\text{mid}}$ between the peaks is no higher than $\varphi/\sqrt{2}$, as shown in figure 4.8. This threshold is somewhat arbitrary, but remember we are only seeking a rough idea of the scaling of the spatial resolution, and this will suffice for that purpose.

For one dipole, the flux detected by the probe is

$$\iint \mathbf{B} \cdot \mathbf{da} = \iint \frac{\mu_0 m}{4\pi r^5} \left(3z^2 - r^2\right) \, dx \, dy. \quad (4.1)$$

The analytic expression for this integral is ultimately not enlightening, so I have included the full expression in Appendix E along with the Maple code used to do this calculation of spatial resolution.

\(^1\)For very low scan heights, specifically when the scan height is much less than the probe dimension, extra peaks are seen that arise from the high inhomogeneity of the field close to the dipoles. For regular use, this is not a concern.
Figure 4.7: A simulation of the magnetic flux through a Hall probe for the scenario shown in figure 4.6. The unit of length for position and height is the dipole separation $a$.

Figure 4.8: An illustration of how the resolving criteria are defined. If $\varphi / \sqrt{2} > \varphi_{\text{mid}}$, then the two dipoles are considered resolved.
In this formulation of spatial resolution, the height above the sample is a more important consideration than probe size, with resolution scaling as \(\sqrt{s^2/1.01^2 + z^2/0.81^2}\), as shown in figure 4.9. However, this changes as I change the constraint on the residual flux. For example, when I consider scans with residual flux greater than \(\varphi/\sqrt{1.3}\) to be not spatially resolved, then we see in figure 4.10 that resolution roughly scales as \(\sqrt{s^2 + z^2}\).

4.2.2 Angle and Height Considerations

In the above discussion, we assumed that the Hall sensor is perfectly parallel to the sample surface. In reality, this is impossible, though we should be able to get within a degree or better. In sensors where a conductor is deposited on the substrate’s surface, it is desirable to have a slight tilt to the probe so that the substrate will take the brunt of collisions rather than the probe and its leads, as shown in figure 4.11. However, this tilt can lead to spatial resolution problems if the Hall device is not designed properly.

For a given tilt angle, the further the Hall cross is from the point of substrate-sample contact, the greater the height of the active region. Placing the Hall cross close to the substrate’s edge helps place the sensor closer to the sample. It would be reasonable to expect the substrate to be polished to a sharp corner so that the Hall probe is 1 \(\mu\)m from the active region of the sensor. This is about as close as you can go using a light microscope to quickly see how much more polishing is required.

Assuming a 1\(^\circ\) tilt and a 1 \(\mu\)m distance from the active region to the edge of the chip, the height of the active region above the surface is

\[
z = d \sin \theta = (1 \, \mu\text{m}) \sin 1^\circ = 17.5 \, \text{nm.}\quad (4.2)
\]

With this extra height, the maximum percent change in resolution is about +24\% for a 50 nm probe. The actual percentage change in resolution varies with the pre-tilt scan height of the probe, as shown in figure 4.12.
Figure 4.9: Plot of height vs. probe width for the resolution limit using the $\phi/\sqrt{2}$ definition of resolution. Both axes are in the same arbitrary units. Probes in the parameter space under the curve can detect magnetic flux with spatial resolution of 1 arbitrary space unit. The simulated points fit to the ellipse $(s/1.01)^2 + (z/0.81)^2 = 1$.

Figure 4.10: Plot of height vs. probe width for the resolution limit using the $\phi/\sqrt{1.3}$ definition of resolution. Simulated points roughly fit the ellipse $(s/1.04)^2 + (z/1.01)^2 = 1$. 

59
Figure 4.11: Illustration of probe tilt geometry.

Figure 4.12: This graph shows the percentage change in spatial resolution that results from tilting a Hall probe 1° away from flat, as a function of scan height, for a selection of Hall probe sizes. It assumes the spatial resolution scales like $\sqrt{(s/1.01)^2 + (z/0.81)^2}$ and the active region is 1 μm from the edge of the chip.
Chapter 5

Hall Probe Design

“Design is a funny word. Some people think design means how it looks. But of course, if you dig deeper, it’s really how it works.”

— Steven Paul Jobs [47].

We have seen that Hall-effect probes have a great deal of potential. Calculations indicate that they can be made with a flux sensitivity comparable to the best flux sensors, with the possibility of supplanting even SQUIDs in some applications due to Hall probes’ superior size scalability. Of particular interest to me and the Broun Lab is making nanoscale Hall sensors for fundamental research purposes, such as studying magnetic features in superconductors.

This chapter will discuss the design of the Hall probes I attempted to make for this thesis. For the conventional design, I briefly talk about the different parts and their purpose. While conventional probes are simple and easy to make, the conventional design has two major weaknesses:

- fragility — bare thin bismuth film is easily damaged by dust particles and other debris that may be on the surface you wish to scan, and even a protective film of germanium tends to get scratched off;
- scan height limits — to avoid damage, the probe must be rastered over the surface at some finite height, which limits spatial resolution.

We will study a novel Hall probe design that I invented, and this novel design solves the above problems using two approaches. First, the active region is thickened so that it can withstand some damage. Second, the orientation of the active region is changed so that the substrate provides more protection during scans. We will look at
the theoretical performance of several variations on this Hall probe configuration for this device, which is considerably more complicated to make than the conventional probes.

5.1 Conventional Probes

The conventional probe design is a 2D planar device shown in figure 5.2. I will explain design choices I made in this section, including placement of the active region, overall shape, decisions regarding the contact pads.

Let us start with the active region. It is close to the edge of the pattern, which, in turn, would be placed on the corner of a square silicon substrate. Assuming the pattern is flush with the edges of the chip, the active region would be one micron from the corner, which we saw in section 4.2 was an important factor to determining scan height and spatial resolution. If there is a gap between the pattern and the edges of the chip, the substrate can be polished to remove the gap, though this is difficult in practice. In principle, the resistance through the outermost pair of Hall probe leads can be monitored to ensure the active region is not polished away.
Figure 5.2: Hall probe design. Going from the bottom left image to the top right image, we zoom in progressively more on the active region of the probe. Length scales are in microns. The size of the active region in this picture is 50 nm. The finer details around the active region (blue) are to be written with high resolution EBL settings while the leads and pads (green) are to be written with high speed settings.
The leads splay out close to the active region to reduce the leads’ resistance, which causes thermal noise that can affect Hall probe performance.

The bismuth contact areas are on the same side of the substrate as the active region. This limits the scan range in one direction, though, as figure 5.3 shows. In any case, secure attachment of current and voltage wires to the pads is obviously necessary. Two standard methods of attaching wires to the bismuth probe, soldering and wire bonding, did not work. When I tried soldering, the solder failed to wet to the bismuth, even with flux. The film also tended to come off, though an adhesion layer may help. Someone else tried wire bonding, and that did not work either.

Figure 5.3: The Hall sensor has complete freedom of movement up, down, into the page, and out of the page, but the left-right range of motion is limited by the placement of the wires connecting to the probe.

Instead, I pressed small pieces of indium onto the bismuth pads and used those to connect the voltage and current wires (see Appendix F.5 for more details). Pressed indium contacts offer a quick and easy way to attach macroscopic wire leads as needed and were secure enough for tests and the Hall probe scanning. That means the pads must be spread far enough apart that human hands can wire up the contacts reliably without overlaps, short circuits, or hitting the active region. A separation of about 1 mm worked well.

The last design element to consider is a protective coating. Scanning probes generally raster very close to the surfaces they image, and surface impact is an omnipresent threat. The protruding edge of the probe’s substrate offers meagre and profoundly unsatisfactory protection against large particles or even significant bumps on the surface to be scanned. The soft underbelly of the Hall probe’s active region needs better armour.

In the latest probes I used for scanning microscopy, I used evaporated thin films of germanium as a protective coating. Pure germanium is a semiconductor with low electrical conductivity, so it does not interfere with the Hall probe’s sensing of magnetic flux. It is a hard substance, certainly much harder than bismuth, which helps it survive surface impacts. Germanium deposition via evaporation allows for excellent and easy control of the layer thickness. The main downside is its high evaporation
temperature, which may cause annealing and a slight doping of the bismuth. Future projects may explore alternative protection measures.

Table 5.1: Hardness of bismuth and germanium [48].

<table>
<thead>
<tr>
<th>Substance</th>
<th>Mohs Hardness</th>
<th>Microhardness (MN/m²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bi</td>
<td>2.25</td>
<td>177</td>
</tr>
<tr>
<td>Ge</td>
<td>6.0</td>
<td>5400</td>
</tr>
</tbody>
</table>

5.2 The Novel Design

Figure 5.5 shows the novel Hall probe design. Essentially, it is a conventional Hall probe with the active region flipped up 90°. The basic structure is a trilayer of conducting leads, an insulating spacer, and conducting leads. The leads join at a junction formed by an active sensing material, a semiconductor or semimetal. The materials I chose for my particular novel Hall probes were aluminum for the leads, silicon dioxide for the insulator, and bismuth for the active region. With this design, the thickness of the active region is easily scalable, while putting the substrate to much better use protecting the probe.

This is an unusual geometry for a Hall probe, but not unusual for hard drive tunnel magnetoresistance (TMR) sensors. In fact, the idea for the novel Hall probe
Figure 5.5: This boomerang-shaped novel Hall probe design detects magnetic flux going into the page. Light grey: bismuth active region. Light blue: aluminum leads. Dark blue: silicon dioxide insulating spacer. Yellow: gold contacts connecting to aluminum leads. Relative size of the active region is exaggerated for clarity. The active region thickness dimension is into the page.

design was inspired by hard drive sensors. To see how the novel design helps make Hall probes more robust, let us get a little bit of insight into the robustness of the well-established hard drive TMR read heads.

The feats achieved by hard drives and their designers are stunning. In daily operation, the read head flies nanometres above a drive platter spinning at thousands of RPM and must be able to withstand a few collisions with that platter. If the thickness of a packaged hard drive read head is on the order of a millimetre, the nanometre-scale flight height of a hard drive sensor would be analogous to a Boeing 747 jet flying tens of microns above the ground, and a 747 would certainly not survive ground contact at flight speeds.

So how were TMR sensors made more robust to withstand the rigors of long-term use in hard drives? Among other things, the devices were made larger than they needed to be, which helps their robustness to damage. TMR sensors are structured as two ferromagnetic thin films separated by a very thin insulating tunnel barrier.
While the thickness and width of the films are constrained by the arrangement of bits on the hard drive, length is not similarly constrained. When damage occurs along the length dimension, there is plenty more undamaged sensor to compensate for the damaged section. It is like having a backup sensor. Indeed, there is evidence that TMR sensors are capable of limited self-repair. Wallash et al. found that when a TMR read head is repeatedly swept over a particle embedded in a disk, the read head’s resistance initially dips almost to 0 in a likely damage-related short circuit, but partially recovers over successive sweeps, perhaps because the damaged section is polished away [49].

This approach can also be adapted to Hall probes in what can be thought of as backup active regions. In practice, this would simply mean a thicker active region, a section of which could be damaged, yet be part of a viable Hall probe. This is an idea that could be implemented on conventional probes, though fabrication of thick conventional probes via lift-off becomes difficult for small probes because lift-off becomes difficult as the thickness-to-width aspect ratio increases. Fabrication by etching may be troubled by undercut issues. Even if such a thick, small conventional probe were made, the tall, thin structures could easily be pushed down by debris, possibly causing short circuits.

The novel Hall probe design, among other things, makes a thick active region much more scalable in fabrication and also better-supported structurally. The novel design achieves this essentially by flipping over the active region 90°, rearranging the current and voltage leads slightly, and adding an insulator between the leads. The change in geometry is shown in figure 5.6. With the novel design, the insulator prevents the leads from being pushed together, and increasing the active region thickness corresponds to an easy change of dimensions laterally across the substrate.

Of course, another way to improve Hall probe robustness is to put sturdy structures between the probe and any harmful debris. The novel design accomplishes this by placing the substrate in front of the probe when in scanning position, as shown in figure 5.7. With the substrate present to protect the sensor, we can now scan much closer to the magnetic surface, where we would get higher spatial resolution and capture more of the magnetic flux. As an added advantage, wiring no longer constrains the range of motion, as it did with the conventional design.
Figure 5.6: A conceptual transformation of a conventional Hall probe to the novel design, going through intermediate steps of a “thick” conventional probe and a rotated thick probe. The active region is highlighted in grey, the voltage leads are labelled, and the axis of magnetic field detection is shown for all designs.

5.3 Simulations

Before trying to make this novel design, I wanted to have some assurances that it would perform similarly to or better than conventional designs. I used COMSOL Multiphysics 4.2 software to perform a series of 2D finite-element simulations of various Hall probe designs to compare their Hall voltage responses. This is mainly to ensure that Hall probes in the novel geometry perform similarly to Hall crosses of conventional geometries, thickness questions aside. In our discussion here, we will first derive the differential equation that will be used to simulate the Hall probe behaviour. Then we will discuss special considerations for the simulations to ensure the comparisons are fair. Finally, we will look at the results of the simulations themselves.
Figure 5.7: Scanning orientation of a novel Hall probe, in blue. The active region is at the bottom, facing downwards to detect flux from the green spins. The grey substrate protects the probe from the ominous black object.

5.3.1 Hall Probe Differential Equation

COMSOL takes input of a differential equation and boundary conditions with adjustable parameters for the conductivity and geometry, then outputs a map of electric potential. When interpreting the results, though, we should keep in mind that the simulations are assuming classical electrodynamics in uniform media. Quantum effects and carrier mean free path do not figure into these calculations. However, for devices that are larger than both \( \Lambda \) and \( \lambda_F \), the classical simulation should be quite adequate.

Let us derive the differential equation first. Consider the equation of continuity for electrostatic charge flow in the steady state:

\[
\nabla \cdot \mathbf{J} = -\frac{\partial \rho_q}{\partial t} = 0,
\]

where \( \mathbf{J} \) is the current density, \( \rho_q \) is the charge density, and \( t \) is time. We assume charge density does not change in the steady state, so the derivative of \( \rho_q \) is 0. Now consider also Ohm’s law and the calculation of electric field in terms of potentials:

\[
\mathbf{J} = \sigma \cdot \mathbf{E}, \quad (5.2)
\]

\[
\mathbf{E} = -\nabla V - \frac{\partial \mathbf{A}}{\partial t}, \quad (5.3)
\]

69
where $\sigma$ is the low-field magnetoconductivity tensor, $V$ is the scalar electric potential or voltage and $A$ is the vector potential. Again, since we are in the steady state, we can neglect the derivative with respect to time. With these substitutions into equation (5.1), we get

$$- \nabla \cdot (\sigma \cdot \nabla V) = 0. \quad (5.4)$$

Next, we must consider the boundary conditions for this differential equation. Obviously, current should not be able to travel outside of the Hall probe structure. We must also ensure that the current-carrying leads conduct the same amount of charge through each probe. To do this, we must come up with a way to get current in terms of the voltage, which is what COMSOL actually calculates. We start with the definition of electric potential, Ohm’s law, and the definition of current density:

$$\mathbf{E} = -\nabla V, \quad (5.5)$$
$$\mathbf{J} = \sigma \cdot \mathbf{E}, \quad (5.6)$$
$$I = \int \mathbf{J} \cdot d\mathbf{A}. \quad (5.7)$$

Then we substitute the first equation into the second, and the second into the third. Assuming current is in the $x$-direction and thickness is measured along the $z$-axis, we get:

$$J_x = \sigma_{xx}(-\partial_x V) + \sigma_{xy}(-\partial_y V), \quad (5.8)$$
$$I = d \int J \, d\ell = -d \int (\sigma_{xx}\partial_x V + \sigma_{xy}\partial_y V) \, dy, \quad (5.9)$$

where $d$ is the thickness of the Hall probe.

### 5.3.2 Simulation Details

The geometries tested, as seen in figures 5.8 and 5.9, span some common conventional Hall probe shapes, in addition to the novel design. I tested several different configurations of the novel design, using parameters appropriate for gold leads\(^1\) and bismuth active regions. Some of the shapes include an anticipated fabrication flaw that would occur due to difficulty in making the novel design’s active region. In most cases, these simulations were done prior to my ultimately settling on aluminum for the leads, but the difference is irrelevant for Hall probe sensitivity. The main point is that bismuth conducts poorly and the leads connecting up to the bismuth device should conduct well. For the sake of choosing a metal, I input parameters that roughly match gold.

\(^1\)These simulations were done prior to my ultimately settling on aluminum for the leads, but the difference is irrelevant for Hall probe sensitivity. The main point is that bismuth conducts poorly and the leads connecting up to the bismuth device should conduct well. For the sake of choosing a metal, I input parameters that roughly match gold.
cases, the Hall probe active region is 50 nm × 50 nm, except for the one that is 50 nm × 100 nm. Though the simulations were in 2D, I assumed an active region thickness of 50 nm for all designs to get Hall voltages in the right units. For the novel designs, the “thick” leads are 50 nm thick while the “thin” leads are 10 nm thick.

The modelling of gold and bismuth materials in the finite element simulations was quite primitive compared to the models of bismuth and lead we will use when we study doped bismuth for use in Hall probes. I used a two-band model appropriate to bismuth at room temperature and a single-electron-band model of the gold. The specific parameters I used are in table 5.2. The specific forms of the conductivities used are shown below:

\[
\sigma = \begin{bmatrix}
\sum_i \frac{e n_i \mu_i}{1 + \mu_i^2 B^2} & \sum_i \frac{e n_i \mu_i^2 \text{sgn}(q_i)}{1 + \mu_i^2 B^2} \\
- \sum_i \frac{e n_i \mu_i^2 \text{sgn}(q_i)}{1 + \mu_i^2 B^2} & \sum_i \frac{e n_i \mu_i}{1 + \mu_i^2 B^2}
\end{bmatrix}, \quad (5.10)
\]

where \(i\) is the band index, \(e\) is the elementary charge, \(\mu\) is the mobility, \(n\) is the carrier concentration, \(\text{sgn}(q_i)\) is the sign of the charge (positive or negative), and \(B\) is the magnetic field.

Table 5.2: Table of parameters used in finite element simulations. The bismuth parameters came from a paper by Kochowski and Opilski [50] for their 50 nm undoped films at room temperature. Their model assumed 2 bands. The gold parameters came from [51]. I used the bulk values. It appears I made a small calculation error in converting from gold Hall coefficient to carrier concentration. This should not materially affect the results, though. The actual carrier concentration should be about \(8.7 \times 10^{28} \, \text{m}^{-3}\).

<table>
<thead>
<tr>
<th>Material and Band</th>
<th>Carrier Concentration (m(^{-3}))</th>
<th>Mobility (m(^2)/V · s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bismuth holes</td>
<td>3 × 10(^{25})</td>
<td>0.04</td>
</tr>
<tr>
<td>Bismuth electrons</td>
<td>1.3 × 10(^{24})</td>
<td>0.12</td>
</tr>
<tr>
<td>Gold electrons</td>
<td>1.01 × 10(^{29})</td>
<td>0.0028</td>
</tr>
</tbody>
</table>

Accurate and precise material models are not necessary here, though, and are beside the point. We are mainly concerned about whether the novel geometry is good for Hall probes and for those purposes, it would be just as useful to input parameters for a “good metal” and a “poor metal” rather than gold and bismuth, so long as we use the same parameters across all simulations.

The voltage at a particular lead was evaluated as average voltage across the edge of that lead. The geometries were discretized to a mesh with maximum feature size of
0.5 nm. The current flowing through each probe was 30 $\mu$A. Limiting the double-area probe to 30 $\mu$A like the others may seem unfair, but this more-or-less compensates for the extra flux that would be going through the extra-large probe.

The results of the simulations are shown in table 5.3.

Table 5.3: Hall voltage simulation results. Larger Hall voltage is better.

<table>
<thead>
<tr>
<th>Probe Geometry</th>
<th>Hall Voltage ($\mu$V)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conventional Bi Hall cross</td>
<td>60.9</td>
</tr>
<tr>
<td>Bi clover leaf cross</td>
<td>59.2</td>
</tr>
<tr>
<td>Novel Bi probe</td>
<td>61.0</td>
</tr>
<tr>
<td>Novel Bi probe, double-area active region</td>
<td>60.9</td>
</tr>
<tr>
<td>Novel Bi probe, Au bottom</td>
<td>29.8</td>
</tr>
<tr>
<td>Novel Bi probe, Au contacts</td>
<td>34.7</td>
</tr>
<tr>
<td>Novel Bi probe, thin Au contacts</td>
<td>57.9</td>
</tr>
<tr>
<td>Novel Bi probe, thin Au contacts, 50 nm defect</td>
<td>55.1</td>
</tr>
<tr>
<td>Novel Bi probe, thin Au contacts, 40 nm defects</td>
<td>52.5</td>
</tr>
<tr>
<td>Novel Bi probe, thin Au contacts, 50 nm defects</td>
<td>52.5</td>
</tr>
</tbody>
</table>

5.3.3 Conventional Designs

Conventional Hall cross and the clover leaf cross are mainly here to set a benchmark level of performance. Being new to COMSOL, I also thought it would make a good introductory exercise to see if the clover leaf cross gave a significantly different Hall voltage response compared to the straight Hall cross. The clover leaf is preferred because splaying out the leads reduces thermal noise. Simulations indicate that the performance difference is very small, if there is one at all, with Hall voltages of 60.9 and 59.2 $\mu$V for the straight and clover leaf designs, respectively.
5.3.4 Novel Designs

The results for the novel designs reveal an interesting pattern. The designs with thick gold layers did not perform as well as the ones without thick gold layers (29.8 and 34.7 $\mu$V vs. 61.0, 60.9, and 57.9 $\mu$V). This appears to be due to the gold layers shorting out the Hall effect, particularly in the case of the design with the continuous gold bottom layer.

Reassuringly, the novel designs perform very similarly to the conventional geometries. The all-bismuth novel designs match the sensitivity of all-bismuth conventional designs. Fabricating all-bismuth sensors was not convenient, though, because the 4D LABS cleanroom did not deposit bismuth in its deposition chambers. Given that the goal was proof-of-concept, I worked on fabricating something more like the design with thin gold contacts.

5.3.5 Novel Design with Fabrication Flaws

Another interesting thing to check is the sensitivity of the Hall voltage to fabrication flaws for the novel probe design. The details of how these flaws may form will be discussed in section 6.2.4. For now, let this brief sketch of the procedure suffice. To make a bismuth active region, I plan to mill out a trench in a fabricated Hall
Figure 5.9: These are the novel probe designs I simulated in COMSOL. The grey and gold indicate that I simulated those areas with approximate models of room temperature bismuth and gold, respectively.
probe boomerang and fill it with bismuth using a lift-off procedure. I wanted to make sure that milling too deep and/or overfilling the trench with bismuth would not ruin Hall probe performance, so I simulated a few novel probe geometries with geometric defects, as shown at the bottom of figure 5.9.

Fortunately, the simulations indicate that defects of this nature have only small effects, if any, producing Hall voltages of 55.1, 52.5, and 52.5 µV compared to 57.9 µV of the novel probe with thin gold contacts and no defects.

5.4 Thickness and Hall Probe Sensitivity

To be clear, adopting this novel design is not entirely beneficial. Thickening the active region of the probe may move a significant fraction of the active region volume much further away from the magnetic feature. This is important because for applications of nanoscale Hall probes, the sources of magnetic field are small or point-like — magnetic dipoles and the ends of flux lines in superconductors, for instance — and
so the field will drop off rapidly as a function of distance. As a result, the volume of active region interacting with the field may be considerably smaller than the physical volume. One of our tasks here is then to quantify how Hall probe sensitivity scales with thickness of the probe’s active region, in particular by looking at an expression like equation (3.31)

\[ B_{\text{min}} = \frac{d\sqrt{4k_B T R \Delta f}}{I R_H}. \]

\( R, d, I, \) and \( R_H \) are the important parameters here. Everything else is independent of Hall probe dimensions.

The first thing to do is define a solvable problem that is a reasonable approximation of the situation. Given the novel probe geometry in figure 5.11, let us define an interaction length \( \ell \) over which the magnetic field contributes to the Hall voltage by, say, shifting some flowing hole carriers in the \( x \)-direction\(^2\). Net positive and negative charges build up on the \(+x\) and \(-x\) edges of the active region, respectively. Let us assume only charges in the bottom slice of the active region, thickness \( \ell \), contribute to the Hall voltage signal. The full thickness of the active region is \( d \).

![Figure 5.11: Sections of the novel Hall probe that are closer to magnetic features will have larger contribution to the Hall voltage. Dividing up the active region into discrete slices, some with strong field interactions (green) and others with no effective field interactions (white), is conceptually useful for the analysis. Leads are in blue. The substrate is grey.](image)

Next, let us assume the electrodes connecting to the active region are perfect conductors and, to first approximation, spread the net charges evenly along the height

\(^2\)This argument works generally for electron carriers or multi-band conductors as well.
of the electrodes. The Hall voltage in such a geometry should scale roughly as the voltage between a pair of parallel plate capacitors that have had their area increased by a factor of $N = d/\ell$. Effectively, this scales down the Hall coefficient $R_H$ by $1/N$. $N$ will be a convenient scaling parameter throughout our discussion.

Let us tackle the scaling of $R$ next. We can split the active region into $N$ slices of thickness $\ell$, all with resistance $R$ and connected in parallel. The effective resistance of such an arrangement scales as $1/N$.

Finally, let us consider the limits of current flow through the novel probe. Each slice of active region is attached to an identically-sized surface area of substrate that can conduct away the same amount of electrical heat. Thus, each slice will be able to support the same maximum bias current through its branch, so the current bias should scale as $N$.

Putting all this together, with the fact that $d$ goes like $N$, of course, we get

$$B_{\text{min}} \propto \frac{N \sqrt{1/N}}{N(1/N)} = \sqrt{N} \propto \sqrt{d}. \quad (5.11)$$

The flux scales similarly.
Chapter 6

Hall Probe Fabrication

“Any customer can have a car painted any color that [she or] he wants so long as it is black.”

— Henry Ford.

In Chapter 5, we studied designs for both the conventional planar Hall probe and the novel Hall probe. Now we get to look at how to make them and how far I got in making them. Commensurate with their sophistication in structure, the electron beam lithography-based fabrication of conventional sensors will only be discussed briefly before we move on to the much more involved novel sensor fabrication procedure, which features photolithography supplemented by reactive ion etching and ion beam milling.

The main goal here was to make a novel probe as a proof-of-concept. While I did not get that far, I made significant progress towards that goal.

6.1 Conventional Probes

I fabricated the conventional probes using standard electron beam lithography (EBL). I used the Raith GmbH e_LiNE system at 4D LABS to make a PMMA mask to make the probe pattern. I then thermally evaporated bismuth and lifted off the excess bismuth by dissolving the mask in acetone. Finally, I attached leads to the probe using contacts made by pressing indium into the bismuth and evaporated a protective germanium layer on the active region of the probe. The interested reader can consult Appendix G for more details of the fabrication procedure.

An example of a device after lift-off and without germanium is shown in figure 6.1.
Figure 6.1: Hall probe with 100 nm active region. It was written with 10 kV electrons at a dose of 132 $\mu$C/cm$^2$. Pattern written August 19, 2009, chip 2.
6.1.1 EBL Resolution Tests

I tested the spatial resolution of the conventional probe fabrication process, varying the EBL electron accelerating voltage, beam aperture, and PMMA electron dose. The pattern I used was a fractal Hall cross pattern with a range of different cross sizes from 2 µm to about 30 nm.

In figures 6.2 and 6.3, we see examples of results of the tests that I did. The 40 nm minimum feature size is quite close to the 30 nm the Hall cross size I specified.

While the tests were done on a pattern without important structures, like leads, that may complicate the attainment of the same EBL spatial resolution in a full device, the results set a benchmark for what is possible for Hall probe fabrication with the e_LiNE system.

6.1.2 EBL Writing Problems

Throughout my testing of Hall probe patterns, I have encountered omissions or deformations in the written exposures. The causes of these defects go to the heart of the operation of the Raith e_LiNE e-beam writer and its software. Such details are more appropriately contained in a training manual for the e-beam lithography system, so I will try to sketch them out briefly, as needed.

One type of defect reproducibly seen in my probe patterns has been the omission of large areas of the designs, up to hundreds of square microns. The defects mainly appeared when the e_LiNE system was instructed to rasterize a design into an uncommonly large number of exposure points. According to the local Raith service technician, Mirwais Aktary, the omissions may be caused by a memory buffer overload from the lithographic computer handling too many exposure points. We can decrease the number of exposure points by increasing the step size between exposed dots and by increasing the write field size.

Another type of defect is the omission or deformation of small parts of a pattern, usually with one linear dimension on the order of tens of nanometres. Aktary suggested that these errors are due to issues with how the e_LiNE software fractures the pattern into sections over which the e-beam rasters. If a fractured section has a dimension on the order of the step size between exposed dots, parts of the pattern may not be written properly. Gaps may also appear in the fractured pattern. I attempted to change the behaviour of the e_LiNE software by breaking up the affected patterns into smaller pieces, which the software would presumably fracture differently. The patterns wrote properly, but it is still not known for certain what is
Figure 6.2: Bismuth image, 30 kV, 10 µm aperture, 220–600 µC/cm² dose. Pattern written August 20, 2007.

Figure 6.3: Bismuth image, 30 kV, 10 µm aperture, 420 µC/cm² dose pattern from above. Minimum feature size is about 40 nm though the design specifies about 31 nm. Width of the biggest cross is 490 nm, though the design specifies 500 nm.
causing the writing errors. The errors’ appearance is intermittent, making them the extremely frustrating and difficult to troubleshoot.

These issues remain unresolved for the duration of my experimental work.

Figure 6.4: The appearance of write errors in EBL patterns is bafflingly inconsistent. All three patterns were made with similar settings on the e-beam writer, yet one is written well and the others are not. The gaps, about 50–100 nm, appear in both PMMA and bismuth patterns.

Figure 6.5: More examples of EBL writing errors. Resist image of patterns exposed April 28, 2009.

6.2 Novel Probe Fabrication: A Blueprint

Making a 3-layer novel probe structure with fabricated interconnections is not a trivial matter. Success of complex, multi-step procedures tends to decrease as the number of steps increases and so it behooves workers in the field to make many such devices in parallel in hopes that low device yields are made irrelevant by the flood of fabricated
The mask patterns that I made would produce 34 Hall probes on a 2-inch wafer if all devices were functional.

The general procedure I present here was developed by me in consultation with staff at 4D LABS: Mary Boysel, Tom Cherng, Nathanael Sieb, and Grace Li. Jamal Bahari from Dr. Ash Parameswaran’s research group at SFU’s School of Engineering Science was helpful specifically with the wafer-dicing process. Precise details of the fabrication procedure can be found in Appendix G.2.

To be clear, the blueprint I present here consists of the steps that I actually carried out or planned to carry out for making the novel Hall sensors. For tasks in which I tried more than one method, I have omitted the “inferior method” for clarity. When I discuss the results of the fabrication in section 6.3, I will include these inferior methods for completeness.

There is some arbitrariness to the processing choices that I made, and they were not necessarily the optimal ones, but they were choices that allowed me to proceed in a timely manner. My goal was proof-of-concept for the novel design, not optimizing the fabrication process for high device yields. Any improvements that I thought of after I ceased fabrication efforts will be discussed in section 6.4, after going through the fabrication procedure and results.

The procedure can be split up into five parts: patterning of bottom gold pads, trilayer patterning, deposition of top gold pads, active region fabrication, and finishing. Like the conventional probe fabrication we discussed earlier, we can categorize parts of the procedure as coarse and fine fabrication steps. The coarse steps would be making the gold pads and the main “boomerang” structure of the probe, which are all fabricated with photolithography. The active region fabrication is a fine fabrication step in which a FIB is used to mill out a trench in the trilayer, which is then filled with bismuth to form the active region. Finishing is fairly minor, involving wiring and probe mounting.

As we go through the processing procedure in detail, the reader may want to refer a few times to figures 6.6–6.8, which show the mask patterns used in the novel probe fabrication.

### 6.2.1 Bottom Gold Pads

The purpose of the bottom gold pads is to provide large, robust, wire-bondable electrical contacts for the bottom Hall probe leads. The contact thickness should taper
Figure 6.6: Outlines of all masks used for novel probe fabrication are overlaid in this figure. Red: first use of gold pad mask. Black: main probe pattern mask. Blue: second use of gold pad mask. Green: 2" wafer to be patterned; ideally, the wafer would be (100) Si and would have its main flat on the left or right so that dicing can occur neatly along cleavage planes. There are two lines of alignment marks placed vertically to allow the masks to be aligned precisely. The two masks used to make this overall pattern are shown separately in figures 6.7 and 6.8.
Figure 6.7: Gold pads pattern mask. Enclosed areas are transparent and are the features to be made. The array of small rectangles is for gold pads. The two vertical lines with tics are for mask alignment. The large rectangles on the outside are viewports for roughly centring the mask over the wafer.

Figure 6.8: Probe “boomerang” pattern mask. Enclosed areas are opaque and are the features to be made. The vertical line of features is for mask alignment. Small squares are placed near the active regions of the probes to allow for tests of the FIB before a milling fabrication step for the active region.
off smoothly so that the aluminum overlapping on the gold will be continuous with
the aluminum on the substrate\(^1\).

The method I devised to ensure continuity was to perform lift-off using a special
lift-off resist (LOR, manufactured by MicroChem) on the bottom and AZ MiR 703
resist on top. After spin-coating the resists on the wafer, the sample is exposed to UV
light through the gold pads mask shown in figure 6.7 and resist is developed. After
that, a 5 nm titanium adhesion layer is deposited before adding a 200 nm gold film,
both by e-beam evaporation. Finally, the resists are dissolved to lift off the undesired
portions of the metals and the wafer is briefly descummed in a plasma etcher.

Let us also take a brief tour of the mask design used here. Looking at figure 6.7,
we see an array of small rectangles for the gold pads. Large rectangles around the
outside serve as viewports so that the user can centre the mask over the wafer by eye.
A small “R” at the top left of the mask pattern tells the user which way the mask
should be facing; the R should be in the correct orientation when the mask is in use.
There are two sets of alignment marks running vertically for the entire width of the
mask pattern. There is one set each for the device and gold pad layers that will go
on top. We will take a closer look at the alignment marks in the trilayer and top gold
pads discussion next.

### 6.2.2 Trilayer

The trilayer forms the base of the novel Hall probe. It is a sandwich of aluminum,
silicon dioxide, and aluminum layers stacked on top of each other, fabricated in the
Hall probe boomerang shape. The bottom aluminum layer connects to the gold
contacts made in the previous step while the top aluminum layer will get its own gold
contacts in a later step.

The trilayer fabrication procedure uses photolithography and reactive ion etching
(RIE). First, the Al/SiO\(_2\)/Al trilayer is sputtered onto the wafer. Then AZ MiR
703 resist is spin-coated on top, exposed using the probe pattern mask in figure 6.8,
and developed\(^2\). With the resist-covered parts protected, we then subject the wafer
to a series of RIE treatments to remove the unwanted portions of the trilayer. The

\(^1\)An astute reader may see problems with connecting aluminum and gold directly, since they form
a brittle and poorly-conducting intermetallic “purple plague” at high temperature. One solution
would be to deposit a thin layer of some other metal, like titanium, between the gold and aluminum.
I did not do this because I was not aware of the problem at the time, but it would be easy to
implement in the future.

\(^2\)The veteran cleanroom user will know that the standard developer for AZ MiR 703 resist is
highly basic and reacts with aluminum metal. I later used an AZ diluted developer that is less basic.
aluminum etchants used are BCl$_3$ and Cl$_2$ while the silicon dioxide etch used is CF$_4$.

The wafer should then be dipped in deionized water to remove any residual chlorine or fluorine radicals that may form acids which can corrode the remaining aluminum. Finally, the remaining resist can be removed by flood exposure to UV light and dissolution by a developer.

Let us now look at the probe mask design, seen in figure 6.8. There is a long line of marks, shown in figure 6.9, for easy coarse alignment of both angle and position of the mask. Smaller rectangles offer slightly finer points of alignment. Finally, the very fine alignment is between the tiny alignment crosses on the wafer, previously deposited by the gold mask pattern, and little L-shaped marks on the mask. Alignment with these markers should be accurate to within a few microns since the linewidth of the crosses is about 10 microns.

![Figure 6.9: Left alignment marks. Red: gold pad pattern. Black: probe pattern. The large rectangles, present on both the gold pad and probe patterns, are for rough alignment while the red crosses are for fine alignment with the black L-shaped pieces.](image)

6.2.3 Top Gold Pads and Dicing

The top gold pads sit just atop the upper aluminum leads, providing a pair of large wire bonding points to the novel probe. We have already covered how to lay down gold pads, with the main difference here being that the lift-off layer is not necessary because we do not need a gradual taper to the slope of the gold pads.

There is one last detail of the gold pad mask design that I would like to address as it pertains to aligning the mask for the second set of gold pads. This is done using the set of alignment marks on the right side of the pad mask pattern. I wanted to make just one mask for the gold pads, so the alignment marks had to be able to align to the same pattern previously laid down in the first pad deposition. I solved that
problem with the design shown in figure 6.10. As the figure illustrates, the alternating pattern of marks makes it easy to align the transparent windows on the mask with the alignment marks on the wafer. The small rectangles are used to mark the offset required for the second set of gold pads relative to the first set.

Figure 6.10: Detail on the right alignment marks. On the left, we see how the mask pattern can be overlaid with the existing wafer pattern for a precise alignment. On the right, we see that the gold contact pads, the large red and blue rectangles, have exactly the same spacing as the small rectangles near the fine alignment marks.

Dicing is fairly straightforward, done by using a wafer scriber that scratches out lines of preferential cleavage in the wafer. It is especially advantageous to use wafers that naturally cleave into rectangles, such as (100) Si.

6.2.4 Active Region

Fabricating the active region is the most difficult step. We need to join the aluminum leads with bismuth to make the active region of the sensor. The method I devised to do this is basically to use a focussed ion beam (FIB) to mill out the active region area and evaporate bismuth into the resulting trench. It is not as simple a matter as
that, though, and we will go through the active region fabrication procedure in more
detail below.

The first step is to spin-coat PMMA on top to prevent short-circuits from excess
bismuth. The second step is to evaporate carbon on top and use carbon paint to
ground the top film. This guards against sample charging, which is bad for scanning
electron microscope (SEM) imaging and FIB milling. The third step is to do the
actual milling through the carbon, PMMA, aluminum, and silicon dioxide layers. To
do this, I used the FEI Strata 235 Dualbeam, a combination SEM, FIB, and an
energy dispersive X-ray spectroscopy (EDX) system. The EDX capability was used
to determine when the trilayer had been penetrated. The final steps are to evaporate
bismuth onto the sample and use acetone to lift off the undesired bismuth.

In our discussion of the active region fabrication procedure, we glossed over two
things. The first is that we are unlikely to make the exact shape we intend for the
active region. It is possible that after performing this procedure, some of the PMMA
will not come off due to overexposure from SEM imaging and secondary electrons
arising from the ion beam milling. Bismuth may adhere to residual PMMA. The
depth of the milled trench is also not too easy to measure, so it would be best to
overfill the trench with bismuth to ensure the Hall probe is viable. Fortunately,
we saw in section 5.3 that simulations indicate that overfilling the trench does not
significantly change Hall probe performance.

There are also a series of caveats regarding the use of EDX to find the milling
endpoint. The first is that spatial resolution of the EDX technique is on the order of
microns because of electron scattering. The scalability of EDX from an active region
width on the order of microns to one on the order of nanometres is questionable,
though I would hope the milling dose of gallium ions is the same for large and small
trenches so that we could rely on dose calibrations using larger trenches. EDX analysis
for a spot in a trench is also potentially inaccurate since X-rays emitted from deep
in the trench are more likely to be absorbed or scattered away. Finally, the EDX
signals from non-surface features may overwhelm the surface X-rays we want to detect.
This can be mitigated somewhat by angling the electron beam and decreasing the
electron acceleration voltage, though the latter may also decrease signals we want to
detect. Nonetheless, we will see later that EDX is still able to give useful qualitative
information about the composition of the surface.
6.2.5 Finishing

The finishing steps are to glue the probe chip to a carrier chip, and wire bond the probe to a carrier chip. These should be relatively straightforward steps. The probe can then be glued to a carrier chip by GE varnish and gold ball-bonded to gold contacts on the carrier chip. The pattern mask for the carrier chips is shown in figure 6.11. The fabrication process for the carrier chips is similar to that of the bottom gold pads. A dicing saw was used to cut up the wafer, but a scriber would do just as good a job.

![Figure 6.11: Photolithographic mask for making Hall probe carrier chips. Shaded in yellow are the parts made of gold, including the pads for soldering to the probes, which are shown in black. In green is the size of the 2-inch wafer on which this would all fit.](image)

6.3 Fabrication Results

With a detailed plan worked out, how far did I get in implementing it? The first set of gold pads were relatively straightforward to lay down. The trilayer took some experimentation to properly fabricate, but I ultimately succeeded in making the probe boomerang shapes. The active region fabrication proved to be quite difficult, and I
was unable to take any novel Hall probes beyond this step. The device yields did not supply many prospective probes, and those that reached this step did not work. However, tests done on non-viable probes and scraps indicated that the procedure could be made to work, in principle. I will discuss the results of the fabrication attempts with a focus on what I learned in the process.

### 6.3.1 Bottom Gold Pads

I started fabrication on 2” silicon wafers with 200 nm of silicon dioxide thermally grown on top. Fabrication of the bottom gold pads was quite successful and largely without problems. The deposited features were correctly sized, as figure 6.12 shows, and I only encountered minor problems with lifted off material re-depositing on the wafer. This problem was easily resolved with more aggressive rinsing of the wafer.

![Figure 6.12: Typical result of photolithography to make bottom gold pads pattern. The dimensions in the image are very close to the ones in the design, which are 10, 100, and 200 microns. Exposed May 30, 2011 (wafer 2). Imaged by light microscope.](image)

### 6.3.2 Trilayer

Making the trilayer was somewhat more difficult. With finer features, etch recipes, and development schemes to test, there is more potential for problems than with the simple lift-off scheme used for the bottom gold pads.
The first task in refining the trilayer deposition process is to tune the exposure, development, and baking times so that the photoresist pattern develops properly. Near the active region, the probe boomerang narrows to the micron scale, which strains the spatial resolution limits of the equipment available to me. I could consistently make whole novel probe boomerangs in resist at 4 s exposures to near-UV light with the OAI Model 800 MBA Mask Aligner.

Depositing the Al/SiO$_2$/Al trilayer at thicknesses of 30 nm/150 nm/30 nm was fairly straightforward. Default sputtering settings were determined by 4D LABS staff and were sufficient for putting down the layers. The quality of the layers is open to questions, though. In particular, I worry about pinhole defects in the insulating SiO$_2$ layer. Later tests of etched trilayers indicate 8 out of 34 novel probes have some conduction past the SiO$_2$ layer.

Table 6.1: Sputtering settings for depositing the trilayer. The parameters shown for aluminum were used for both layers.

<table>
<thead>
<tr>
<th>Layer Deposited</th>
<th>Thickness (nm)</th>
<th>Power (W)</th>
<th>Gas Pressure (mTorr)</th>
<th>Deposition Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al</td>
<td>30</td>
<td>350</td>
<td>3 (100% Ar)</td>
<td>220</td>
</tr>
<tr>
<td>SiO$_2$</td>
<td>150</td>
<td>350</td>
<td>3 (97% Ar, 3% O$_2$)</td>
<td>3900</td>
</tr>
</tbody>
</table>

The actual chemical composition of the nominal SiO$_2$ layer is of some importance because we know SiO$_2$ is insulating, but some other related compound may not be. To have some assurance that the PVD procedure grew SiO$_2$, I performed X-ray photoelectron spectroscopy (XPS) analysis to figure out the oxidation state of the silicon atoms, which should be 4+ for SiO$_2$. The binding energies for various oxidation states of silicon are shown in table 6.2.

Table 6.2: Binding energies for Si 2p electrons for different oxidation states of silicon, from Hao et. al. [52]. These values are for a C 1s calibration signal at 284.5 eV.

<table>
<thead>
<tr>
<th>Species</th>
<th>Si$^0$</th>
<th>Si$^{1+}$</th>
<th>Si$^{2+}$</th>
<th>Si$^{3+}$</th>
<th>Si$^{4+}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binding Energy (eV)</td>
<td>99.3</td>
<td>100.3</td>
<td>101.0</td>
<td>101.9</td>
<td>103.3</td>
</tr>
</tbody>
</table>

As figure 6.13 shows, the XPS data are consistent with there only being the Si$^{4+}$ oxidation state present in the film. We can identify a charging peak that decreases as a function of X-ray flux. Increasing dwell time does not seem to affect the overall shape of the spectrum, but does decrease noise.

The next step was to prepare the photoresist template on top of the trilayer, and that is when we see the first significant problem in the novel probe production: a
Charging Effects on Si 2p XPS Spectra

(a) Sample charging as a function of both X-ray flux (a good proxy is the e-beam current incident on the X-ray source, in mA) and dwell time for any data point (in ms).

(b) Decomposition of Si 2p peak area into Si 2p peak and charging peak. The only Si 2p electron signal seen is for a 4+ oxidation state. The “fitted curve” is the sum of fitted Si$^{4+}$ and charging Gaussian curves. E-beam current 1 mA, dwell time 500 ms.

Figure 6.13: Section of XPS spectra for nominal 50 nm of SiO$_2$ sputtered onto a glass substrate at 4D LABS. Film growth settings were the same as in table 6.1 except sputtering time was 1300 s. Sample 111216SiO$_2$.

chemical reaction between the original developer I used and the aluminum in the trilayer. I observed a white substance emanating from the surface, possibly tiny hydrogen bubbles. An example of the damage to aluminum is shown in figure 6.14. The developer, AZ 300 MIF, is very basic and dissolved the aluminum.

Fortunately, I was able to access AZ Developer Diluted 1:1$^3$ that does not dissolve much aluminum. When tested, I found using a profilometer that the developer dissolves 4 nm of aluminum in 10 minutes, and a wafer is typically immersed in developer for only about a minute.

Etching was the next step, and the next point of significant difficulty. My original procedure called for a three-step wet etch procedure to remove the unwanted parts of the trilayer: Transene Aluminum Etchant A$^4$ for the Al layers and buffered oxide etch (BOE) diluted 10:1, a hydrofluoric acid etch, for the SiO$_2$ layer. Sequential use

$^3$Special thanks to Xin Zhang and the CIO Group for use of their supply of developer.

$^4$Again, courtesy of Xin Zhang and the CIO Group.
of the aluminum etchant, BOE, and aluminum etchant resulted in severe undercut, as shown in figure 6.15. With this degree of undercut, sections of the aluminum leads going to the active region were destroyed and the probes were not viable. At this point, I turned to RIE as a more anisotropic alternative etch.

So then I needed to find good RIE recipes for etching away the Al and SiO$_2$ layers, for which I used BCl$_3$/Cl$_2$ and CF$_4$, respectively, as the process gases. Tests of the RIE process consisted of applying an etch, then using a Cascade Microtech M150 Probe Station to check whether the etched surface conducts. Of course, an Al surface is expected to conduct while a SiO$_2$ surface should be insulating, though this simple picture is complicated by the fact that the probes can scratch through the surface and electrically contact the film underneath instead.

From these tests, I worked out some basic etch parameters for the BCl$_3$/Cl$_2$ and CF$_4$ RIE processes, shown in table 6.3, but did not make much effort to minimize the etch times. I did a bit of experimentation with the BCl$_3$/Cl$_2$ etch, which indicated that the etch time can be decreased to 60 s, possibly lower, but time constraints prevented me from doing more.

Testing of the RIE procedure revealed two problems. The first is that after the three RIE treatments (BCl$_3$/Cl$_2$, CF$_4$, and then BCl$_3$/Cl$_2$ again), I noticed some corrosion of the aluminum films. In one case, I even observed bubbling under the resist. I strongly suspect that some residue remaining from the RIE treatments had

Figure 6.14: Effects of AZ 300 MIF developer on an Al/SiO$_2$/Al trilayer. The top layer of aluminum appears to be severely damaged. Sample 110325, imaged by light microscope March 6, 2012.
(a) Portion of novel probe boomerang after trilayer deposition and photoresist patterning for probe shapes. (b) Novel probe boomerang after wet etches to remove excess portions of trilayer. The resist appears intact, but aluminum has been etched away.

Figure 6.15: Trilayer and patterned resist, representative probe pattern from March 14, 2012 resist patterning for the 110325 wafer. Images taken with light microscope.

Table 6.3: RIE etch test parameters used for most of the Al and SiO$_2$ etching I did. SCCM is a standard cm$^3$ per minute.

<table>
<thead>
<tr>
<th>Layer to Remove</th>
<th>Process Gas</th>
<th>Flow Rate (SCCM)</th>
<th>Power (W)</th>
<th>Pressure (mTorr)</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al</td>
<td>Cl$_2$</td>
<td>10</td>
<td>100</td>
<td>50</td>
<td>90</td>
</tr>
<tr>
<td>BCl$_3$</td>
<td>35</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SiO$_2$</td>
<td>CF$_4$</td>
<td>30</td>
<td>100</td>
<td>30</td>
<td>270</td>
</tr>
</tbody>
</table>

reacted with moisture in the air to form acids, which reacted with the aluminum to form hydrogen gas. The effects of such reactions are pictured in figure 6.16. A deionized water rinse after each RIE treatment was sufficient to stop the apparent acid corrosion.

The second problem I encountered was a translucent residue that appeared to be preventing the final BCl$_3$ etch from removing the bottom Al layer. I suspect that the CF$_4$ etch caused a layer of teflon to form on top of the bottom Al layer. The carbon-based etchants tend to polymerize and leave coatings on the etched sample. While this may be desirable for protecting the sidewalls from undercut, it can also impede etching down into the wafer. Teflon is particularly bad because it is famously inert to most chemical attack, so when using RIE, only the physically ablative action is effective against it. Another possibility is a reaction between the bottom Al layer
Figure 6.16: A probe pattern after trilayer deposition and three RIE etches to remove excess parts. There is apparently some damage from RIE acid residues. Light microscope image from 121114 test wafer.

and the fluorine-based etchants to form AlF$_3$ or similar products. Again, intensifying the ion impact component of RIE is probably the best option as AlF$_3$ is unlikely to be very reactive. Time constraints prevented me from experimenting with remedies for this problem, but I will suggest several approaches in section 6.4.

6.3.3 Top Gold Pads and Dicing

As with the deposition of the bottom gold pads, the fabrication of the top gold pads was routine and uneventful. Mask alignment for the top gold pads was straightforward and the pad placement was adequately accurate. The residue that remained after the etching and top gold pad deposition did not seem to affect the patterning of the novel Hall probes since the photolithographic steps that needed to be precise had already been done.

Tests with the probe station showed that 10 out of a possible 34 trilayer boomerangs on the wafer were properly-structured, with I-V characteristics that are roughly exponential between the two aluminum layers. 8 had ohmic conduction past the insulating SiO$_2$ layer. 21 did not have conducting paths through the constriction where the active region would go.

At this point, I wanted to dice the wafer into individual probe pieces. I coated the wafer of probe boomerangs with about 4000 Å of PMMA in hopes of protecting the
Figure 6.17: Light microscope image of sample 110428 after lift-off for the second set of gold pads. The pattern appears to be continuous.

Figure 6.18: Current versus voltage of an Al/SiO$_2$/Al trilayer, with 9.5 mm$^2$ area. The current is tiny throughout, with a possible breakdown of the dielectric happening as the I-V characteristic slopes up. The data are from a single measurement on a single sample. It is strange that there appears to be two curves offset from each other.
wafer from damage during the dicing. For the dicing, I am grateful for the assistance of Jamal Bahari from Dr. Ash Parameswaran’s research group. Bahari operated a wafer dicer in the Engineering Science cleanroom for me. The dicer’s diamond-coated blade is constantly sprayed with water to lubricate and cool it as the blade cuts across a wafer crystal-bonded to a large block.

Unfortunately, the dicing process appeared to be very damaging to the probes on the wafer. A few cuts to the wafer caused parts of the probe films to flake off. At this point, I stopped the dicing. It is possible that other forms of damage occurred to the boomerangs as well, such as oxidation or erosion from the water. Damage related to heat from using crystal-bond is also possible. It is not clear precisely what caused the most damage, but the combination of water flow, vibration from the saw, the cutting, and heat was sufficient to destroy all but one of the trilayer probe patterns.

Bahari was very helpful in the aftermath of this dicing setback, suggesting a different way of dividing up the rest of the wafer using a specialized scriber. Later, I tested the technique on a 2-inch wafer using the Tempress 1713-10C Wafer Scriber in the 4D LABS cleanroom and was highly successful in precisely dividing the test wafer. Scribing and cleaving is an ideal technique for delicately dicing wafers of Hall probes.

### 6.3.4 Active Region

At last, we come to the hardest part: milling out the active region and filling it with bismuth. Recall that I first need to coat the prospective probes in PMMA and carbon, then mill out the active region using a FIB, thirdly evaporate bismuth into the trench, and finally dissolve the PMMA to lift off the undesired bismuth.

If I were doing this from scratch with the scribe-cleave style of dicing, the sample would not need to be re-coated with PMMA. However, I suspect the PMMA on the remaining good prospective probe was damaged by the dicing saw, so I re-spun 4000 Å of PMMA on the diced probe chip.

Next, I evaporated carbon on the chip. The carbon evaporator I used does not do so in a well-controlled way, so there is not a great deal of value in explaining the experimental parameters, but for what it is worth, I evaporated carbon for about 45 s and estimate this thickness to be less than 200 nm.

---

5I believe it was graphite, but I do not remember clearly. In any case, the material is not particularly important. The block mainly served as a stable platform to mount the wafer while not damaging the saw as the blade cut through the wafer and block.
I did some FIB tests on expendable samples to get a rough idea of the settings to use, though these will vary because the thicknesses of resist and carbon deposited will vary from sample to sample. After that, I attempted to find a suitable place for milling out an active region on a practice sample. Trilayer structures under the resist and carbon were difficult to see using SEM imaging, though a 52° sample tilt and 15 kV electrons under the SED detector helped. Gold is still easily seen by the SEM due to strong backscatter of electrons, even under a layer of PMMA, so it may be helpful to find the gold pads first.

To see how much milling is required, I then performed a series of gallium dose tests, milling a different carbon/resist/trilayer test area each time, and then performing EDX analysis to check the composition. The results of the EDX analysis are shown in figures 6.19 and 6.20. The horizontal scales for all those graphs depict the amount of gallium charge per unit area fired at each test area. This is analogous to electron exposure from e-beam lithography, so I am going to call the corresponding gallium quantity the “gallium exposure” or “gallium dose.”

Note that I did not calibrate the EDX signals with any known standards before using the SEM software to calculate a composition, so the numbers shown in the EDX composition graphs are not to be taken as absolute. We should interpret them as relative shifts in the surface composition as one mills away more and more material. More importantly for us, we can still interpret the lack of aluminum, for example, as actually being a lack of aluminum. We merely want to know when we have milled away the trilayer and have reached the wafer material.

So how should we interpret the curves? The carbon curve should be strongly correlated with the presence of the carbon and PMMA layers on top. The presence of oxygen could be indicative of resist, the sputtered SiO₂ layer, oxidized aluminum, or the native oxide on the wafer. The aluminum signal is the most straightforward; it should only be present when the milling has reached one of the aluminum layers. A silicon signal could be detected either from the SiO₂ layer or the wafer. The gallium signal, of course, measures the presence of embedded gallium atoms. We should not give too much credence to vague trends in the carbon or oxygen signals, though. The X-ray-generating electron relaxations in carbon and oxygen involve bonding electrons, and such interactions reduce the quantitative accuracy of the results.

In figure 6.19, we see all of the data plotted on the same graph. Note that I did not take data for gallium exposure below about 1.5 nC/µm² since we are mainly interested in detecting when we have milled through to the substrate.
Figure 6.19: Uncalibrated EDX data from ion-milling tests. A series of 1 micron by 5 micron rectangles were milled and subsequently checked for composition using EDX. The beam current drifted slightly during data collection from 48 to 49 pA.

(a) The first 13 data points taken in the data shown in figure 6.19. The “early” graph.

(b) The last 7 data points taken in the data shown in figure 6.19. The “late” graph.

Figure 6.20: The data from figure 6.19 make more sense when plotted separately, with averaging when there were two data points for a particular exposure. The plots overlap for doses greater than 2.8 nC/μm². The data are consistent with the combined thickness of the carbon, resist, and trilayer being smaller for the second graph than the first graph. Perhaps the structure being milled was not flat or the repeated milling and imaging eventually did damage over a wide area.
Before we do any detailed analysis of the graph, notice that there is much more variation in composition on the high exposure end of the graph compared to the low exposure side. I think this is due to some kind of damage in the sample. If we separate the data of the first 13 points and the last 7 points, ordered chronologically, we get the graphs shown in figure 6.20, with minor data averaging in the separated data sets. Let us call them the “early” and “late” graphs, respectively. The data are much less variable separated in this manner. I would argue that the late graph is for FIB performed on a damaged sample, so the milling reached the silicon substrate; we see this in a silicon EDX signal plateau. The same dose in the early graph does not show the silicon plateau.

I did observe damage to the surface, but only when I took the temporal last data point. I was focussing very closely on the surface and I think the SEM's electron beam charged the surface enough for the top layer to explode. It is also possible that ion beam imaging damaged the surface, though I confined its use to taking quick single images for focussing. It is unlikely that the carbon/resist/trilayer films varied enough to cause all of the effects observed since 3 of the “late” points were spatially close to the “early” points.

There are at least two plausible broad interpretations of what is happening in the early graph. The first is that we see the last bit of resist being milled away at approximately the 1.5 nC/µm² dose, then the gallium ions remove the Al, SiO₂, and Al layers sequentially until the substrate is reached. The second is that the ion beam is already blasting through SiO₂ at 1.5 nC/µm² and we only see the subsequent removal of the SiO₂ and bottom Al layer before the wafer is reached. However, in a previous experiment, I made more qualitative observations from EDX data that indicated a transition from a carbon- and oxygen-dominated EDX signal to one with much stronger aluminum and silicon signals around 1.4–1.7 nC/µm², so I favour the first explanation.

Table 6.4: EDX expectations vs. actual EDX signal changes for the scenario where we see resist and trilayer milled through in the EDX data.

<table>
<thead>
<tr>
<th>Event</th>
<th>Exposure Range (nC/µm²)</th>
<th>Expected EDX Signal Changes</th>
<th>Actual EDX Signal Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resist milling</td>
<td>1.5–1.7</td>
<td>C ↓, O ↓, Al ↑</td>
<td>C ↓, O ↑, Al ↑</td>
</tr>
<tr>
<td>Al milling</td>
<td>1.7–2.2</td>
<td>Al ↓, Si ↑, O ↑</td>
<td>Al ↓, Si ↑, O ↑</td>
</tr>
<tr>
<td>SiO₂ milling (trilayer)</td>
<td>2.2–2.6</td>
<td>Si ↓, O ↓, Al ↑</td>
<td>Si →, O ↓, Al ↑</td>
</tr>
<tr>
<td>Al milling</td>
<td>2.6–3.1</td>
<td>Al ↓, Si ↑, O ↑</td>
<td>Al ↓, Si ↑, O ↑</td>
</tr>
<tr>
<td>SiO₂ milling (wafer)</td>
<td>3.1–3.4</td>
<td>Si ↑, O ↓</td>
<td>Si ↑, O ↓</td>
</tr>
</tbody>
</table>
We should try to account for the anomalies in the data that do not correspond well to the first account, though. The most glaring one is the weak first Al peak at about 1.7 nC/µm², at least compared to the second Al peak at about 2.6 nC/µm². Perhaps the top layer of aluminum was thinned through chemical attack despite steps taken to prevent developer and RIE residue reactions. Another possibility is that the spacing of the gallium ion doses unfortunately missed the first Al peak, but found the second Al peak, although given the much larger width of the second peak, this does not seem like a complete explanation.

Speaking of the width of the second Al peak, I also wonder about why it is so wide. Again, it is possible that the spacing of gallium doses was too coarse to get a precise width, but what if it is actually very wide? As the gallium ions mill deeper, we would expect sample charging to become an issue since only the top layer is grounded. This may make the ion beam less focused and make the milling process more inhomogeneous in the trench. The inhomogeneity would worsen the deeper the trench is, which would widen the second Al peak. An alternate explanation would be that we are seeing signal from the FIB sputtering aluminum in the trench. As the trench deepens, more material will tend to re-deposit inside the trench.

The other interesting anomalies are the increasing oxygen signal during the resist-milling phase and the increasing Si signal throughout. Both are likely due to penetration of electrons through the surface layer to generate X-rays underneath.

As for the late graph, there is not a great deal to say about it since there is not much data there, but we can say that the points at about 2.8 and 3.1 nC/µm² have similar compositions as the points around 3.0 and 3.3 nC/µm² on the early graph. It is plausible to consider the late graph a continuation of the early graph with some offset. In that case, we can then conclude that the thermal oxide on the wafer is milled through at about 3.3 nC/µm² on the early graph, or on the undamaged sample.

In any case, the main result of the milling tests is that we can detect the point at which the trilayer is completely milled through. From the early graph, this point is between 3–3.15 nC/µm² of gallium ion exposure mills through about 400 nm of PMMA and a 30 nm/150 nm/30 nm trilayer of Al/SiO₂/Al.

The last part of the active region fabrication that I tested was evaporating the bismuth into milled trenches and lifting off the excess bismuth. We can evaluate the success of this particular step in two ways: the quality of the deposited bismuth film and the desired versus actual dimensions of the bismuth feature. If any of the probes worked, we would certainly be evaluating magnetotransport properties too.
(a) Nominal 5 micron by 1 micron trenches. Test wafer 120328, from July 11, 2012 FIB test.

(b) Nominal 5 micron by 0.2 micron trenches. Test wafer 120328, from July 11, 2012 FIB test.

Figure 6.21: Bismuth trench-filling experiments for nominally 5 micron by 1 micron trenches. Note these images were obtained with a 52° sample tilt, which distorts the dimension going up and down the page. The numbering corresponds to trench numbers in table 6.5. The beam current used was 40 pA. SEM images of test wafer 120328, from July 11, 2012 FIB test.

The results of the milling and bismuth filling procedure can be seen in figure 6.21. The black parts, aside from some artifacts due to charging, appear to be the bare test wafer. We see some success in filling the milled structures with bismuth, shown in grey, but we also see white features around the milled spaces, which may prevent lift-off of some of the excess bismuth. This white material is probably PMMA that cross-linked from being adjacent to gallium ion exposure and remained undissolved after lift-off. In the 5 × 1 trenches, the bismuth appears to form a good quality film, but this is not the case for the 5 × 0.2 trenches. While it is unclear whether the lower sections of the small trenches are similarly affected, the top layer is definitely a low-quality film, perhaps because the aspect ratio of the trench is too high for efficient evaporation without bismuth becoming stuck to the sidewalls.

It is also interesting to look at the dimensions of the bismuth structures. For this, let us look specifically at the structures made using gallium doses around the trilayer breakthrough of about 3.15 nC/µm². The data are collected in table 6.5. The “Actual Dimensions with PMMA Border” given the table are hard to pin down precisely since some segments of PMMA seem to have moved after evaporation, have variable widths,
and are sometimes difficult to distinguish from the bismuth. To impose some kind of regularity to the measurement, I ignored the parts of the PMMA walls that appear to spread out sideways, which appear to have a Λ-shaped cross-section. Nonetheless, the dimensions with PMMA border measurements should be viewed with a small degree of scepticism.

Table 6.5: Various measurements of the sizes of trenches milled through carbon, PMMA, and the Al/SiO$_2$/Al trilayer. Note the PMMA border came off the three smaller trenches, so it could not be measured. Trench numbers correspond to those in figure 6.21.

<table>
<thead>
<tr>
<th>Trench Number</th>
<th>Dose (nC/µm$^2$)</th>
<th>Nominal Dimensions (µm × µm)</th>
<th>Actual Dimensions (µm × µm)</th>
<th>Actual Dimensions with PMMA Border (µm × µm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.88</td>
<td>5 × 1</td>
<td>4.4 × 0.69</td>
<td>4.8 × 0.94</td>
</tr>
<tr>
<td>2</td>
<td>3.12</td>
<td>5 × 1</td>
<td>4.6 × 0.82</td>
<td>4.7 × 1.1</td>
</tr>
<tr>
<td>3</td>
<td>2.88</td>
<td>5 × 0.2</td>
<td>4.6 × 0.20</td>
<td>—</td>
</tr>
<tr>
<td>4</td>
<td>3.12</td>
<td>5 × 0.2</td>
<td>4.5 × 0.20</td>
<td>—</td>
</tr>
<tr>
<td>5</td>
<td>3.36</td>
<td>5 × 0.2</td>
<td>5.0 × 0.25</td>
<td>—</td>
</tr>
</tbody>
</table>

While the measurements of the actual trench dimensions appear to be closer to the nominal dimensions for the smaller trenches, I suspect the PMMA borders are obscuring some deposited bismuth. We see a hint of this in trench number 4, where the PMMA has been stripped away from the top. Though I lack clear evidence for all the examples studied, I suspect all the actual trench dimensions are fairly close to their nominal sizes, within perhaps 100 nm. The measurements including the PMMA border, which are quite close to the nominal sizes, give an estimate of the actual trench sizes under the assumption that my suspicions are correct.

Should a probe not work after the milling and bismuth filling steps, it may be possible to save the probe with annealing as a last resort. This was where I was at after the one trilayer pattern that survived etching and dicing steps, but did not electrically connect after the milling and filling (and lift-off). Annealing should be avoided if possible since it may make undesired changes to the crystallite sizes or melt the insulating PMMA into the active region$^6$. A series of hot plate annealing treatments of 63 minutes at 100°C, 40 minutes at 150°C, and 31 minutes at 200°C, with conductivity tests in between, failed to make the probe work.

That brings us to the end of my experimental work on fabricating the novel Hall probe design. I established working procedures that made the trilayer boomerang pattern and gold contacts using photolithography, lift-off, and etching. I also figured

$^6$PMMA melts at about 160°C.
out a useful approach to fabricating the novel Hall probe active region using FIB milling and lift-off. I expect the untested finishing steps — gluing the probe to the carrier chip and wire-bonding to gold electrodes — to be relatively straightforward. My work here lays a firm foundation for a major evolution in Hall probe design.

### 6.4 Process Improvements

This is the first time that the fabrication of the novel Hall probe design has been attempted. As is typical in such instances, there are many avenues for improving the procedure. I will present here some process modifications that I would like to have tried given the time. They can be categorized as follows: material deposition, etching, and active region fabrication.

There are two things that I would suggest changing with respect to material deposition: using aluminum contact pads instead of gold pads and heating the substrate during the SiO$_2$ deposition. Both actions have potential advantages and drawbacks, though.

With the change to aluminum contact pads, there is no longer risk of gold-aluminum intermetallics forming between the contact pads and trilayer. While this can also be accomplished by adding protective chromium layers, using aluminum pads also has the added benefit of enabling room-temperature aluminum wire-bonding, so that the bismuth in the sensor will not be annealed unnecessarily, as it would in gold thermosonic bonding. While making a solderable carrier chip will be more complicated with aluminum wire bonding, fabrication of a wafer of such chips is a one-time event and should be achievable with one mask along with a little opaque tape\(^7\). There may be an increase in the contact resistance between the pad and the aluminum in the trilayer, but it should be a thin native oxide; I would not expect it to cause many problems.

Heating the substrate during SiO$_2$ deposition has the possible benefit of decreasing the number of defects in the SiO$_2$ layers. A few of the trilayer boomerangs were found to have short circuits between the Al layers, possibly due to pinhole defects in the SiO$_2$. Substrate heating during deposition should help improve the quality of the

\(^7\)For example, the wire-bondable pads and solderable pads can have a common aluminum base, the pattern created by one mask on positive resist for lift-off. Everything except the solderable pads can then be taped over on the mask and used to expose positive resist again, this time to allow deposition of chromium and gold, with lift-off to remove the excess.
insulating film, though it may also increase oxidation of the bottom Al layer in the probe.

For the etching-related improvements, there are certainly a few fine-tuning steps that can be taken, such as altering gas ratios or gas pressure. Etching times should be minimized to prevent unnecessary undercut. While it is unclear what caused some trilayer boomerangs to fail after the RIE steps, whether through non-optimal photolithography or RIE undercut, I would err on the side of more undercut prevention. With that in mind, I would suggest trying some different etch gases. CHF$_3$, for example, has a stronger tendency than CF$_4$ to form polymers.

Polymers may also have been a problem in my work, though, preventing the removal of the bottom aluminum layer in some places. Polymer removal can be enhanced for either the SiO$_2$ etch or second aluminum etch by strengthening the physical etching component of RIE, perhaps by increasing the fraction of argon present and decreasing the gas pressure. Adding oxygen gas will also help remove polymers chemically, but at the cost of also destroying the photoresist mask. The oxygen may also oxidize the aluminum parts of the trilayer boomerang, though if enough resist remains, there should only be a small amount of oxidation from the sides. Finally, etch time for the second aluminum procedure can be increased.

It is also possible that the bottom aluminum layer remained not due to polymer buildup, but from reactions with fluorinated gas to make aluminum fluoride. The same remedies as for the polymer case apply: increasing physical etching, perhaps adding oxygen, and increasing the second aluminum etch time.

The last etch-related thing I would suggest is more use of descumming for removal of residual resist and ashing for removal of all resist. It may be particularly useful to descum after the boomerang pattern photoresist development and after the FIB milling step. Both descumming and ashing use oxygen plasmas, with lower input power for the descumming compared to ashing. Perhaps descumming would even help with the possible polymer problem. Again, I would worry about aluminum oxidation but with careful tuning of parameters, it may be possible to minimize damage.

Finally, I have three suggestions to improve the active region fabrication. The first idea is to reduce the ion current used in milling out the novel probe active region. Although the trenches milled in figure 6.21 did not appear to have gross irregularities, I noted in the discussion that the differing widths of the aluminum EDX peaks could be interpreted as inhomogeneity of milling. This could be alleviated by reducing the ion current, which should reduce the severity of substrate charging. In any case, reducing the ion current would not, in principle, reduce the milling efficacy.
The other two ideas are for reducing the sidewall bismuth deposits and perhaps improving the film quality. Thinning the layer of resist should make it easier for evaporated bismuth to reach the bottom of the milled trench, though making the resist too thin risks some material not lifting off properly. It would also be good to find a resist less sensitive to exposure due to gallium ion bombardment. Ideally, the resist would still wash away after the trench milling, taking the sidewall deposits with it.
Chapter 7

Doped Bismuth as a Hall Probe Material

Oprah Winfrey: “Did you ever... dope... to enhance... performance?”
Lance Armstrong: “Yes.”
— from the TV show Oprah’s Next Chapter, season 2, episode 15.

In Chapter 3, we looked at the Hall probe theory in some depth and found that we have enormous freedom to choose different materials for Hall sensors, from conventional semiconductors all the way up to good metals. Preliminary calculations showed bismuth may be superior to semiconductors in Hall sensors, though we glossed over some complications. At least in small bismuth devices, we may see much less telegraph noise compared to semiconductor sensors. It is unclear whether bismuth is the best possible material to work with and modify to make Hall probes, but it is a good one to try. The lack of in-depth bismuth Hall probe work may be obscuring significant gains to be made.

In this chapter, I will present my work on the magnetotransport properties of doped bismuth. I measured resistivity and Hall effect in bismuth films for a variety of lead dopings in an effort to optimize bismuth for use in Hall sensors. The problem is a very complex one. Even undoped bismuth is best modelled by three bands of carriers, each with their own concentrations and mobilities, and with their own interactions with the environment. I would have liked to develop a fuller understanding of magnetotransport in doped bismuth, but time constraints prevented me from doing

If anyone is reading this decades later, Lance Armstrong was a competitive cyclist who took banned drugs to improve his cycling performance.
more. Such is the plight of the solid state physicist: each solid state system is a
universe in itself, with enough mysteries to consume a career, if one lets it.

This is necessarily a somewhat sprawling chapter as we go through a few different
approaches to make sense of the data and extract as much information as possible.
First, we will go over what is known about the band structure of doped bismuth
from the literature. Then we will use the frame of a three-band model to interpret
what is likely happening to the carrier concentrations and mobilities. Finally, I will
present some educated guesses as to what else may be happening in the doped bismuth
samples and what ramifications all of this may have for making bismuth Hall sensors.

7.1 Bismuth: Basic Properties

In this section, we will walk through a gallery of bismuth’s properties, starting with a
short historical section, then moving on to structural properties, band structure, and
effects of mixing with lead.

Bismuth is a grey semimetal. Its practical uses tend to be in small niches, though
the similarity of its physical properties to lead have made low-toxicity bismuth in-
creasingly popular as a replacement for toxic lead, which is being phased out of some
consumer applications like solder in electronics. Bismuth is a relatively unremark-
able substance except for the fact that relatively low magnetic fields are needed to
quantize electronic states, which helped make it the first element to have its Fermi
surface determined experimentally [53]. More interestingly for us, bismuth also has
the strongest classical Hall effect of all the elements, which is likely what drew inter-
est from Sandhu et al [39] when they first made bismuth Hall sensors in an effort to
escape the miniaturization problems they had with semiconductor Hall sensors.

The crystal structure of bismuth is categorized as a rhombohedral lattice system,
with an axis of 120° rotational symmetry. According to a bismuth crystal structure
review by Donohue [54], the average of measured lattice constants of bismuth at room
temperature are $a = 4.7460 \text{ Å}$ and $\alpha = 57.23^\circ$ for a primitive rhombohedral cell. For
a non-primitive hexagonal cell, the dimensions are $a = 4.5461 \text{ Å}$ and $c = 11.8623 \text{ Å}$.

The Fermi surface of pure bismuth features six half-ellipsoid electron pockets at the
$L$-points and two half-ellipsoid hole pockets at the $T$-points; the two carriers overlap
in energy, as seen in figure 7.2. The cumulative volumes of the electron pockets and
hole pockets are equal, making undoped bismuth a compensated semimetal. Light
hole pockets can be opened at the $L$-points at sufficiently high temperature and/or
doping. Experiments have shown that both the electrons and the light holes are low-effective-mass carriers while the other holes are relatively heavy [55–57].

The carrier compensation is detrimental to Hall probe performance, though. This is most clearly seen in the Hall coefficient for a two-band model of bismuth:

\[ R_H = \frac{p\mu_p^2 - n\mu_n^2}{e(p\mu_p + n\mu_n)^2}. \]  

(7.1)
There is at least a partial cancellation in the numerator, which decreases the Hall coefficient. In fact, my data in figure 7.3 show a zero crossing in the Hall coefficient at low temperature, making bismuth sensors useless in that regime. For applications like superconductor research, it is essential that the Hall probes work at low temperature. Ideally, the sensors would work well over a wide temperature range.

![Undoped Bismuth Hall Coefficient](image)

**Figure 7.3**: Hall coefficient for undoped bismuth. The zero crossing makes the sensor useless close to that temperature.

Doping bismuth to eliminate this cancellation is key to improving the performance of bismuth Hall probes. There is some work on magnetotransport in single-crystal doped bismuth, notably by K.K. Dey *et al.* [59, 60] and on doped bismuth films, notably by Schnelle and Dillner [61–63]. To my knowledge, there are no papers discussing the use of doped bismuth for Hall sensors, and that is one of the major gaps in knowledge that I seek to fill with this thesis.

For this thesis, I have chosen to remove all electron carriers by doping, or compensate electrons, using lead because lead has a vapour pressure relatively close to that of bismuth, making my evaporative sources last longer and the films less variable between deposition runs. For reference, see table 7.1 for details of the vapour pressures of pure bismuth and lead at various temperatures.
Table 7.1: Vapour pressure of pure bismuth and lead as a function of temperature. The numbers are taken from a graph in a review article by Honig [64].

<table>
<thead>
<tr>
<th>T (K)</th>
<th>$P_{\text{Bi}}$ (mm Hg)</th>
<th>$P_{\text{Pb}}$ (mm Hg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>600</td>
<td>$1.5 \times 10^{-8}$</td>
<td>$7.0 \times 10^{-9}$</td>
</tr>
<tr>
<td>700</td>
<td>$3.0 \times 10^{-6}$</td>
<td>$9.0 \times 10^{-7}$</td>
</tr>
<tr>
<td>800</td>
<td>$1.5 \times 10^{-4}$</td>
<td>$4.5 \times 10^{-5}$</td>
</tr>
<tr>
<td>900</td>
<td>$4.0 \times 10^{-3}$</td>
<td>$9.0 \times 10^{-4}$</td>
</tr>
<tr>
<td>1000</td>
<td>$5.0 \times 10^{-2}$</td>
<td>$1.0 \times 10^{-2}$</td>
</tr>
<tr>
<td>1100</td>
<td>$3.0 \times 10^{-1}$</td>
<td>$7.5 \times 10^{-2}$</td>
</tr>
</tbody>
</table>

The solubility of lead in bismuth appears to not particularly well-studied at low concentrations, though, with Gokcen [65] estimating a maximum of 0.5 at.-% of lead at 125°C. This is primarily of concern during evaporation of a bismuth-lead alloy. Bismuth and lead in the evaporated film may have time to phase-separate if the lead is insoluble, though it may be possible to mitigate this by cooling the substrate and freezing the lead dopants in place on deposition.

### 7.2 One Band Is Best

It is clear that having electron and hole carriers simultaneously in a Hall probe material is detrimental to performance, but how about having two hole bands of different mobilities? That is less clear, though the title spoils the punchline. The approach we will look at to show one band is best is somewhat cumbersome mathematically, so I will simply set up the problem and refer the interested reader to Appendix A.3 for more details. The end result of the calculation is that we always want to remove a lower mobility carrier and replace it with a higher mobility carrier until there is only one band.

Our approach will be to work out the material-dependent parameters in $B_{\text{min}}$ and then optimize the equation with respect to the fraction of carriers of one kind.

The symmetry of the problem dictates that there are only two distinct cases: both bands the same (two hole bands or two electron bands) and both bands different (one electron band, one hole band). As an example, let us check the case of two hole bands. The math is similar for two electron bands or one electron and one hole band.

We begin with the expression for $B_{\text{min}}$ from equation (3.32):

$$B_{\text{min}} = \frac{R_d}{R_H} \sqrt{\frac{P_{\text{th}}}{P_{\text{bias}}}}.$$
We have the following two-band expressions for Hall coefficient, electrical resistivity, and resistance:

\[
R_H = \frac{p_1 \mu_1^2 + p_2 \mu_2^2}{e(p_1 \mu_1 + p_2 \mu_2)^2}, \tag{7.2}
\]

\[
\rho = \frac{1}{p_1 \mu_1 + p_2 \mu_2}, \tag{7.3}
\]

\[
R = \frac{\rho \ell}{A} = \frac{\ell}{A} \frac{1}{p_1 \mu_1 + p_2 \mu_2}. \tag{7.4}
\]

\(p\) is the carrier concentration. Subscripts 1 and 2 denote the different hole bands.

If we substitute the electron heat conductivity from equation (3.50) into our heat conduction equation (3.45) and then use the Hall coefficient, resistivity, and resistance relations I just introduced, we calculate the minimum detectable field to be

\[
B_{min} = C_\alpha \frac{p_1 \mu_1 + p_2 \mu_2}{p_1 \mu_1^2 + p_2 \mu_2^2} \sqrt{\frac{1}{\kappa_{ph} + C_\beta(p_1 \mu_1 + p_2 \mu_2)}}, \tag{7.5}
\]

where \(C_\alpha\) and \(C_\beta\) are parameters independent of changes in carrier concentrations and \(\kappa_{ph}\) is mostly independent of carrier concentrations\(^2\). We should be careful about how we optimize \(B_{min}\), though. To find the optimum balance between bands 1 and 2, if there is one, we should find the derivative of \(B_{min}\) with respect to, say, band 1 while keeping the total number of carriers constant. From here, the math is straightforward, but messy.

The main result is that single-band material gives superior sensitivity compared to any of the two-band carrier configurations. For systems with both electrons and holes, the reason for this is clear: electrons and holes tend to cancel each others' influence. For systems with two hole bands or two electron bands, there is an interesting little kernel of insight. The band with the higher mobility should be the one that wins out. We saw in the single band picture that this corresponds to a larger carrier mean free path and/or Fermi wavelength since \(\mu = e \Lambda \lambda_F / \hbar\) from equation (3.36).

I refer the interested reader to Appendix A.3 for more calculational details. The very interested reader may even solve the general case for an arbitrary number of bands using Lagrange multipliers.

\(^2\)The phonon component of heat conductivity should not change much unless the material is drastically changed, say, with a massive amount of doping.
7.3 Optimal Doping and Bismuth Band Structure

So we want to dope to a one-band material. How feasible is that? Will doping bismuth to one band remain optimal for more than just low temperature use? To answer such questions, we need to know a bit about the band structure of bismuth. Figure 7.4 shows the band structure of bismuth as a function of energy.

With time and effort, we could probably compensate the conduction band with acceptor impurities, such as lead, without opening up a second hole band\(^3\). The gap between the conduction band and the possible second hole band is about 13.5 meV, according to Vecchi and Dresselhaus [66], so lowering the Fermi level into that gap would allow the conduction band to be repopulated with electrons at finite temperatures. At low temperatures, though, a doped bismuth probe could be made single-band. It may also be sufficient to be dominated by a single band of carriers.

Doping, however, introduces complications to the problem. At sufficiently high dopings or temperatures, bismuth goes from an effectively two-band material to a three-band material. To find an optimum doping for bismuth Hall probes so that they work at a given temperature range, we must model the behaviour of the system. Building such a model is the goal of the next section.

\(^3\)In hindsight, perhaps it would have been better to electron-dope and avoid the extra hole band complication altogether.
The starting point will be a Drude model of all charge carriers in the Hall probe, with carrier concentration and mobility as the variables. We will account for temperature and doping-dependent changes in carrier concentration using Fermi-Dirac statistics and the Fermi surface of bismuth. The temperature-dependence of mobility will be correlated with different scattering regimes.

### 7.4 A Three-Band Model of Bismuth

As mentioned before, at the right temperature or doping, bismuth becomes effectively three-band. In addition to the existing heavy hole and light electron bands, a light hole band also arises.

We can derive our model briefly from Drude theory. Start with the steady state force balance for a particular charge carrier with isotropic scattering and solve for $v$:

$$\frac{m v}{\tau} = qE + qv \times B \quad (7.6)$$

$$v = \frac{q\tau}{m} (E + v \times B) = \mu (E + v \times B) \quad (7.7)$$

where $v$ is the drift velocity, $m$ is the mass, $B$ is the magnetic field, $E$ is the electric field, $\tau$ is the mean time between collisions, and $q$ is the charge.

We know the solution for $c$ to an equation of the form $c = a + c \times b$ is

$$c = \frac{a + a \times b + b(a \cdot b)}{1 + b^2} \quad (7.8)$$

So then we get

$$v = \frac{\mu E + \mu^2 E \times B}{1 + \mu^2 B^2} \quad (7.9)$$

This applies to each band in the model, so we can then get the current density by summing over the contributions from each band:

$$J = \sum_i n_i e v_i = \sum_i \left( \frac{n_i e \mu_i E}{1 + \mu_i^2 B^2} + \text{sgn} \, q_i \frac{n_i e \mu_i^2 E \times B}{1 + \mu_i^2 B^2} \right) \quad (7.10)$$

Note that the sign of charge $q_i$ changes the sign of the second term because hole and electron carriers produce conventional currents in different directions.

Assuming $B$ is in the $\hat{z}$ direction, we can rewrite this as a matrix equation of the form

$$J = \sigma E \quad (7.11)$$
where $\sigma$ is the magnetoconductivity matrix:

$$\sigma = \begin{bmatrix} \sigma_L & \sigma_H B \\ -\sigma_H B & \sigma_L \end{bmatrix} \quad (7.12)$$

with the following definitions of $\sigma_L$ and $\sigma_H$:

$$\sigma_L = \sum_i \frac{n_i e \mu_i}{1 + \mu_i^2 B^2} \quad (7.13)$$

$$\sigma_H = \sum_i \text{sgn} q_i \frac{n_i e \mu_i^2}{1 + \mu_i^2 B^2} \quad (7.14)$$

$\sigma_L$ is a longitudinal conductivity term while $\sigma_H B$ is the Hall conductivity. The usual formulation of Hall probe physics is a resistivity tensor, though, which is just a straightforward transformation of the conductivity tensor:

$$\mathbf{E} = \rho \mathbf{J} \quad (7.15)$$

$$\rho = \begin{bmatrix} \rho_L & -R_H B \\ R_H B & \rho_L \end{bmatrix} = \sigma^{-1} \quad (7.16)$$

$$\rho_L = \frac{\sigma_L}{\sigma_L^2 + \sigma_H B^2} \quad (7.17)$$

$$R_H B = \frac{\sigma_H B}{\sigma_L^2 + \sigma_H^2 B^2} \quad (7.18)$$

$$\sigma_L = \frac{\rho}{\rho^2 + R_H^2 B^2} \quad (7.19)$$

$$\sigma_H B = \frac{R_H B}{\rho^2 + R_H^2 B^2} \quad (7.20)$$

where $\rho_L$ and $R_H$ are the longitudinal resistivity and Hall coefficient, respectively.

In the low-field limit for bismuth where magnetoresistance is negligible ($\mu_i B \approx 0$ for all bands), we can use equations (7.13), (7.14), (7.17), and (7.18) to recover the perhaps familiar multi-band resistivity and Hall coefficient expressions:

$$\rho = \frac{1}{e(p_1 \mu_1 + p_2 \mu_2 + n \mu_n)} \quad (7.21)$$

$$R_H = \frac{p_1 \mu_1^2 + p_2 \mu_2^2 - n \mu_n^2}{e(p_1 \mu_1 + p_2 \mu_2 + n \mu_n)^2} \quad (7.22)$$

where $p$ and $n$ are carrier concentrations for holes and electrons. $\mu$ is a mobility. The subscripts 1, 2, and $n$ refer to the numbering of hole bands and the electron band.
We want to be able to calculate the carrier concentrations, look at $\rho$ and $R_H$ data as a function of temperature and doping, and infer the $T$-dependent behaviour of $\mu$ to understand the system.

### 7.4.1 Crystallite Size and Orientation

It was glossed over a bit in the Drude theory derivation, but I explicitly made an assumption of an isotropic model. I would like to justify that here with the help of some X-ray diffraction data and a bismuth film SEM image that appear to show incomplete crystal alignment. These data are also useful for determining the crystallite sizes, which we will do briefly.

As we know, for single crystals, the magnetotransport characteristics of bismuth are determined by the Fermi surface and include anisotropic behaviour. Bismuth evaporated onto silicon is polycrystalline, as we first saw in SEM images from Chapter 6. X-ray diffraction on undoped bismuth reveals a diffraction pattern of a somewhat-oriented film. As seen in figure 7.5, the diffraction peaks largely correspond to (00$\ell$) in a hexagonal unit cell so that the three-fold symmetry axis is perpendicular to the substrate. It is not apparent, though, from SEM images that there is other alignment of the polycrystals. In a system of non-aligned polycrystals, an isotropic approximation would work.

From the diffraction data, the Scherrer formula can be used to determine the grain size, $\tau$, in the out-of-plane direction from the Scherrer constant, $K = 0.9$, the X-ray wavelength, $\lambda$, the line-broadening at full-width half-maximum, $\beta$, and the X-ray incident angle, $\theta$:

$$\tau = \frac{K\lambda}{\beta \cos \theta} \quad (7.23)$$

We can correct for instrumental broadening using the silicon substrate data using $\beta = \beta_M - \beta_{Si}$. We also need to correct for the broadening due to there being two wavelengths of X-rays from the copper K-alpha 1 and K-alpha 2 emissions. A quick approximation would be to calculate the angles at which we expect to see X-ray intensity peaks for a particular diffracting plane separation and take the difference of angles to be the non-monochromatic X-ray broadening. Subtracting that off gives us a rough estimate of the line-broadening from finite grain size.

Using the data in table 7.2, we get a grain size of about 11 nm for the 22.54° peak (Bi 003) and 5 nm for the 45.91° peak (Bi 006). The film is about 65 nm thick. The actual grain size may vary since the calculation does not factor in broadening due
Figure 7.5: X-ray diffraction graphs for a silicon substrate (top) and undoped bismuth on silicon (bottom). Bismuth peaks are identified from Sailer and McCarthy’s data in the Powder Diffraction File [67] while the silicon peak information comes from Swanson’s data in the Powder Diffraction File [68].
to strain. A more careful correction for non-monochromatic X-rays may also offer additional insight.

Table 7.2: X-ray diffraction line-broadening parameters for out-of-plane grain size calculation. The X-ray source is primarily copper K-alpha 1 ($\lambda = 1.54060$ Å), with significant intensity of copper K-alpha 2 ($\lambda = 1.54443$ Å).

<table>
<thead>
<tr>
<th>$\beta_M$ (degrees)</th>
<th>$2\theta$ Angle (degrees)</th>
<th>Non-Monochromatic Broadening (degrees)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.22</td>
<td>22.54 (Bi)</td>
<td>0.057</td>
</tr>
<tr>
<td>0.30</td>
<td>45.91 (Bi)</td>
<td>0.121</td>
</tr>
<tr>
<td>0.15</td>
<td>69.15 (Si)</td>
<td></td>
</tr>
</tbody>
</table>

To get further information on the crystal alignment, we can look at SEM images, such as in figure 7.6 and observe that there is no clear in-plane orientation for the crystallites, which vary in size between 50–100 nm in the plane of the film. While there is no apparent difference between the crystallites of the two images of different dopings, it is possible that if lead phase-separates from the bismuth, we could see different shapes of crystallites or a stronger differential emission of electrons.

7.4.2 Carrier Concentrations

Going back to the Drude model explored earlier, it would be useful to estimate the carrier concentrations in doped bismuth. To do that, we need to know about the Fermi surface and band structure of bismuth, which will be the subject of this subsection.

According to Jones and Shoenberg [53, 69], the Fermi surface of bismuth consists of a heavy hole ellipsoid surrounded by three tilted light electron ellipsoids arranged symmetrically around the hole pocket. In undoped bismuth, the Fermi energy is set such that the number of electron carriers is equal to the number of holes. With sufficient doping or heating, light hole pockets will be created near the light electron pockets.

Lax’s Model of the Bismuth Fermi Surface

To do calculations, we need a mathematical model of bismuth’s Fermi surface. The work of Jones and Shoenberg has since been built on by Lax et al. [70], who introduced the standard closed-form approximation for the shape of a bismuth electron pocket, which has an ellipsoid-like form, known as the Lax model or ellipsoidal non-parabolic model:

$$E_n \left(1 + \frac{E_n}{E_g}\right) = \frac{\hbar^2}{2} \mathbf{k} \cdot \mathbf{M}^{-1} \cdot \mathbf{k},$$

(7.24)
Figure 7.6: SEM image of crystallites in patterned bismuth thin films. On top is a nominally undoped sample while the bottom has a nominal 0.1% lead doping.
where $E_n$ is the electron energy, $E_g$ is the energy gap between the light holes and electrons, $M_n^{-1}$ is the inverse of the electron effective mass tensor, and $k$ is the wavevector.

The heavy hole pocket is usually approximated as an ellipsoid:

$$E_{p1} = \frac{\hbar^2}{2} k \cdot M_{p1}^{-1} \cdot k,$$  

(7.25)

where the $M_{p1}^{-1}$ denotes the heavy hole inverse mass tensor.

We can obtain the density of states for each carrier type using these equations for Fermi surface shape to calculate the $k$-space volume occupied\(^4\), which is proportional to the number of states, and taking the derivative with respect to energy. Assuming these forms for the Fermi surface shapes, we get the following density of states for the electrons and heavy holes:

$$D_n(E) = C_1 \frac{d}{dE} \left[ E \left( 1 + \frac{E}{E_g} \right) \right]^{3/2},$$  

(7.26)

$$D_{p1}(E) = C_2 \sqrt{E},$$  

(7.27)

where $C_1$ and $C_2$ are constants. The density of states for the light holes is analogous to that for electrons, with the same constant.

<table>
<thead>
<tr>
<th>Table 7.3: Selected bismuth band and transport parameters.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_1$</td>
</tr>
<tr>
<td>$C_2$</td>
</tr>
<tr>
<td>$\mu_1$ at 90 K, about 40 nm thick</td>
</tr>
<tr>
<td>$\mu_2$ at 90 K, about 40 nm thick</td>
</tr>
<tr>
<td>$m_1$ for $H \parallel$ bisectrix, bulk</td>
</tr>
<tr>
<td>$m_2$, $m_n$ for $H \parallel$ bisectrix, bulk</td>
</tr>
</tbody>
</table>

**Carrier Concentration Calculations**

Ideally, the concentration of each carrier type in bismuth at a temperature $T$ is calculated by integrating the density of states with charge carriers distributed according

\(^4\)The volume of an ellipsoid defined by an equation of the form $k \cdot A \cdot k = 1$ is given by $\frac{4}{3} \pi \sqrt{\text{det}(A^{-1})}$. 121
to the Fermi-Dirac function\(^5\):

\[
n_i = \int_{-\infty}^{\infty} \frac{D_i(E)}{e^{(E-\mu)/k_BT} + 1} \, dE,
\]

(7.28)

where \(n_i\) is the carrier density for band \(i\), \(D_i(E)\) is the density of states for band \(i\), \(E\) is energy, \(\mu\) is the chemical potential, \(k_B\) is Boltzmann’s constant, and \(T\) is temperature. In practice, I integrate from the bottom of the band to \(10^7 \, k_B T\) to save time. The upper limit of integration is orders of magnitude above the chemical potential for the temperatures used in the calculations, so error from not doing an infinite integral is negligible.

For a concrete example of how the calculation works, let us look at the heavy hole band at \(T = 0\). The carrier density is given by

\[
p_1 = \int_0^{E_F} D_{p1}(E) \, dE = \frac{2}{3} C_2 E_F^{3/2}.
\]

(7.29)

The constant \(C_2\) can be calculated using known carrier concentrations\(^6\) in pure bismuth at low temperature. The concentrations of carriers in other bands are calculated similarly. The carrier densities of all bands are related to the density of electron acceptors \(N_A\) in the material by

\[
p_1 + p_2 - n = N_A.
\]

(7.30)

In my calculations, I chose particular values of \(N_A\), then varied \(p_1\), \(p_2\), and \(n\) by tuning the chemical potential until equation (7.30) was true.

For my particular calculations, I have used \(E_g = 13.5\) meV from Vecchi and Dresselhaus [66], along with \(E_n = 30\) meV and \(E_{p1} = 12\) meV from Shoenberg pp. 232–233 [72].

We should bear in mind the limits of such calculations, though. The measurements of band structure parameters were done with pure bismuth, and the model may not hold for significant dopings. In addition, Shoenberg, in reviewing the bismuth band parameters for section 5.3.5 of his book [72], notes that the error in the energy gap may be as large as 10%.

\(^5\)For holes, the signs of the energies and chemical potential are flipped since we are dealing with carriers of positive instead of negative charge.

\(^6\)This turns out to be \(p_1 = n = 3.015 \times 10^{23} \, \text{m}^{-3} \) [71]. The values of \(C_1\) and \(C_2\) I used are in table 7.3.
7.4.3 Mobility and Scattering

So far, we have examined a theoretical account of how resistivity and Hall coefficient are influenced by carrier concentrations and mobilities. We also know how to calculate carrier concentrations as a function of temperature. Using resistivity and Hall coefficient data, we would then be able to figure out the behaviour of the mobilities of the three carrier types in bismuth.

In particular, the limiting behaviour of the mobilities as a function of temperature can tell us what scattering mechanisms are dominant in the bismuth films, allowing us to make adjustments to the fabrication procedure and optimize Hall probe sensitivity. With sufficient data across temperatures and dopings, we could create a “phase diagram” showing different scattering regimes.

In this section, we will look at the likely scattering mechanisms at work and identify the limiting temperature dependence for each kind of scattering.

Phonon Scattering

Lattice vibrations in bismuth can scatter electrons and holes. The vibrations are largely frozen out at low temperature, but become important at higher temperatures, as more phonons are excited. This decreases mobility as $\mu \sim T^{-3/2}$.

Grain Boundary Scattering

As an electron attempts to cross a grain boundary, it must scatter into an accessible electron Fermi surface pocket on the other side or else it will not cross. Random orientations of bismuth crystallites could easily make entire light electron pockets inaccessible to any crossing attempts, which greatly impedes light electron conduction. A similar situation arises for light holes.

Grain boundary scattering is not expected to have any temperature dependence, waxing and waning in relative influence as temperature changes decrease and increase the importance of other scattering mechanisms.

Charged Impurity Scattering

The lead dopants (impurities) in the bismuth films accept electrons, so they become charged and have greater ability to scatter charge carriers. The faster a carrier passes a charged impurity, the less of a scattering effect is exerted on the carrier. This becomes important when increasing temperature causes the speed of carrier particles
to increase significantly. However, this carrier speed effect is only significant when the Fermi level is near a band edge, basically when $k_B T \sim E_F$. In this regime, the mobility scales as $\mu \sim T^{3/2}$ for non-degenerate materials.

Table 7.4 shows Fermi temperatures for various carrier concentrations for bismuth, assuming the bands are rigid with doping.

Table 7.4: Concentrations of various carriers in bismuth as a function of band-specific Fermi temperature. These values were calculated from the Lax model band parameters, with each band separately filled to $E_F$ from the bottom of the band. $p_1$ and $p_2$ are the light and heavy hole bands, respectively.

<table>
<thead>
<tr>
<th>$E_F$ (meV)</th>
<th>$n$ or $p_1$ (m$^{-3}$)</th>
<th>$p_2$ (m$^{-3}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.3</td>
<td>$5 \times 10^{21}$</td>
<td>$6.5 \times 10^{22}$</td>
</tr>
<tr>
<td>8.6</td>
<td>$1.7 \times 10^{22}$</td>
<td>$1.8 \times 10^{23}$</td>
</tr>
<tr>
<td>17.2</td>
<td>$8 \times 10^{22}$</td>
<td>$5.2 \times 10^{23}$</td>
</tr>
<tr>
<td>25.9</td>
<td>$2.1 \times 10^{23}$</td>
<td>$9.5 \times 10^{23}$</td>
</tr>
</tbody>
</table>

Bismuth is not likely to see charged impurity scattering due to screening from its charge carriers. Phonon scattering and grain boundary scattering are probably important in bismuth films, though.

### 7.5 Doping Methods

The bismuth used came from two sources: Alfa Inorganics 1/2 inch rod with 99.999% pure bismuth (metals basis) and Alfa Products 2.5 mm shot with 99.99995% pure bismuth (metals basis). The lead flakes used were manufactured by BDH Chemicals, with impurities of antimony, tin, arsenic, bismuth, copper, iron, nickel, and silver at or below 10 ppm.

I melted together these materials to make bismuth evaporative sources with approximately 0%, 0.1%, 0.5%, 1%, and 4% nominal atomic percentages of lead. The films evaporated are all 65–70 nm thick.

#### 7.5.1 Melting in Sealed Glass Container

For this method, I placed bismuth and lead in a modified round-bottom flask with narrowed neck just large enough to allow bismuth pellets and lead flakes to enter. I pumped out the flask with a turbo pump to prevent oxidation of the material during the mixing. While pumping on the bismuth and lead, I used a heat gun to drive off some gases in the flask. It is not desirable to prematurely melt and evaporate the
bismuth during the pumping phase, though possible to melt the bismuth using heat
gun, as I discovered.

To close off flask, a propane torch softened the glass neck of the flask, which was
twisted and pulled away as it melted. I then left the flask in a furnace at 400 °C for
an hour.

This method was used on the 0.1 nominal at. % doping.

7.5.2 Melting in Vacuum Chamber

For this method, I placed the bismuth and lead in an evaporation boat inside a
vacuum chamber. The capacity is around 1.5 g of material. When chamber pumped
down to about $2 \times 10^{-6}$ Torr, I melted the mix and deposited at less than 1 Å/s for
30 minutes to mix the materials.

This method was used for all other doped materials.

7.6 Measures of Doping

Before we begin a detailed discussion the data, we should make careful distinctions
between all the different ways we can estimate the level of doping in bismuth Hall
probes. In particular, we should make distinctions between the concentration of
dopant atoms in the evaporative source, the concentration of dopant atoms in the
evaporated film, and the actual active acceptor concentration in the film.

The most straightforward estimate of doping is to figure out the concentration
of lead in the bulk doped bismuth material that is used to evaporate thin films.
In the thesis, I refer to this as the nominal dopant concentration or nominal lead
concentration. Of course, the source material can be significantly different from the
deposited material. The main influences are thermodynamic effects that cause partial
pressures of lead and bismuth to be different, which leads to an altered composition
of evaporated material. We saw earlier in section 7.1 that there could be anywhere
from two to four times the concentration of bismuth in the deposited films compared
to the evaporative sources.

Another measure of doping level is the concentration of dopant atoms in the
evaporated film. While this is a better measure of doping than the lead concentration
in the evaporative source, there are still problems with it. First, it is a difficult and
time-consuming task to figure out the dopant concentration in films. Secondary ion
mass spectroscopy (SIMS) uses an ion beam to eject ions for identification via mass
spectroscopy. While it has good ion and depth sensitivity, the ion milling process takes a long time and requires good calibration samples. Energy-dispersive x-ray (EDX) spectroscopy is another possibility, but also requires careful preparation of calibration films with known dopant concentrations. Due to time constraints, I did not pursue either of these techniques.

Finally, the measure of doping that matters the most is the film’s active acceptor concentration, or simply the acceptor concentration in this thesis, but is the most difficult to measure directly. Even if we know the lead concentration in a doped bismuth film, we do not necessarily know the degree of elemental phase separation in the film. If the lead phase-segregates out of the bismuth, the lead will not dope the bismuth like an isolated lead atom in a sea of bismuth atoms. When we use only the acceptor concentration in equations, we are excluding all the lead atoms that are inert as dopants due to chemical reactions, phase segregation, or some other means. Often, the acceptor concentration is deduced using a fit to a theoretical model.

With all of the above factors considered, the nominal dopant concentration may be significantly greater than the acceptor concentration. Nevertheless, as the main purpose of this study is to establish the optimal doping conditions for deploying bismuth as a Hall sensor, the empirical correlation between $R_H$ and nominal lead concentration is extremely useful.

### 7.7 Tools for Estimating Acceptor Concentration

While we have a band model that can be used to calculate the carrier concentrations in bismuth as a function of doping, we should have data-based checks that match up to guesses regarding acceptor concentration in the lead-doped bismuth films. One-band theory and Raoult’s law can be used to do this.

#### 7.7.1 One-Band Theory

One thing we can do is assume that there is only one dominant hole band in the material and infer an approximate total hole concentration from the Hall coefficient data. Hole concentration is not acceptor concentration, of course, but in the high doping limit, the two converge. This approximation can be significantly wrong if either there is a significant population of electrons or the minority hole carrier has a much larger mobility than the majority hole carrier. If electrons are significant, we could then use one-band theory to get an upper bound on the hole concentration.
Otherwise, one-band theory should give us a good approximate value for the total hole concentration.

Mathematically, we are making the following approximations:

\[
R_H = \frac{1}{e p_{\text{eff}}},
\]
\[
\rho = \frac{1}{e p_{\text{eff}} \mu_{\text{eff}}},
\]

(7.31)  
(7.32)

where \( p_{\text{eff}} \) is the effective total hole concentration.

We can then rearrange these expressions to get \( p_{\text{eff}} \) and \( \mu_{\text{eff}} \) (\( \mu_{\text{eff}} \) is included only for completeness). To make transparent what we are actually calculating with these approximations, let us also write them in terms of the concentrations and mobilities of the actual carriers in the bismuth:

\[
p_{\text{eff}} = \frac{1}{e R_H} = \frac{(p_1 \mu_1 + p_2 \mu_2 + n \mu_n)^2}{p_1 \mu_1^2 + p_2 \mu_2^2 - n \mu_n^2},
\]

(7.33)

\[
\mu_{\text{eff}} = \frac{R_H}{\rho} = \frac{p_1 \mu_1^2 + p_2 \mu_2^2 - n \mu_n^2}{p_1 \mu_1 + p_2 \mu_2 + n \mu_n},
\]

(7.34)

We see that this model breaks down when electrons and holes both have a significant influence on magnetotransport in the films. In particular, we would see “\( p_{\text{eff}} \)” go to infinity and “\( \mu_{\text{eff}} \)” go to 0 in the extreme case. We can still use the one-band model to look at the effective conductivities at low temperature to get an estimate of total carrier concentration in the system, comparing this to carrier concentrations calculated at various dopings, assuming we do not see signatures of model breakdown.

### 7.7.2 Raoult’s Law

While one-band theory can be used to check the acceptor concentration in bismuth films, it would be good to have an independent check on it. Raoult’s law works well for that. Using concentrations of lead and bismuth in the evaporative sources, we can estimate the lead concentration in the deposited films.

The main technique here is to estimate the relative amounts of lead and bismuth evaporated by looking at the partial pressures due to each element. Each partial pressure should be proportional to the number of atoms of each element. We use Raoult’s law, which simply states that the vapour pressure emanating from a mixed material is proportional to the fraction of each component multiplied by the vapour
pressure of each pure component. Mathematically, it is this:

\[ P = \sum_{i} P^*_{i} x_{i}, \]  
(7.35)

\[ 1 = \sum_{i} x_{i}, \]  
(7.36)

where \( P \) is the total pressure, \( P^*_{i} \) is the partial pressure of pure material \( i \), and \( x_{i} \) is the molar fraction of material \( i \) in the evaporative source. For our case, the two things being mixed are lead and bismuth, so we get

\[ P = P^*_{\text{Bi}} x_{\text{Bi}} + P^*_{\text{Pb}} x_{\text{Pb}}, \]  
(7.37)

\[ 1 = x_{\text{Bi}} + x_{\text{Pb}}. \]  
(7.38)

For practical considerations, it is really not so much a law as it is a linear approximation that works best at low dopings. In the context of our evaporating lead-bismuth mixture, the Raoult’s law assumption is that the interatomic and intermolecular forces between bismuth and lead are the same as those between bismuth and bismuth or lead and lead. This is not strictly true, of course, but should still be a good approximation, particularly because lead and bismuth have strong similarities in chemical properties.

### 7.8 Measurements and Data Processing

With our three-band model of bismuth established, we are now ready to process the data, which consists of a series of voltage-to-current ratios, done from 4 K to 300 K over many samples, which were all between 65 and 70 nm thick. Here, we will look at how to change the raw data to the conductivities that we desire, as well as estimate the doping levels.

#### 7.8.1 Resistivity and Hall Coefficient Measurements

For the resistivity measurements I did on my bismuth films, I used the van der Pauw method [73, 74], a four-point probe technique. The van der Pauw method can be applied to flat samples if the following conditions are met: the contacts are on the edge of the sample, the contacts take up much less area than the sample, the sample
thickness is uniform, the sample material is uniform, and there are no holes in the sample.

The standard measurement process for both resistivity and Hall coefficient is to take voltage-to-current ratios for many different permutations of contacts to calculate $\rho$ and $R_H$. An example of the full procedure is seen on web site of NIST’s Physical Measurement Laboratory [75]. For this thesis, let us consider only the simplified version used for my measurements, which omit some redundant measurements that were used as checks.

Figure 7.7: Numbering of contacts for van der Pauw method magnetotransport measurements.

Let us first orient ourselves to the geometry of the problem and introduce some notation. Suppose the sample is as shown in figure 7.7 with electrical contacts numbered as shown. We define $V_{a,b}$ as the voltage drop from contact $a$ to contact $b$ and $I_{c,d}$ as the current flowing from contact $c$ to contact $d$. Then we define the resistance-like $V/I$ quotient $R_{a,b,c,d}$ to be

$$R_{a,b,c,d} = \frac{V_{a,b}}{I_{c,d}}. \quad (7.39)$$

Van der Pauw found that resistivity can be calculated by finding the solution to the following equation:

$$\exp(-\pi R_{1,2,3,4}d/\rho) + \exp(-\pi R_{2,4,1,3}d/\rho) = 1, \quad (7.40)$$

where $d$ is the sample thickness. Basically, for one $R_{a,b,c,d}$ quotient, the $V$ and $I$ leads are aligned left-right; in the other configuration, the leads are aligned up-down. The equation can be solved using a computer, typically using Newton’s method.
I measured Hall coefficient by calculating the average of $R_{a,b,c,d}$ quotients in two configurations of leads at different magnetic fields and performing a linear fit in $B$ for the following equation:

$$\frac{R_{3,2,1,4} + R_{4,1,3,2}}{2} = \frac{R_H}{d} B + R_{\text{offset}},$$

(7.41)

where $R_{\text{offset}}$ is a constant that arises from misalignment of wires in the Hall probe junction. Keep in mind that $R_{3,2,1,4}$ and $R_{4,1,3,2}$ are measurements that change with magnetic field.

It is notable that the Hall coefficient for one configuration (say, the 3, 2, 1, 4 arrangement) can be better than that of another configuration (say, the 4, 1, 3, 2 arrangement); when using a probe for scanning, one should use the configuration that provides the stronger voltage response to a given magnetic field and excitation current.

### 7.8.2 Data Collection and Processing

Let us now look at how I collected data so that we can see what forms of pre-processing are necessary to properly analyze the data. I measured the $R_{a,b,c,d}$ quotients mentioned in the previous section using a continuous flow cryostat. The curious reader can find details of the cryostat system in Appendix B.

The $R_{a,b,c,d}$ quotients required for resistivity and Hall measurements were taken sequentially for each sample, with delays between measurements on the order of seconds. As a computer recorded the data, the temperature was swept from 300 K to 4 K and then back to 300 K, with about 10 hours taken for a sweep in each direction. The magnetic field was kept constant at $-1$, 0, or $+1$ T for each cycle of sweeping down and up in temperature.

The theory of how to calculate the resistivity and Hall coefficient assumes the various $R_{a,b,c,d}$ measurements are taken at the same temperature, but the measurements were done independently at different temperatures. To get data at matching temperatures, I interpolated the $R_{a,b,c,d}$ measurements as a function of temperature and performed a moving average to smooth out statistical fluctuations. Then I calculated resistivity and Hall coefficient as a function of temperature using the equations from section 7.8.1. The resulting graphs are shown in figure 7.8.

The resistivity and Hall coefficient measurements are not enough to solve for all variables in our model of bismuth, though. We have three bands of carriers, each
Figure 7.8: Processed data showing resistivity and Hall coefficient of bismuth films at various dopings. Kochowski and Opilski’s data on 66-nm-thick undoped bismuth for resistivity and a 64-nm-thick film for Hall coefficient are included for comparison [50].
with carrier concentrations and mobilities. We discussed how to calculate the carrier concentrations for the model, which leaves us with three unknowns: the mobilities. However, we only have two measurements: $\rho$ and $R_H$, which leaves the system underdetermined in general. It would be easier to figure out the role of each carrier if their contributions to the electrical properties were simply additive instead of more complicated quotients of terms, like $\rho$ and $R_H$.

The conductivity matrix elements we derived in section would be helpful. It is straightforward to use equations (7.19) and (7.20) to convert the $\rho$ and $R_H$ measurements into conductivities for easier data analysis. The resulting conductivity graphs are shown in figure 7.9.

For both sets of graphs, I have included data from Kochowski and Opilski for undoped bismuth thin films of similar thicknesses to the films I grew [50]. A couple of possibilities for the differences between my data and that of Kochowski et al. would be thickness measurement uncertainties and dopants contaminating the samples. The curve shapes are consistent with each other, though.

### 7.8.3 Acceptor Concentration Estimates

After processing the data, it would be helpful for the discussion to have some idea of the doping level of each sample. Let us pick up the tools we introduced earlier for estimating the acceptor concentration from section 7.7.

Given the vapour pressures of pure bismuth and lead in table 7.1, we can use Raoult’s law to estimate the proportion of lead in the evaporated film. For our lead-bismuth system, if the amount of lead is very small, the proportion of lead in the film is

$$x_{\text{film}}^\text{Pb} = \frac{P^*_{\text{Pb}}x_{\text{Pb}}}{P^*_{\text{Bi}}x_{\text{Bi}} + P^*_{\text{Pb}}x_{\text{Pb}}} \approx x_{\text{Pb}} \frac{P^*_{\text{Pb}}}{P^*_{\text{Bi}}}. \quad (7.42)$$

Looking at the values of the vapour pressures in table 7.1, we get estimates of film lead concentration anywhere from $1/5$ to $1/2$ of the nominal lead concentration. While evaporation temperature was not well-controlled for the bismuth film deposition, one would have to be rather unlucky to not catch the average evaporation temperature in the large range in the table.

The one-band method of estimating acceptor concentration can be applied to the low-temperature Hall coefficient data to get the values shown in table 7.5.
Figure 7.9: Processed data showing conductivities of bismuth films at various dopings. I interpolated Kochowski and Opilski’s data from two different samples to generate approximate conductivity curves for comparison [50].
Table 7.5: A comparison of various estimates of acceptor concentration in the doped bismuth films. Concentrations are expressed as a percentage of the number density of atoms in pure bismuth.

<table>
<thead>
<tr>
<th>Evaporator Source Nominal Dopant Concentration (%)</th>
<th>One-Band Data Inversion Hole Estimate of Acceptor Concentration (%)</th>
<th>Raoult’s Law Estimate of Acceptor Concentration (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.04–0.09</td>
<td>0.02–0.05</td>
</tr>
<tr>
<td>0.5</td>
<td>0.12–0.14</td>
<td>0.1–0.25</td>
</tr>
<tr>
<td>1</td>
<td>0.17–0.20</td>
<td>0.2–0.5</td>
</tr>
<tr>
<td>4</td>
<td>0.68</td>
<td>0.8–2</td>
</tr>
</tbody>
</table>

7.9 Discussion

Bit by bit, we have put together the theoretical machinery that we need to analyze the data. Now is a good time to take stock of what we have done and how it all works.

We have been working with a three-band model of bismuth, with three carrier concentrations and three mobilities. The carrier concentrations can be estimated using the Lax model of bismuth’s Fermi surface and the one-band approximation, but we only have two measurements ($\rho$ and $R_H$) and three temperature-dependent unknown mobilities that we want to find. By necessity, the mobility analysis will be somewhat qualitative.

For the 4% and 1% nominal dopings, I assume there are no electron carriers remaining due to the high concentration of lead impurities. This reduces the transport model to four parameters. I also assume that the carrier concentrations of the light and heavy holes are independent of temperature since the films are doped to the point of showing metal-like behaviour. So now the model has two temperature-dependent mobilities and two measurements. We can guess reasonable carrier concentrations and invert the data to get mobilities. My goal for those data sets was to find physically plausible mobility curves.

For the 0% and 0.1% nominal dopings, inverting the data is not possible because there likely are still electron carriers in those systems. For the 0.5% nominal doping measurements, inverting the data with reasonable constant carrier concentration guesses did not give plausible mobilities. It is possible that thermally-driven changes in carrier concentrations made the method not applicable.

So for the 0%, 0.1%, and 0.5% data sets, I adopted a systematic method of curve-matching to try to understand the physics involved. I calculated temperature-
dependent carrier concentrations using the Lax model and a rigid band approximation. I then set the three temperature-independent mobilities as a constant parameters and plotted the resulting conductivity graphs to see if the calculated curve shapes match those of the data. Ideally, the calculated curves would also be close in absolute magnitude — within a factor of 10.

7.9.1 4% Nominal Doping

The 4% nominal doping analysis is perhaps the most straightforward of the cases at hand. At high dopings, we expect electron band to be completely compensated and for thermal excitations to have negligible effect on carrier concentrations. The data in figure 7.11 are consistent with this picture.

Let us first just look at the data in the graphs. If electrons are compensated and mobilities are low, as you would expect from a heavily-doped sample, we can at least say from the longitudinal conductivity graph that the decrease is due to a decrease in either the balance of high and low mobility holes or due to overall decrease in hole mobility. The Hall conductivity graph seems to confirm that this is largely due to a decrease in hole mobility since Hall conductivity goes like $\mu^2$ and decreases more sharply than the $\mu$-scaling longitudinal conductivity at low temperature. The cause of the levelling off of the conductivities is unknown, but perhaps a more sophisticated analysis can shed some light on the matter.

With electrons compensated, we are left with four parameters in our problem as a function of temperature: carrier concentrations and mobilities for both the remaining hole bands. As mentioned before, the carrier concentrations can be calculated from Lax’s model of bismuth band structure, which brings us down to two undetermined parameters: the mobilities. Since we have two measurements $R_H$ and $\rho$ as a function of temperature with two unknowns, we can simply invert the data to get the mobilities for the light and heavy holes.

However, inversions using the Lax model carrier concentration calculations resulted in negative mobilities, so I used Table 7.5 to provide estimates of carrier concentration and let the hole carrier concentrations be adjustable temperature-independent parameters. I adjusted the hole carrier concentrations until I got a “good” inversion of the data, where the mobilities would be monotonically decreasing with temperature. There are few physical mechanisms that would produce an increase in mobility as a function of temperature, and none of them are suspected to be at work here. For instance, ionized impurity scattering can cause an increase in mobility with temper-
ature, but the 4% nominal doping sample should be doped well past the top of the light hole band, where we would expect the effect to be important.

When inverting the conductivities with hole carrier concentrations as parameters and no electrons, a family of plausible solutions emerges, all with high-mobility minority holes and low-mobility majority holes. An example of the results of the inversion is shown in figure 7.11. In the Lax model carrier concentration calculations, this corresponds to a high-doping regime with majority light holes and minority heavy holes, which seems to contradict the results from inversion. If the light holes were somehow more strongly affected by scattering off dopants, that might make sense. Another possibility is that the Lax model band parameters change significantly when going from zero doping to high doping, and so my assumption of rigid bands in calculating carrier concentrations may not be valid.

Considering just the inverted mobilities, we see that to a first approximation, the minority hole mobility does not change with temperature. If these are indeed the light holes, perhaps this is due to grain boundary scattering being dominant compared to phonon scattering. Then we could look at the majority hole mobility and say that those are the heavy holes, which are less affected by grain boundary scattering, but decrease mobility with temperature due to phonon scattering.

7.9.2 1% Nominal Doping

Like the 4% nominal doping, the 1% nominal doping is also expected to be largely free of electron carriers due to the large acceptor population. The shapes of the 1% conductivity data curves resemble those of the 4%, suggesting similar physics at work here. An interesting difference is the behaviour at high temperature, specifically an uptick in $\sigma_L$ as opposed to the slight downturn in $\sigma_H$. Recall that changes in carrier concentration tend to have stronger effects on $\sigma_L$ and changes in mobilities tend to have stronger effects on $\sigma_H$. Perhaps the increase in $\sigma_L$ is driven by movement of low-mobility holes to the high-mobility hole band while the $\sigma_H$ changes are caused by a decrease in mobilities with temperature.

As with the 4% nominal doping data, I performed an inversion using a model that has two hole bands with constant carrier concentrations. Again, using calculated Lax model carrier concentrations did not give good results. When I used the Lax model carrier concentrations, Mathematica was unable to invert the equations successfully. Instead, I set the hole carrier concentrations as temperature-independent parameters.
Figure 7.10: The conductivity data (left) for 4% nominally-doped films, shown with calculated conductivities (right) from inverting the data in a two-hole-band model to obtain mobilities using a hole concentration of $5.6 \times 10^{20}$ m$^{-3}$, 10% of which are high-mobility holes.

Figure 7.11: The mobilities from the 4% data inversion look reasonable, with monotonically decreasing mobilities.
and adjusted them by hand to see if anything plausible arose. An example of a plausible result is shown in figure 7.12.

Use of one-band theory estimates the carrier concentration to be 0.19%, which is well above the threshold for electrons becoming effectively compensated. Raoult’s law calculations indicate the carrier concentration should be around 0.2–0.5%. I used overall carrier concentrations in the same ballpark for my inversion calculations.

A typical mobility graph from that inversion is shown in figure 7.12. The curve shapes for the mobilities have a strong resemblance to those from the 4% nominal doping, with larger mobilities, of course. The graphs depict mobilities that increase with temperature in some parts, but I suspect this is from a limitation of my assumption of constant carrier concentrations. A small change in the balance of low-mobility and high-mobility holes may make the difference.

As in the 4% nominal doping case, the mobility curves would suggest that the high-mobility minority carriers are not affected much by temperature, possibly because of a high degree of grain boundary scattering. We also see a low-mobility majority carrier band that is strongly affected by temperature, with the decrease suggestive of phonon scattering.

### 7.9.3 0% Nominal Doping

The positive low-temperature Hall conductivities for some of the nominally undoped films suggest that these films may not actually be undoped. The bismuth may be doped by elements present in the evaporation chamber, which is made of aluminum, a likely hole dopant. Conducting posts in the chamber are made of copper. Other elements are likely present from past evaporations, including germanium and gold. Germanium is probably a hole dopant.

That said, the curves have the same shape as the available undoped bismuth Kochowski data, though, so perhaps the doping is not very high and the physics is not significantly changed.

Assuming mobilities do not increase, longitudinal conductivity increases certainly mean higher carrier concentrations from thermal excitations.

In undoped bismuth, the number of holes equals the number of electrons at all times. At low temperature, the electrons’ mobility is certainly higher than that of the heavy holes since the Hall conductivity is negative. As temperature increases, Lax model calculations indicate that light holes increase in influence from 0% of holes at low temperature to about 10% of holes at 293 K. It seems plausible that this
Figure 7.12: The conductivity data (left) for 1% nominally-doped films, shown with calculated conductivities (right) from inverting the data in a two-hole-band model to obtain mobilities using a carrier concentration of $1 \times 10^{26} \text{ m}^{-3}$, 26% of which are high-mobility holes.

Figure 7.13: The mobility curves from the inversion look mostly reasonable, though “increases” in mobility here may come from inaccuracies from assuming constant carrier concentration.
drives at least part of the increase in Hall conductivity. At high temperature, Hall conductivity levels off while longitudinal conductivity continues to increase steadily. Neglecting the magnetoresistance terms, Hall conductivity goes like \( \mu^2 \), so is more sensitive to mobility changes. I suspect mobility changes are behind the levelling off of Hall conductivity at high temperature.

Unfortunately, with a three-band system here, inverting the data is no longer an option because the equations are underdetermined. So I shift to trying to match the conductivity curves using some simple assumptions. To see if the above situation is at least plausible, I used Lax model carrier concentration calculations along with constant carrier mobilities. The results are shown in figure 7.14. There is a correspondence between the calculated curves and the data at low temperature. For high temperature, I suspect a decrease in the heavy hole mobility drives the levelling off of the Hall conductivity and the decrease in longitudinal conductivity slope. It is also possible, though, that light hole mobility is decreasing significantly with temperature.

To summarize, a model with constant light hole and electron mobilities and decreasing heavy hole mobility seems to generate the same behaviour as the data. This would correspond with a picture where grain boundary scattering dominates with light holes and electrons, but where phonon scattering is most important for heavy holes.

### 7.9.4 0.1% Nominal Doping

In the 0.1% nominal doping data we see in figure 7.16, we see the conductivity data. Like we saw for the undoped film data, we see a steady increase in \( \sigma_L \) in temperature, driven by increases in total carrier population.

Interestingly, the curves appear to be slightly offset from the 0% nominal doping graphs. It is tempting to conclude that the physics is likely similar for the two cases, but that is not certain. It is particularly intriguing that the \( \sigma_L \) graphs for the 0.1% doping are lower than those for the 0% doping while the \( \sigma_H \) graphs for the 0.1% doping are higher than those for the 0% doping. One neat and tidy explanation for this would be the removal of high-mobility electrons and partial replacement with holes, many of which are likely lower-mobility. Another explanation would be that scattering off dopants has decreased mobilities across the board, decreasing \( \sigma_L \), but compensating some electrons has increased \( \sigma_H \).

I do not think, though, that electrons have a significant influence here. According to Lax model carrier concentration calculations, electrons are largely compensated
Figure 7.14: $\sigma_L$ and $\sigma_H$ graphs for undoped bismuth films. The calculated curves use
$\mu_1 = 0.9 \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$, $\mu_2 = 0.395 \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$, $\mu_n = 0.400 \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$, and Lax model
carrier calculations assuming no doping. At high temperature, the calculated curves show
different behaviour from the data, but this is likely due to a breakdown of the constant
mobility assumption.
below 100 K by a replacement of 0.02% of bismuth atoms with lead, and we see differences across all temperatures. At best, we should see a few percent electron carriers at high temperature, and the electrons’ importance is limited to that regime. Even a generous reading of the one-band or Raoult’s law estimates indicates that electrons do not play a large role in making the 0.1% nominal doping samples different from the undoped ones. It is possible that there are still electrons in the system, though. Using one-band theory to calculate $p_{\text{eff}}$ as a function of temperature, we see relatively large variations for the 0.1% nominal doping samples in figure 7.15, as may be expected for a sample with electron carriers.

![Film Data: One-Band Model Estimates](image)

**Figure 7.15:** $p_{\text{eff}}$ as a function of temperature for bismuth films using the one-band model.

We see in the data an increase in $\sigma_H$ with temperature that does not appear to be entirely explainable by the electron mobility decreasing. I think that this increase in $\sigma_H$ is caused by the increase in the relative population of light holes. If the 0.1% doping has not changed the physics of the material significantly, this is also what is happening in the undoped material.

An attempt at matching the curve shapes of the data is also shown in figure 7.16. I was able to approximate the shapes of the curves over the whole temperature range with Lax model carrier concentration calculations and constant mobilities for all carriers. For my curve matching, I was able to adjust the electron mobility so that the rising concentration of electrons at high temperature caused the Hall conductivity to level off at high temperature. This might happen or mobilities could decrease with temperature or both. It is hard to tell.
Figure 7.16: $\sigma_L$ and $\sigma_H$ graphs for 0.1% nominally-doped bismuth films. The calculated curves use $\mu_1 = 0.26 \, \text{m}^2 \, \text{V}^{-1} \, \text{s}^{-1}$, $\mu_2 = 0.06 \, \text{m}^2 \, \text{V}^{-1} \, \text{s}^{-1}$, $\mu_n = 0.31 \, \text{m}^2 \, \text{V}^{-1} \, \text{s}^{-1}$, and Lax model carrier concentrations assuming 0.02% acceptor concentration, at the lower end of the Raoult’s law estimate.
7.9.5 0.5% Nominal Doping

The data for the 0.5% nominal doping films in figure 7.17 are very interesting. To a first approximation, the conductivities are unchanging with temperature. One would expect electrons to be completely compensated here, even taking into account our imprecise measures of doping in these experiments.

My guess is that the hole carriers are redistributing from the heavy holes to the light holes while one or both of the hole mobilities are decreasing with temperature. That would account for \( \sigma_L \) holding steady while \( \sigma_H \) increases slowly. I am not sure how to explain the increase in \( \sigma_L \) and levelling off of \( \sigma_H \) at high temperature. It is possible that the Lax model band parameters I used are not good for this level of doping and temperature, but hard to tell without further experiments. In any case, as long as mobilities are not increasing, we must be getting more thermally-excited carriers.

My curve-matching attempts here are perhaps the least convincing of them all. My calculations resulted in very weakly temperature-dependent curves that compare well enough at low temperature, but deviate significantly from the data at high temperature. It would be great to have some more data to flesh out what is happening.

7.9.6 Overall Discussion

Now having looked at each data set in detail, now let us look at all the data sets together so that we can come to some tentative conclusions. The parameters used are summarized in table 7.6.

The most interesting thing to me is how similar the longitudinal and Hall conductivity experiment graphs look for the 0%, 0.1%, and 0.5% nominal acceptor concentration data, as shown in figure 7.9. For the 0% and 0.1% Hall conductivity experiments, the graphs even look like simple offsets of each other. Given how the 0.1% nominal acceptor concentration samples likely have electrons compensated beyond any importance and that high-mobility carriers are more important for Hall conductivity, I expect the cause of the offset is mostly replacement of electron carriers with light holes. The uniformity of the offset with temperature suggests that the mobilities of electrons and light holes are similarly uniform in temperature, probably due to grain boundary or edge scattering effects.

The highly-doped end, with 1% and 4% nominal acceptor concentration, appears to be composed of two bands: majority heavy holes and minority light holes. The heavy holes have decreasing mobility with respect to temperature while the light holes
Figure 7.17: $\sigma_L$ and $\sigma_H$ graphs for 0.5% nominally-doped bismuth films. The calculated curves use $\mu_1 = 0.03 \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$, $\mu_2 = 0.001 \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$, $\mu_n = 0.1 \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$, and Lax model carrier concentrations assuming 0.1% acceptor concentration.
Table 7.6: Parameters used to generate graphs from the doped film magnetotransport data analysis. These are by no means the only possible parameterizations of the data. They are examples of values that lead to reasonable-looking results.

<table>
<thead>
<tr>
<th>Doping</th>
<th>Carrier Concentrations</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>From Lax model, 0% acceptors</td>
</tr>
<tr>
<td>0.1%</td>
<td>From Lax model, 0.02% acceptors</td>
</tr>
<tr>
<td>0.5</td>
<td>From Lax model, 0.1% acceptors</td>
</tr>
<tr>
<td>1%</td>
<td>Holes: $1 \times 10^{26} \text{ m}^{-3}$, 26% high-$\mu$; $n = 0$</td>
</tr>
<tr>
<td>4%</td>
<td>Holes: $5.6 \times 10^{26} \text{ m}^{-3}$, 10% high-$\mu$; $n = 0$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Doping</th>
<th>Mobilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>$\mu_1 = 0.9 \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$, $\mu_2 = 0.395 \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$, $\mu_n = 0.400 \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$</td>
</tr>
<tr>
<td>0.1%</td>
<td>$\mu_1 = 0.26 \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$, $\mu_2 = 0.06 \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$, $\mu_n = 0.31 \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$</td>
</tr>
<tr>
<td>0.5%</td>
<td>$\mu_1 = 0.03 \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$, $\mu_2 = 0.001 \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$, $\mu_n = 0.1 \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$</td>
</tr>
<tr>
<td>1%</td>
<td>Graphed from inverting data</td>
</tr>
<tr>
<td>4%</td>
<td>Graphed from inverting data</td>
</tr>
</tbody>
</table>

have nearly constant mobility at all measured temperatures. This would be in line with the expectation that light holes, like the electrons with their highly anisotropic Fermi surface, would be highly susceptible to grain boundary scattering or would be swamped by impurity scattering at high doping. It is also possible that the light holes are more susceptible to impurity scattering than heavy holes such that we get the same results.

The most interesting case is if the most important carrier for Hall conductivity is indeed in a regime where grain boundary scattering dominates. Remember from section 3.6.4 that we want to be able to tune the mean free path of carriers to around the size of the Hall probe active region. If grain boundary scattering dominates, we can do this simply by varying the substrate temperature during evaporation. The evidence suggests that we can, at least for low dopings, like 0.1% nominal acceptor concentration.

Indeed, we may not need to dope to 0.1% nominal acceptor concentration. We just need to compensate one type of carrier, and depending on the concentration of lead atoms that are actually active acceptors, that may not even require 0.1%. According to theory, Hall probe sensitivity is constrained by length scales, and doping beyond what is required to remove cancellation of the Hall effect while satisfying constraints is going to be counterproductive. Further doping will decrease the Fermi wavelength more than required and decrease the Hall probe sensitivity.
Compensating the holes with electron donor atoms would be an interesting alternative that avoids the complication of two active hole bands in the probe. Tellurium, with its high vapour pressure relative to bismuth at a given temperature, would be a candidate for doping via co-evaporation. Tanaka, for instance, has done some interesting work on tellurium-doped bismuth single crystals [76].
Chapter 8

Bismuth Hall Probes

“If it weren’t for the last minute, nothing would get done.”
— Rita Mae Brown.

The previous chapters have laid lots of groundwork for us to understand Hall probes. In Chapter 3, we built a model to understand Hall probes in general. In Chapter 5, we learned how to make conventional probes. In Chapter 7, we put together a framework to understand how doping affects magnetotransport properties of Hall devices. We are finally ready to look at some magnetotransport data on a couple of Hall probes and also a Hall probe scan I performed. More work is required to comprehensively characterize the parameter space of bismuth Hall probes, but the Hall sensor data examined here is a helpful beginning to such efforts.

We will analyze magnetotransport measurements to compare the following: unpatterned bismuth films and patterned Hall probes, both doped and undoped; bismuth films with and without a germanium capping layer, for both doped and undoped bismuth; and bismuth Hall devices with and without a germanium capping layer, for both doped and undoped probes.

The probes were fabricated according to the procedures outlined in Chapter 5. The films and the probe patterns discussed in this section all have thicknesses in the range 65–70 nm.

8.1 Hall Probes: Films and Patterned Sensors

I would expect the magnetoconductivity data for thin films and patterned sensors to not differ much from each other. Both patterned sensors have active regions
Figure 8.1: Top: SEM micrograph of the active region of Hall probe 120229P6C2. Bottom: SEM micrograph of probe 120529P6’s active region.
approximately $320 \times 320 \text{ nm}^2$. The physics problem from a film to a patterned probe does not change much, except the geometry may reduce the mean free path of some carriers.

In other words, graphs of longitudinal conductivity and Hall conductivity should be similar between films and probes, though we should be careful to not put too much weight in these results, which come from a single undoped probe and just one doped probe. Any number of anomalies may have struck the samples, including contamination of the pure bismuth source, contamination of the substrate, or some other unnoticed variation in the deposition procedure. The deposition chamber was prone to flaking of films stuck to internal surfaces. Reproduction of these results would be very helpful.

### 8.1.1 Discussion for Films and Sensors

Now we are ready to look at the graphs in figures 8.2 and 8.3, which show longitudinal and Hall conductivity as a function of temperature.

We see in figure 8.2 that for an undoped probe and film, the longitudinal conductivity appears to be the same. The Hall conductivity graph shows the film having a lower value at high temperature, though they are still within the same order of magnitude. The overall curve shapes are still the same. We should analyze the data with a bit of wariness because we only have measurements on one probe each for undoped and doped bismuth.

Consider the nominally undoped material first. Longitudinal conductivities are basically the same, though there appears to be a small difference in Hall conductivity, though the film and the probe Hall conductivities are still within the same order of magnitude. Assuming no major contamination issues, carrier concentrations should be similar between the film and the probe. That would imply that in the probe compared to the film, either the mobility of electrons is dropping significantly with temperature, the mobility of holes is increasing significantly with temperature, or both. It would be helpful to do further experiments, decreasing or increasing the size of the Hall probe active region to see if there is a consistent pattern to the changes in magnetotransport properties. Preferably, bismuth would be deposited for probes and films simultaneously for easy direct comparison.

In the 0.1% nominal acceptor concentration data, the curves largely resemble each other, as expected. The probe Hall conductivity dips slightly at high temperature compared to the film. If this is a significant difference, in our model, this would appear
to be a hole mobility decrease at high temperature since electrons should mostly be compensated.

### 8.2 Effects of Germanium Capping Layer

As mentioned before, the purpose of the germanium capping layer is to protect the Hall probe from any debris that may be on a surface scanned by the probe.

Evaporating germanium on top of a bismuth Hall probe may substantially change the system in several ways that we should consider: the germanium may conduct, the bismuth may be doped, the bismuth may be annealed.

While pure bulk germanium is a semiconductor and should not affect the conduction of a Hall probe substantially, a hot germanium film in contact with bismuth may become doped and provide a competing conduction channel for electrons moving through the Hall probe. This should be detrimental to the performance of a Hall probe. Indeed, a major reason for using bismuth is to avoid the electronic noise that arises from doped semiconductors. If the germanium were conducting well, we would expect germanium deposition to increase the longitudinal conductivity due to the additional conduction channel. The deposited germanium, which is electron-dominated over a wide temperature range [77], would decrease the Hall coefficient.

If the germanium dopes the bismuth, we would naïvely expect it to hole-dope bismuth. For undoped bismuth, this should increase the Hall conductivity and perhaps decrease the longitudinal conductivity due to compensating some high-mobility electrons. For bismuth with 0.1% nominal acceptor concentration, we would expect increased longitudinal conductivity and increased Hall conductivity due to extra hole carriers providing more conduction.

The bismuth may also be unintentionally annealed from germanium evaporation since it has low vapour pressure unless a great deal of heat is applied. For carriers that are constrained by grain boundary scattering, this should increase the mobilities of the affected carriers. If my suspicions about bismuth are correct, this would mean the light holes and electrons should see greater mobilities. For undoped bismuth, we should see a decreased Hall conductivity and increased longitudinal conductivity since the electrons dominate. For the film doped to 0.1% nominal acceptor concentration, we would expect increased Hall conductivity and increased longitudinal conductivity since the light holes dominate.
Figure 8.2: A comparison of magnetotransport properties for a film and a probe made with undoped bismuth.
Figure 8.3: A comparison of magnetotransport properties for a film and a probe made with 0.1% nominal acceptor concentration bismuth.
In the event that these mechanisms adversely affect bismuth Hall probe performance, there are at least a few possible remedies. Intermixing of germanium and bismuth can be prevented by putting a passive layer between the two (sputtered SiO₂, for instance), by using a different capping material altogether, or by cooling the Hall probe substrate with liquid nitrogen during germanium deposition so that the atoms cannot diffuse very far.

### 8.2.1 Discussion for Germanium-Capped Bismuth Films

Let us now look at some results from bismuth films that I capped with 100 nm of thermally-evaporated germanium.

Consider the undoped bismuth film data in figure 8.4. The differences between the patterned and unpatterned films are small. The germanium deposition decreases the longitudinal conductivity, and more so at low temperature. The germanium capping layer also slightly decreases Hall conductivity at low temperature and increases it a bit at high temperature in one sample. From the mechanisms outlined above, only germanium hole-doping would decrease longitudinal conductivity.

The Ge layer causes a decrease in Hall coefficient at low temperature that is not consistent with a picture where germanium hole-dopes bismuth, though annealing of the bismuth may have decreased the Hall coefficient. The germanium may also be slightly conductive, which would also decrease the Hall coefficient. However, Hall conductivity for the Ge-capped films increases compared to the bare bismuth films at high temperature.

Now consider the 0.1% nominal acceptor concentration bismuth data in figure 8.5. The germanium deposition increases the longitudinal conductivity at high temperature and decreases it at low temperature while increasing Hall conductivity fairly evenly across all temperatures. Considering the low-temperature longitudinal conductivity, this decrease with the Ge capping layer is perhaps again due to compensating electrons. It is possible that at 0.1% nominal acceptor concentration, there remains a small electron pocket that can be compensated by Ge atoms. I am not sure what to make of the increase in longitudinal conductivity with the Ge layer. The Hall conductivity data are highly reminiscent of the doped film data from Chapter 7 in that the Ge-capped samples have an almost constant offset from the uncapped samples at all temperatures. I tentatively conclude this has a similar cause of Ge hole-doping.

With somewhat ambiguous results and many plausible scenarios, it is hard to be definitive in making conclusions about this data. It seems very likely, however, that
Figure 8.4: A comparison of magnetotransport properties for undoped bismuth films without and with germanium capping layers.
Figure 8.5: A comparison of magnetotransport properties for 0.1% Pb-doped bismuth films without and with germanium capping layers.
the bismuth is annealed, given the material has a melting point of 271.4 °C. During the germanium deposition process, the evaporation boat was radiating so much heat that adjacent parts of the bell jar could not be touched for very long without pain. This is despite there being an aluminum enclosure surrounding most of the solid angle that would receive germanium deposition.

8.2.2 Discussion for Germanium-Capped Bismuth Sensors

Now let us do the same analysis with germanium-capped bismuth sensors, again with 100 nm of Ge.

For the undoped probe data, seen in figure 8.6, both the longitudinal and Hall conductivities increase across all temperatures, with larger increases at high temperature. This is consistent with hole doping from the germanium. It is also possible that mobilities have increased due to annealing of the bismuth, though it is difficult to discern between these scenarios.

For the doped probe data, seen in figure 8.7, the addition of germanium causes a decrease in longitudinal conductivity at low temperature and an increase in Hall conductivity at high temperature. I am not sure what causes the temperature dependence. I leave the data here as a reference for future empirical inquiry.

At least we know that protecting 0.1% nominal acceptor concentration sensors with germanium does not drastically affect the magnetotransport properties. These Hall sensors will need all the protection they can get.

8.3 Hall Probe Microscope Image

Ultimately, I did secure a Hall probe microscope image, though the time for which the probe worked was rather fleeting. The image is shown in figure 8.8. I used a SRS810 Lock-In Amplifier in series with a 100 kΩ resistor to drive the excitation current through the probe and to measure the subsequent voltage from the Hall probe. My records of the exact lock-in amplifier settings are regrettably not good, but from memory and the data files I have, here are the settings I believe I used for the scans: scan height of 200 nm, horizontal step size of 200 nm, 100 scans averaged, signal input A - B, time constant of 10 ms, reference output of 3 V rms at 990 Hz, and 10 ms wait time before measurement.

Looking at the Hall probe image, it does indeed resemble the worm domains shown in figure 8.9. The Hall probe image shows the width of a worm domain to be
Data: Undoped Probe Longitudinal Conductivity

Data: Undoped Probe Hall Conductivity

Figure 8.6: A comparison of magnetotransport properties for undoped bismuth probe without and with a Ge capping layer.
Figure 8.7: A comparison of magnetotransport properties for 0.1% nominal acceptor concentration bismuth probe without and with a Ge capping layer.
Figure 8.8: A Hall probe scan of worm domains in an iron garnet film. The units of the colour map are in microvolts, and for this probe, a change of about 73 microvolts is a change in magnetic field of a tesla. The field detected is on the order of a millitesla.

Figure 8.9: A magneto-optical Kerr effect microscopy image of an iron garnet film’s worm domains. According to T. Chouinard [78], the domains are $10.6 \pm 1.8 \, \mu m$ wide. Image courtesy of Chouinard.
about 8 microns while the MOKE microscopy image indicates domains are about 10.6 microns wide. It shows my Hall probe microscope working with a probe I made for the first time. The image is low-resolution compared to a scan published by Sandhu et al. [39], but the quality should improve with future iterations of Broun Lab Hall probes and Hall probe microscope software.
Chapter 9

The End of a Beginning

“Now, the name of this talk is ‘There is Plenty of Room at the Bottom’ — not just ‘There is Room at the Bottom.’ What I have demonstrated is that there is room — that you can decrease the size of things in a practical way. I now want to show that there is plenty of room. I will not now discuss how we are going to do it, but only what is possible in principle — in other words, what is possible according to the laws of physics.

— Richard Philips Feynman, There’s Plenty of Room at the Bottom [1].

More than half a century has passed since Feynman gave that seminal talk. Perhaps in some ways, we have surpassed what he envisioned at the time, with the proliferation of scanning probe microscopy techniques and the stunning decades-long exponential growth of transistor density. Yet, we have gotten to this point by use of top-down processes of large machines making small devices, not by Feynman’s suggestion of small machines making smaller fabrication tools making yet smaller instruments, and so on.

There is still much to be done in the nanotechnology world. With some more advances, we can imagine a democratization of nanoscale fabrication tools like we have seen with 3D printers at larger scales. Perhaps artists will print and program their own interactive wallpapers and windows. Maybe average people will print clothing that could change its look on command. Who knows what a generation native to such tools will build? There remains plenty of room at the bottom.

Following in the trail blazed by commercial nanofabrication technologies, there is also plenty of room for scanning Hall probe microscopy to improve. I would like to end off my thesis looking to the future, at possible improvements to the field. We have come across many of them already in this thesis. I gather them here for a somewhat
coherent look at exciting prospects for scanning Hall probe microscopy, specifically for microscope development, Hall device design and fabrication, and innovations in materials science.

9.1 Hall Probe Microscope

The Broun Lab’s Hall probe microscope has several useful features: a long-range coarse motion stage, a fine motion stage capable of sub-nanometre motion, piezo-controlled double gimbals for probe rotation, a lock-in amplifier for Hall voltage measurements, and LabVIEW software to make it all work together.

It can still be improved, of course. The gimbals’ piezos need software written for them to work, and that needs to be integrated into the existing Hall probe microscope software. Some precise method of aligning the Hall probe tip to the surface to be scanned would be helpful. The current method is to use a dissecting microscope to peer through gaps in the casing around the sample stage. A small, focussed light, like an LED, may be helpful to illuminate the sample stage too. If the software were highly advanced, perhaps it could also model the curvature of the sample surface and adjust the tilt of the Hall probe to stay a constant distance from the surface. That is probably a very demanding task, but let us dream a bit!

A capacitive position readout for the coarse motion stage would also be very helpful. Knowing the position change would allow very large area scans by stitching together smaller area scans. T. Chouinard did some proof-of-principle experiments for such a system, but full integration into the Broun Lab’s Hall probe microscope would require another capacitance bridge or some capacitance measurement multiplexing.

Finally, development of time-resolved Hall probe capability would open up many other physical systems to study with a sensor that already has high flux sensitivity and high spatial resolution. The Broun Lab has a picosecond pulse generator, with possible access to another. The software and apparatus would need upgrading, of course.

9.2 Hall Device Design and Fabrication

As discussed in detail in chapter 6, I developed a novel probe design that, in principle, is more robust than conventional probes. This robustness allows for operation closer
to the sample surface, which improves spatial resolution and makes it easier to detect weak, localized signals.

Fabrication progressed as far as making the probe shape as a metal-insulator-metal trilayer with gold contacts. I attempted unsuccessfully to use a gallium focussed-ion beam and evaporated bismuth to make the active region of the probe. Perhaps using a different protective layer on top of the probes during ion beam milling would be helpful. I was using PMMA, but the resist around the milled hole appeared to be overexposed and stayed in place after the rest of the PMMA was dissolved. A sputtered insulating layer with an etch that is compatible with the trilayer may work well.

A uniform conducting layer on top of the resist may also be helpful for precision milling. I used a carbon layer on top to reduce charging of the resist during milling, but gold is an alternative that may perform better.

Another method to try would be helium ion milling. Helium ions would damage PMMA or some other layer less. Spatial resolution of the milling will be better. Conveniently, 4D LABS acquired a helium ion microscope.

Design can go in some intriguing directions. For instance, Weaver’s group at the University of Glasgow has put a Hall probe on an AFM tip [79]. This also puts the Hall probe very close to the sample surface. They developed an innovative “float” method of resist-coating to apply resist evenly to non-flat samples, like their AFM tips [80].

### 9.3 Hall Probe Materials

As we saw in chapter 7, doped bismuth is a significant improvement over undoped bismuth for Hall sensors, and it may be a further improvement over semiconductors. My work with lead-doped bismuth can certainly be improved upon with precise measurements of dopant concentration in the evaporated films, with more dopings tested, and more magnetotransport measurements taken to better constrain the parameters in the model. High-field measurements may be particularly helpful since they simplify the three-band model significantly, though the parameters themselves may vary significantly with field.

*n*-doping bismuth may be a convenient way to sidestep the three-band complications. With a bit of work, co-evaporation of bismuth and an *n*-dopant, like tellurium, can be set up in the evaporation chamber.
Further miniaturization and testing of shape effects on bismuth Hall probes would be helpful. This would drive further improvements to spatial resolution.

Studying the effects of low-temperature deposition would be helpful for tuning the mean free path of charge carriers in bismuth, a key parameter in the model we built in chapter 3.

Using different thicknesses of bismuth may also have interesting effects, depending on how the crystallites form and how large they are.

An adhesion layer of, for example, very thin titanium with high sheet resistance might help keep conventional probes intact under mild abrasion.

Finally, a more durable capping layer for conventional Hall probes may be useful in the interim, as the novel probe undergoes development.

\section*{9.4 An End}

This is it: the end of one trip through the nanoworld of Hall-effect sensors. It is but one region in a simultaneously vast and tiny territory. The road down here was carved by Silicon Valley entrepreneurs who saw the potential of computers. My project certainly follows in their path somewhat. We are like explorers in a new world\footnote{Thankfully, no indigenous peoples were harmed in this exploration.}. We hacked through the weeds, found a couple of nice lakes, and saw a lovely view of the valley below. At some point, it got dark, and now we have to go home. Thanks for following along with my scientific adventure.

There remains much unexplored territory, though. Perhaps you will look around yourself some more? One thing is for sure. There are going to be a lot more people mapping the place and finding all sorts of cool stuff. There should be no worry about it getting crowded, though. There’s plenty of room at the bottom.
Bibliography


[37] C. W. Hicks, L. Luan, and K. A. Moler, High resolution scanning GaAs/AlxGa1−xAs Hall probes (2006), APS March Meeting.


[68] F. Swanson, Powder diffraction file 00-005-0565 (1953).


Appendix A

Supplemental Calculations

I have placed additional calculations here that are referenced in the main text of the thesis. They are generally technical points of interest to experts and are presented here for completeness.

A.1 Telegraph Noise in Hall Probes

We can model telegraph noise as a Poisson process where charges hop off a defect with average time $\tau$. The probability that a charge goes from the defect (state 2) to the Fermi sea (state 1) in time $t$ is

$$P_{2\rightarrow1}(t) = 1 - e^{-t/\tau}. \quad (A.1)$$

with probability density $\rho_{2\rightarrow1}(t) = \tau^{-1}e^{-t/\tau}$. The charge hopping causes the voltage between two leads to jump between, say, 0 and $x_0$ volts. The fluctuations $y(T)$ at time $T$ have the form

$$y(T) = \begin{cases} x_0 & \text{for } 0 < T < t; \\ 0 & \text{otherwise}. \end{cases} \quad (A.2)$$

By Carson’s theorem, the power spectral density $G(f)$ of the noise is Lorentzian:

$$G(f) = \frac{2}{\tau} \int_0^\infty |\mathcal{F}\{y(T)\}|^2 \rho_{2\rightarrow1}(t) \, dt = \frac{4x_0^2\tau}{1 + 4\pi^2f^2\tau^2}. \quad (A.3)$$

This is for one impurity. For multiple impurities, we must integrate over the distribution of hopping rates. Assuming that the average hopping times $\tau$ are exponentially distributed by the impurity-conductor distance $z$ ($\tau \propto \exp[\alpha z]$), that there is a continuous range of relaxation times $[\tau_1, \tau_2]$, and that $z$ is uniformly distributed, we get
a probability distribution $\rho_\tau(\tau)$ for the relaxation times [81]:

$$
\rho_\tau(\tau) = \frac{1}{\tau \ln(\tau_2/\tau_1)}.
$$

(A.4)

Figure A.1: Impurities (blue dots) are close enough to the 2DEG for charge carriers to hop onto and off of them.

With the further assumption of identical fluctuations of size $x_0$, we integrate over $\rho_\tau$ to get the following average noise power spectral density per impurity:

$$
\bar{G}(f) = \int_{\tau_1}^{\tau_2} \rho_\tau(\tau) G(f) \, d\tau = \frac{2x_0^2 [\arctan(2\pi f \tau_2) - \arctan(2\pi f \tau_1)]}{\ln(\tau_2/\tau_1) \pi f}.
$$

(A.5)

When the measurement frequency $f$ is well within the range of hopping rates, that is, when $1/\tau_2 \ll 2\pi f \ll 1/\tau_1$, we end up with $1/f$ noise:

$$
\bar{G}(f) \approx \frac{x_0^2}{\ln(\tau_2/\tau_1)f}.
$$

(A.6)

A.2 Hall Probes: Heat Conduction Along Leads

The goal is to estimate the temperature increase in the Hall probe active region that arises from electrical heating, taking into account two conduction paths: flow from the active region directly into the substrate and flow from the active region to the leads.

Let us conceptually split the Hall probe structure into a set of squares as shown for simplicity of calculation because the important quantities for my calculation are the sheet electrical resistance and the sheet thermal resistance, which are resistances per square of material. This should underestimate the level of cooling since it ignores additional surfaces that can conduct heat away from the active region. The scheme for splitting up a probe is shown in figure A.2.

Let us make a further simplification that the heat flow problems for each case, heat flow into the substrate and heat flow through the leads, are effectively one-dimensional with the average temperature difference between two surfaces driving the heat flow.
Figure A.2: A scheme to split a planar Hall probe of thickness $d$ into red and blue squares for heat flow calculations. The dimensions of each small square are $s \times s \times d$. The large squares are $3s \times 3s \times d$.

For the in-plane heat flow through one of the small red squares, we have

$$Q^\parallel = \frac{\Delta T^\parallel}{R^\parallel_T} = \frac{\Delta T^\parallel A}{\rho_T L} = \frac{\Delta T^\parallel s d}{\rho_T s} = \frac{\Delta T^\parallel}{R^\square_T},$$

(A.7)

where $\parallel$ means in-plane, the $\Delta T$ is the average temperature difference, $Q$ is the heat flow, $R_T$ is the thermal resistance, $\rho_T$ is the thermal resistivity, and $R^\square_T$ is the thermal sheet resistance. This shows that the heat flow parallel to the plane is independent of the size of the square, so this applies to the blue squares too.

For the into-the-plane heat flow through one of the small red squares, we have

$$Q^\perp = \frac{\Delta T^\perp}{R^\perp_T} = \frac{\Delta T^\perp A}{\rho_T L} = \frac{\Delta T^\perp s^2}{\rho_T d} = \frac{\Delta T^\perp s^2}{R^\square_T d^2},$$

(A.8)

where $\perp$ denotes into-the-plane quantities. Here, heat flow into the plane is proportional to the area of the square, so we would need to scale this up with the blue squares.

The calculation of electrical currents flowing through the Hall probe wires is analogous to the in-plane heat flow problem, except we replace $Q^\parallel$ with electric current $I$, $R^\square_T$ with electrical sheet resistance $R^\square$, and $\Delta T$ with electric potential difference $\Delta V$:

$$I = \frac{\Delta V}{R^\square},$$

(A.9)
This implies that the heat power generated by the electrical current is independent of the size of the square, so this also applies to blue squares.

With these results in hand, we can now build a rough model of heat flow in the system. Consider the heat flow problem from the red zone to the blue zone in figure A.2. We seek the average temperature increase in the red zone due to running a current \( I \) from the left blue square to right blue square. We ignore the electric heating in the blue squares since they have a large area compared to the red squares and have at least nine times the capability to conduct away that heat.

We can model the heat flow as a thermal resistance circuit with the Hall probe substrate as the thermal ground, as shown in figure A.3. The red zone has 5 parallel into-the-plane connections to thermal ground each with thermal resistance \( R_1 = R_{\square}d^2/s^2 \). The red zone has 4 parallel in-the-plane connections to blue squares each with thermal resistance \( R_2 = R_{\square}T \). The 4 blue squares have each have connections to thermal ground with thermal resistance \( R_3 = R_{\square}d^2/9s^2 \). The net effect is that \( R_2 \) and \( R_3 \) are connected in series and there are 4 parallel arrangements of \( R_2 \) and \( R_3 \), resulting in the diagram seen in figure A.3.

\[
\begin{align*}
\Delta T &= + R_1 \quad \frac{R_2}{5} \quad \frac{R_3}{4} \\
\end{align*}
\]

Figure A.3: Thermal resistance circuit model for Hall probe heat conduction.

The heat flow \( Q \) that can be conducted away is a function \( \Delta T \), the temperature increase of the red zone above the substrate temperature, and \( R_{\text{eff}} \), the effective thermal resistance of the thermal circuit. The relation is

\[
\Delta T = Q R_{\text{eff}} = Q \frac{R_1(R_2 + R_3)}{4R_1 + 5(R_2 + R_3)} = Q R_T \frac{(d^2/s^2)(1 + d^2/9s^2)}{4d^2/s^2 + 5(1 + d^2/9s^2)}. \tag{A.10}
\]

In steady state, the electrical heating is equal to the heat flow out of the red zone, which has 3 squares that carry electric current, each providing \( I^2 R_{\square} \) of thermal power, so

\[
Q = 3I^2 R_{\square}. \tag{A.11}
\]

Thus, we get

\[
\Delta T = 3I^2 R_{\square} R_T \frac{(d^2/s^2)(1 + d^2/9s^2)}{4d^2/s^2 + 5(1 + d^2/9s^2)}. \tag{A.12}
\]

Now we need to evaluate the fraction with \( d^2/s^2 \) terms. Let us look at two limits: \( s \gg d \) and \( d = s \). The first case corresponds to a very thin layer of material composing
the Hall probe, which has little thermal resistance to dumping heat to the substrate, which allows more current to be put into the system. Conduction of heat through the leads is negligible compared to conduction of heat directly to the substrate. Our assumption of red and blue zones breaks down, in fact, and only the active region matters with respect to heat flow, which mainly happens directly to the substrate. If \( d \) is very small, a further assumption of zero thermal resistance between the probe and substrate may also break down, equation (A.12) no longer holds.

The second case corresponds to a logical limit to fabrication: thickness equal to width of a wire so that there is no risk of the wire being so tall that it may fall over, in addition to being relatively easy to make. Conduction of heat through the leads and conduction directly through to the substrate are both important, and equation (A.12) should be important. Let us evaluate equation (A.12) for \( d = s \):

\[
\Delta T \approx 0.35 I^2 R^\Box R^\Box T. \tag{A.13}
\]

To put this in perspective, let us compare this to the temperature change that would occur if the electric heating could only be conducted directly to the substrate through the Hall probe active region, of area \( s^2 \) and thickness \( d \):

\[
\Delta T = QR_T = (I^2 R)(R_T) = (I^2 R^\Box)(R^\Box d^2 / s^2) = I^2 R^\Box R^\Box T. \tag{A.14}
\]

The leads offer about a factor of 3 improvement in conducting away heat from the active region.

Further improvements are possible if the difficult task of fabricating high-aspect-ratio nanostructures can be done.

### A.3 Hall Probes: Optimizing Minimum Detectable Field for Carrier Types

Recall from equation (7.5) that the minimum detectable field for a Hall probe is

\[
B_{\text{min}} = C_\alpha \frac{p_1 \mu_2 + p_2 \mu_2}{p_1 \mu_1^2 + p_2 \mu_2^2} \sqrt{\frac{1}{\kappa_{\text{ph}} + C_\beta (p_1 \mu_1 + p_2 \mu_2)}}. \tag{A.15}
\]

We seek to optimize \( B_{\text{min}} \) with respect to \( p_1 \) while keeping the total number of carriers constant. Mathematically, we enforce this condition by defining the constant total number of carriers as \( N_{\text{tot}} = p_1 + p_2 \), so \( dp_1 = -dp_2 \). The substitution \( p_2 = N_{\text{tot}} - p_1 \) gives us

\[
B_{\text{min}} = C_\alpha \frac{p_1 \mu_1 + (N_{\text{tot}} - p_1) \mu_2}{p_1 \mu_1^2 + (N_{\text{tot}} - p_1) \mu_2^2} \sqrt{\frac{1}{\kappa_{\text{ph}} + C_\beta (p_1 \mu_1 + (N_{\text{tot}} - p_1) \mu_2)}}. \tag{A.16}
\]
Now optimize $B_{\text{min}}$ with respect to $p_1$ for the case where $p_1$ and $p_2$ are both hole bands:

$$\frac{dB_{\text{min}}}{dp} = -C_\alpha [\mu_1 - \mu_2] \frac{2\kappa_{\text{ph}} N_{\text{tot}} \mu_1 \mu_2 + C_\beta (p_1^2 \mu_1^3 + p_1 \{2N_{\text{tot}} + p_2\} \mu_1^2 \mu_2^2)}{[2(p_1^2 + p_2^2)^2(\kappa_{\text{ph}} + C_2(p_1 \mu_1 + p_2 \mu_2))]^{3/2}}. \quad (A.17)$$

Note that all of the terms in the numerator in braces {} must be positive or zero since $N_{\text{tot}} \geq p$, so all terms in the numerator are positive. The denominator is also positive. Note also the overall factor of $-\lbrack \mu_1 - \mu_2 \rbrack$ in the equation.

We seek to decrease the minimum detectable field, so if $\mu_1 > \mu_2$, then we should increase $p_1$. In other words, minimizing the minimum detectable field requires us to keep increasing the relative share of the larger mobility carrier, $p_1$, until the other carrier is gone. We should go single-band. This result applies not just to the case with two hole bands, but is also symmetric with the case of two electron bands.

Now consider the case of one hole band $p$ and one electron band $n$. Optimization of $B_{\text{min}}$ with respect to $p$ gives us

$$\frac{dB_{\text{min}}}{dp} = -C_\alpha \frac{2\kappa_{\text{ph}} N_{\text{tot}} \mu_p \mu_n (\mu_p + \mu_n) + C_\beta (p^2 \mu_p^4 + p \{N_{\text{tot}} + 2n\} \mu_p^2 \mu_n^2)}{[2(p^2 \mu_p^2 - n^2 \mu_n^2)^2(\kappa_{\text{ph}} + C_3(p \mu_p + n \mu_n))]^{3/2}}. \quad (A.18)$$

Note that, as before, all terms in the braces are positive or zero, the numerator is positive, and there is an overall negative sign in front of the fraction.

We must be careful in our interpretation of this result, though. A naïve interpretation would be that we should always increase the share of hole carriers since this decreases $B_{\text{min}}$. The astute reader will notice that $B_{\text{min}}$, as defined, is negative when electrons have the largest influence over the Hall effect. This is because for a given magnetic field, the Hall voltage reverses when the dominant charge carrier goes from holes to electrons or vice versa. To stay consistent, we should only apply this result when holes have the greatest influence on Hall voltage, that is, when $p \mu_p^2 > n \mu_n^2$. Again, we see that decreasing the minimum detectable field means we must increase the relative share of $p$ until there are no more $n$ carriers remaining. The situation is symmetric if electrons dominate; then increasing field sensitivity implies that we should eliminate holes.

To summarize, for maximum Hall probe field sensitivity, one band is better than two bands.
Appendix B

Continuous Flow Cryostat

A magneto-transport measurement flow cryostat was constructed to measure resistivity and Hall coefficient for four samples simultaneously, in magnetic fields up to 9 T and at temperatures ranging from 4–300 K.

B.1 Overview

Wendell Huttema, David Broun, and I constructed a continuous flow cryostat designed to take magnetotransport measurements using Jeff Sonier’s superconducting magnet. Its most notable features are as follows:

- 4-point probe of resistivity and Hall coefficient
- temperature control from ~4–300 K
- magnetic field control up to 9 T
- 4 samples measurable in a temperature sweep
- connection to helium recovery system
- LabVIEW measurement automation.

As the name suggests, this continuous flow cryostat boiled a stream of liquid helium using an electrical heater, controlled by PID loop, to smoothly sweep the temperature as it took measurements on bismuth films and Hall probes made for this thesis.
B.1.1 Physical Design

Helium flow in the cryostat goes from bottom to top, with a thermometer and heater set up at the bottom, used to finely control the temperature of the helium vapour. As the vapour goes up the inside of the cryostat, it encounters the sample can and samples, cooling them. The sample can has a hole at the bottom to allow the gas to diffuse into the can, allowing for better thermal contact. The hole has a pipe thread to allow for sealing, and the can has an indium seal at the top. As the helium gas continues to go up, it is slowed by baffles to allow for more heat exchange, creating a thermal gradient between the cold parts at the bottom and the room temperature parts at the top. Finally, helium exits through the helium outlet and a roughing pump propels the gas to the helium recovery system. Thermometry and temperature control of the helium is done by an Oxford Instruments ITC-503S unit, connected through a port on the flow cryostat.
To take measurements, the cryostat insert has twisted-pair wiring going from Fischer connectors\(^1\) at the top to a sample stage and thermistor in the bottom can. The wiring is cooled via heat-sinking to copper posts attached to the vacuum can. The thermistor is also thermally anchored to the vacuum can, allowing for better sample temperature measurements than the thermometer at the bottom of the cryostat. This wiring and calibration of the thermometer was a major part of my contribution to the cryostat construction.

Stretching away from the Fischer connectors marked “A” and “B” are cables I made connecting to a complicated wiring box I made to interface with the Lakeshore 3716L Low Resistance Scanner, which allows us to switch between different configurations of wires to perform 4-point probe measurements of resistivity and Hall coefficient. A Lakeshore 370 AC Resistance Bridge is connected to the scanner for those measurements.

Stretching away from the Fischer connector marked “T” is a cable I made to interface with a Cryo-Con 62 Cryogenic Temperature Controller, which we use only for temperature measurements.

Jeff Sonier’s superconducting magnet, vital for Hall effect measurements, has the following instrumentation attached:

- Magnet power supply: Oxford Instruments IPS 120-10
- Nitrogen level meter: Oxford Instruments ILM-211S
- Helium level meter: Cryomagnetics Inc. LM-500
- Vacuum space turbo pump\(^2\)

The sample stage, designed by Wendell Huttema, holds four samples simultaneously and is also thermally anchored to the vacuum can. The stage can be switched between two different orientations: samples perpendicular or parallel to the magnetic field. Samples are mounted with double-sided tape on circuit board “chips” that can easily be plugged or unplugged from the sample stage, as seen in figure B.2. Although each chip has nine contact pads, only four are used. The back-side connectors of the chips are used to provide thermal contact with the vacuum can.

### B.1.2 LabVIEW Software

The LabVIEW program for data collection, written by Huttema, begins with a hand-coded measurement ranging algorithm on the resistance bridge, then repeatedly iter-

---

\(^1\)Fischer connectors are great. Solidly-built and well strain-relieved, they are a definite connector of choice in experiments like this.

\(^2\)The vacuum space between the nitrogen jacket and the liquid helium-cooled magnet leaks slightly and needs to be constantly pumped to maintain good vacuum.
ates over user-selected lead configurations, calculating the ratio of voltage and current measurements and writing the data to a file for analysis by a dedicated Mathematica notebook, which is to be discussed in later sections of the thesis.

The basic algorithm for data collection is this:

1. Start temperature sweep
2. Initialize excitation currents for each measurement channel
3. Auto-range all the channels to be measured to the appropriate resistance ranges
4. Measure resistance for each channel and temperature for thermistor located close to samples
5. Write data to a file
6. Repeat 3–5 until temperature sweep is completed.

Temperature sweeps happen very slowly, typically taking 10 hours to sweep from room temperature to 4 K.

B.1.3 Operational Notes

Jeff Sonier’s 9 T magnet generates a field pointing up when a positive current is pushed through it.

The helium transfer siphon should have its needle valve opened completely and have helium vented through it for a minute before being placed in the liquid helium dewar, with helium still flowing. This prevents ice or solid $\text{CO}_2$ from forming, obstructing helium flow in the tube.

Occasionally, the Lakeshore resistance bridge has had problems changing scales and gave erroneous results around 2.00 and 6.32 of each measurement decade, e.g. 2.00 $\Omega$ and 6.32 $\Omega$ or 20.0 $\Omega$ and 63.2 $\Omega$. This problem should be rectified with the latest LabVIEW VI for running the flow cryostat measurements.
Figure B.2: Two sample stage configurations on the left, two sample chips on the right. Looking at the top sample chip, only the top and bottom contacts on the left and right sides of the chip are used in measurements. Drawing by Pawel Kowalski of the SFU Electronics Shop.
Appendix C

Metal Evaporation

In the preceding discussion, we glossed over how the device material is deposited onto the substrate. This can be done in many ways, including sputtering, chemical vapour deposition, and evaporation, but evaporation was chosen because it is easy to use, cheap, reliable, and allows for careful, reproducible tuning of all relevant Hall probe properties. Since the evaporator is used only by the Broun Lab, impurities picked up from shared equipment are avoided.

C.1 Evaporator System

The evaporator system is simple. The vacuum chamber is formed by a glass bell jar with a metal bottom, which has feedthroughs for electricity and water, as needed by instruments inside. A mechanical roughing pump and a water-cooled diffusion pump are sufficient to reach pressures in the $10^{-6}$ Torr range, as measured by the Veeco Instruments RGLL-7 Ionization Gauge Controller. An evaporation boat with bismuth inside is heated electrically to produce bismuth vapour. A schematic can be seen in figure C.1.

The vapour can condense on substrates at either the multi-sample carousel (capacity: 8 standard microscope slides) or the cryogen-cooled low temperature stage. Cooling helps control crystallite size, which is important in preventing Hall effect quenching. A cooled stage shrinks crystallites by reducing the amount of time allowed for hot, freshly evaporated bismuth atoms to rearrange themselves on the substrate. This allows us to change the mean free path of charge in the film. Impurities can be added to the evaporation boat to dope the evaporated films, changing the carrier concentration, and therefore, the Fermi wavelength of the films. As noted in section 3.6.3, Fermi wavelength and mean free path are the fundamental material parameters in Hall probes, so in principle, we have complete control of the properties of the patterned sensors.
Figure C.1: A schematic of the evaporator used to make bismuth films. Bismuth (in blue) is evaporated by electrically heating the semimetal. A bell jar keeps the apparatus under vacuum to allow hot bismuth particles (blue dots) to travel to the thickness monitor, carousel, and cold stage without air resistance or air cooling while preventing film oxidation and fire.

There are two places to mount independently-powered evaporation boats and manually-operated shutters for both. Co-deposition is possible with this configuration, though this capability was not used for this thesis. Perhaps future projects on doped bismuth Hall probes would find this useful.

The thickness of the evaporated film is measured using a quartz crystal microbalance, which is controlled by a Sigma Instruments SQM-160 Rate/Thickness Monitor. Because the crystal is not at the same spot as the substrate, the microbalance must be calibrated with a tooling factor that compensates for the difference between the substrate and crystal deposition amounts.

### C.2 Tooling Factor

The tooling factor was calculated using profilometer measurements on thin films evaporated onto standard glass microscope slides on the evaporation chamber carousel. A permanent marker was used to mask off a section of each slide; the ink was subsequently dissolved using ethanol, lifting off bismuth to create a step edge. A Bruker Dektak XT profilometer was then used to measure the height of that step edge.
Mathematically, the tooling factor is simply

\[ f_{\text{tool}} = \frac{\text{film thickness on glass slide}}{\text{film thickness on crystal microbalance}}. \] (C.1)

The profilometer data can be difficult to interpret at times. Accurate determination of step height requires good levelling of a scan, where one tries to define one region of the scan as flat. Occasionally, the substrate section and the film section slope away from each other, and it is not clear how one should level the scan. There may also be local substrate or film defects.

![Profilometer Scan for March 8, 2012 Bismuth Sample #1](image)

**Figure C.2:** A sample profilometer scan of a step edge on a bismuth sample. Ideally, there would be no gradients for the sections of film on either side of the bare substrate patch. It is not clear whether the gradients are real or not, perhaps due to substrate movement from the lack of attachment points to secure the substrate. Certainly, the gradient reaching past the apparent substrate level on the right cannot be real.

In my analysis, I measured the sizes of the step edges of two films 10 times for each film and calculated the average step height. I eliminated two large outliers in recognition of the uncertainties I outlined above, then calculated the tooling factors for each film. I further averaged the two film tooling factors to get a final tooling factor of 0.90.
I also did similar calculation for a germanium film. The results are summarized in table C.1.

Table C.1: Tooling factor calculation results. The uncertainties listed for profilometry are the standard deviations of the measurements.

<table>
<thead>
<tr>
<th>Sample Details</th>
<th>Profilometry Average Thickness (nm)</th>
<th>Microbalance Film Thickness (nm)</th>
<th>Result Tooling Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bi film 1</td>
<td>68.4 ± 2.2</td>
<td>75.7</td>
<td>0.904</td>
</tr>
<tr>
<td>Bi film 2</td>
<td>68.0 ± 3.5</td>
<td>75.1</td>
<td>0.905</td>
</tr>
<tr>
<td>Ge film</td>
<td>129 ± 16</td>
<td>140</td>
<td>0.92</td>
</tr>
</tbody>
</table>
Appendix D

Hall Probe Microscope

The Hall probe microscope has been discussed in detail already in chapter 4. The only thing I will mention here is to beware of temperature effects when running the microscope at ambient temperatures. Fluctuations in the temperature of the room are sufficient to significantly change the touchdown voltage for the $z$-piezos. This is hazardous to the health of Hall probes, particularly for long scan series. One way to mitigate this problem is to insulate the microscope by attaching the vacuum can and placing it in a cryogen dewar with packing peanuts surrounding everything.
Appendix E

Spatial Resolution Simulations

This section is documentation of the commented source code for the Maple worksheet used to calculate the spatial resolution of a Hall probe with a square active region. For clarity, the variable $AA$ is reproduced here:

\[
AA = \frac{\mu_0 m}{4\pi} \left[ \frac{x_1 y_1 (x_1^2 + 2z^2 + y_1^2)}{(y_1^2 + z^2)(x_1^2 + z^2)\sqrt{x_1^2 + y_1^2 + z^2}} - \frac{x_1 y_0 (x_1^2 + 2z^2 + y_0^2)}{(y_0^2 + z^2)(x_1^2 + z^2)\sqrt{x_1^2 + y_0^2 + z^2}} \right]
- \frac{x_0 y_1 (x_0^2 + 2z^2 + y_1^2)}{(y_1^2 + z^2)(x_0^2 + z^2)\sqrt{x_0^2 + y_1^2 + z^2}} + \frac{x_0 y_0 (x_0^2 + 2z^2 + y_0^2)}{(y_0^2 + z^2)(x_0^2 + z^2)\sqrt{x_0^2 + y_0^2 + z^2}}. \tag{E.1}
\]

This is the result of integrating equation (4.1) over the active region. For your reference, equation (4.1) is reproduced here:

\[
\iint \mathbf{B} \cdot d\mathbf{a} = \iint \frac{\mu_0 m}{4\pi r^3} \left(3z^2 - r^2\right) \, dx \, dy.
\]

The program calculates the flux detected by a Hall probe as it scans in the $x$-direction over a flat surface with two magnetic dipoles pointing in the $z$-direction. When size of the two highest flux peaks, presumably directly over the two dipoles, is smaller than some user-defined criterion times the flux exactly between the peaks. While the algorithm may not appear to be doing this directly, consideration of the even symmetry of the problem will make it clear that this is indeed what happens.

Users of the program should take care to use enough test points for $x$, $z$, and probe size $s$ to ensure that the simulation is accurate. The algorithm also relies on even symmetry of the spatial flux profile. Chopping out a peak or trough on one side and not the other will result in errors. Note that block text comments are in variable width font while in-line comments are in fixed-width font with a “#” in front.
Figure E.1: Flow chart for Maple resolution sim program.
restart;
with(plots):
with(LinearAlgebra):

AA is total flux through a probe of width \( x_1 - x_0 = y_1 - y_0 \) for a dipole at \( (0, 0, 0) \).
BB is total flux through a probe of width \( x_1 - x_0 = y_1 - y_0 \) for a dipole at \( (a, 0, 0) \).
CC adds AA and BB together while constraining the probe to be of size \( s \) and assigning a length of 1 to the dipole-dipole separation.
DD gets rid of some constants that do not affect the calculation.

\[
AA := \mu_0 \frac{m}{4 \pi} \left( \frac{x_1 \cdot y_1 \cdot (x_1^2 + 2 \cdot z^2 + y_1^2)}{(y_1^2 + z^2)(x_1^2 + z^2)(x_1^2 + y_1^2 + z^2)^{1/2}} - x_1 \cdot y_0 \cdot (x_1^2 + 2 \cdot z^2 + y_0^2) / ((y_0^2 + z^2) \cdot (x_1^2 + z^2) \cdot (x_1^2 + y_0^2 + z^2)^{1/2}) - x_0 \cdot y_1 \cdot (x_0^2 + 2 \cdot z^2 + y_1^2) / ((y_1^2 + z^2) \cdot (x_0^2 + z^2) \cdot (x_0^2 + y_1^2 + z^2)^{1/2}) + x_0 \cdot y_0 \cdot (x_0^2 + 2 \cdot z^2 + y_0^2) / ((y_0^2 + z^2) \cdot (x_0^2 + z^2) \cdot (x_0^2 + y_0^2 + z^2)^{1/2}) \right);
BB := \text{subs}(x_1 = x_1 - c, x_0 = x_0 - c, AA):
CC := \text{subs}(y_1 = s/2, y_0 = -s/2, x_1 = x_0 + s, c = 1, AA + BB):
DD := CC / \mu_0 / m / \pi:

This is the giant part of the program that calculates the values of \( z \) and \( s \) at which the two dipoles are barely resolvable as separate entities. The program simulates the flux seen by a Hall probe at height \( z \) over a sample and centred over the two dipoles in the y-direction, scanning in the x-direction and using an adjustable factor to define whether the dipoles are spatially resolved. With the dipole separation distance set to 1, all other distances are set relative the size of this dipole separation distance.

# adjustable parameters:
numzs := 46: # number of z (height) points to take
numss := 91: # number of s (sensor size) points to test
maxss := 1: # maximum sensor size
minss := 0.1: # minimum sensor size
maxzz := 1.0: # maximum sensor height
minzz := 0.1: # minimum sensor height
criterion := sqrt(2): # criterion for deciding resolution
xmax := 5:   # max scan coordinate
xmin := -5:   # min scan coordinate
numsteps := 201:    # number of data points

# store z and s points for resolution of 1
zs := Array(1..2,1..numzs):

# loop over the possible z values
for a from 1 by 1 to numzs do
  # initialize a parameter to break out of the s loop
  breakout := false:
  # height above sample
  zz := minzz + (a - 1) * (maxzz - minzz) / (numzs - 1):
  # loop over the possible s values
  for b from 1 by 1 to numss do
    if breakout = false then
      # probe size
      ss := minss + (b - 1) * (maxss - minss) / (numss - 1):
      # conversion to coordinates used in equation DD; xmax and xmin are relative to the centre of the Hall probe’s active region
      x0max := xmax - ss/2:
      x0min := xmin - ss/2:

      xx0 := Array(1..numsteps):
data := Array(1..2,1..numsteps):
flux := Array(1..numsteps):

      # calculate the x-positions of the back edge of the Hall probe and find the flux measured at these positions
      for i from 1 by 1 to numsteps do
        data[1,i] := x0min + (i - 1) * (x0max - x0min) / (numsteps - 1) + ss / 2:
        data[2,i] := simplify(subs(s = ss, z = zz, x0 = data[1,i] - ss/2, DD)):
      end do:
gtbefore := false:
localmincount := 0:
localmaxcount := 0:
localmins := Array(1..2,1..10):
localmaxs := Array(1..2,1..10):

# find local mins and maxes
for i from 1 by 1 to (numsteps - 1) do
    if evalf(data[2, i + 1]) >= evalf(data[2, i]) then
        gt := true:
    else
        gt := false:
    end if:

    if (gt = false and gtbefore = true) then
        localmaxcount := localmaxcount + 1:
        localmaxs[1, localmaxcount] := data[1, i]:
        localmaxs[2, localmaxcount] := evalf(data[2, i]):
    end if:

    if (gt = true and gtbefore = false) then
        localmincount := localmincount + 1:
        localmins[1, localmincount] := data[1, i]:
        localmins[2, localmincount] := evalf(data[2, i]):
    end if:

    gtbefore := gt:
end do:

maxcandidates := Array(1..2, 1..2):
mincandidate := Array(1..2):

if localmaxcount > 1 then
    # look for biggest max
    for i from 1 by 1 to localmaxcount do
        if localmaxs[2,i] >= maxcandidates[2,1] then
            maxcandidates[1,1] := localmaxs[1,i]:
            maxcandidates[2,1] := localmaxs[2,i]:
        end if:
    end do:

    # look for second-biggest max
    for i from 1 by 1 to localmaxcount do
        if localmaxs[1,i] <> maxcandidates[1,1] then
            # code continues...
if localmaxs[2, i] >= maxcandidates[2, 2] then
    maxcandidates[1, 2] := localmaxs[1, i]:
    maxcandidates[2, 2] := localmaxs[2, i]:
end if:
end if:
end do:

mincandidate[1] := maxcandidates[1, 1]:

# check if the middle must be a min or a max
if localmaxcount mod 2 = 0 then
    # use middle min between the two maxima
    mincandidate[1] := localmins[1, 1 + (localmincount - 1) / 2]:
    mincandidate[2] := localmins[2, 1 + (localmincount - 1) / 2]:
else
    # look for middle max
    mincandidate[1] := localmaxs[1, 1 + (localmaxcount - 1) / 2]:
end if:

# if the middle point is sufficiently large, then store (z, s) as 
# a data point
if (evalf((maxcandidates[2, 1] / criterion)) < mincandidate[2])
    breakout := true:
    zs[1, a] := ss:
    zs[2, a] := zz:
end if:
else
    # if there’s only one peak, then store z, s as a data point
    breakout := true:
    zs[1, a] := ss:
    zs[2, a] := zz:
end if:
end if:
end do:
end do:

Output the simulation results to a text file. First column is a set of s values. Second 
column is a set of z values.
writedata(fileout, convert(Transpose(convert(zs, Matrix)), matrix), float);
fclose(fileout):

Plot the results.

pointplot(zs, labels=["Hall probe size s", "Height above sample surface z"]);
Appendix F

Various Points of Technique

F.1 Profilometry

Profilometry is the measurement of the height of some surface along a scan path; it is performed by dragging a sharp needle across a sample.

A sharp step edge is desirable for accurate profilometry. This can be achieved by applying tape or using a felt pen to draw a line on the sample prior to material deposition. For a felt pen, be sure to use a thick layer of ink. After deposition, remove the tape or perform lift-off with the pen ink.

Tape tends to leave adhesive residues, making profilometry more difficult. Felt pen is desirable because it does not leave behind residues, though lift-off can fail due to marker lines that are too thin or due to sidewall deposits.

Thickness measurements should be taken close to the step edges. Farther from the step edges, the measurements tend to have greater uncertainty due to bowing of the substrate.

F.2 The Delicate Art of Cleaving Silicon

Ideally, conventional Hall probe substrates cleave cleanly and accurately, with sharp and straight edges. The goal is to have substrates that make perfectly cleaved 90° corners and place the probe as close to those corners as possible.
Silicon is not an ideal Hall probe substrate. It cleaves, but often in a jagged way and with lots of dust. I would much rather use gallium arsenide, which barely requires any effort to cleave perfectly. Gallium arsenide is expensive, though, and silicon is cheap because it is just refined sand. However, with careful practice and good technique, it is possible to reliably cleave silicon substrates for Hall probes.

F.2.1 (100) Si

(100)-oriented Si has cleavage planes that are perpendicular to each other. This is nice because if you know the dark art of teasing apart Si just right, you can cleave perfect corners for conventional Hall probes or cleave apart an array of novel Hall probes. Other orientations of Si might have cleavage planes at some odd angle to each other, which is not ideal.

F.2.2 Tools

- sapphire plate
- a ledge (0.5 mm of sticky notes works well. Not only do the sticky notes stick in place, but you can also easily change the height of the ledge by adding or removing sticky notes)
- a flat piece of backing paper so that you don’t scratch the table on which you work
- Kimwipes, to hold the substrate and avoid touching it with your dirty, oily hands

F.2.3 Scratch, Nick, and Snap

This technique provides perfect cleaves for samples longer than 4 cm in the cleavage direction, which is great for conventional Hall sensors to get the perfect corner and also for the novel sensor design to just separate the sensors. It does not provide perfect cleaves for samples shorter than 4 cm in the cleavage direction, but is still good for separating individual novel Hall sensors. The actual location of the cleave may shift about 0.5–1 mm from the score mark.
Use a piece of sapphire or a scribe to nick a straight edge of the substrate. While holding an edge with a Kimwipe, start the mark 2 or 3 mm from a straight edge of the wafer, and scratch the wafer towards the edge, applying firm pressure until your scratching device is very close to the wafer edge. Then finish the scratch by gently sliding the scratching device over the edge, which should remove a sliver of Si from the edge (on smaller pieces, the Si will just snap into two pieces, so try to be very gentle). When the cleavage plane of the wafer is placed over a small cliff, you should be able to cleave the wafer with gentle pressure on the middle of the sides of the two pieces you wish to cleave. (100) silicon will cleave perpendicular to the straight edge if you nick it enough times.

F.2.4 Nick

You might need to nick it for a couple of minutes in order to see a crack propagate. Once you see the crack, it is a good idea to be gentle, especially when using the razor blade to either make small pieces of silicon or to cleave an already small piece. Small pieces tend to fly up and flip over when too much force is applied. You don’t want to try to make silicon pieces smaller than about 1 cm with this method because the substrate tends to fragment.

F.2.5 Score and Snap

Another method is to use a very hard object, like a diamond scribe or sapphire plate, and firmly score the substrate along the line where you want to cut. Next, you can snap it by hand, by breaking it over the edge of a table or book, by the razor blade method above, or any other way you might like.

For large cleaves, score (doesn’t have to be very deep; superficially with light pressure from the sapphire plate is fine) a small distance (∼2 mm) along a cleavage plane, then press down in the vicinity of the score mark until cleaving begins.

For small chips, score the chip as described before, but place the chip so that the cleave you want to make is over a shallow cliff (1 mm works well). Push on the middle of edges of chip like trampoline, forming and propagating cracks through the chip, hopefully producing a clean and straight split. When “trampolining,” beware of pressing on chips too hard, which may cause the chip to flip over and/or make a
rough cleave. The lower limit on size for cleanly cleaving small chips is about 6 mm on either side of the cleave.

With sufficient practice and attention to technique, you should be able to get chips good enough for a Hall probe about 75% of the time.

F.2.6 Rough Cleaves

Sometimes, a Hall probe chip (about 1 cm × 1 cm) must be trimmed to fit on the probe characterization apparatus. You can cleave off small slices of silicon as small as 2 mm wide, but the method described here will generate lots of dust and debris, so it should only be used when the probe is complete.

First, use the sapphire piece to firmly score a line along a cleavage plane all the way across the silicon substrate. Next, hold the piece with the undesired piece on the left, held with a Kimwipe in your left hand. The score line should be just to the right of the ledge (about 0.5 mm). Hold tweezers with the tips oriented to the left and the top of the tweezers digging into the palm of your right hand so that the tweezer tips cannot slip to the right and destroy your sample. Use the tips to push on the chip you want to keep, pushing about 0.5 mm to the right of the score mark. The chip should cleave nicely under moderate pressure.

When cleaving off a small slice of silicon (about 2–3 mm wide), it may be helpful to hold down one end of the chip with something hard and unbending, like a piece of metal or the butt end of a set of tweezers. Then pushing down on the other side of the piece of silicon will result in a cleave.

With this method, you should be able to reliably perform straight cleaves practically all the time.

F.3 Piezo Cutting

For the piezo benders, the Hall probe microscope uses sheets of piezo ceramic (material code PSI-5H4E) made by Piezo Systems Inc. (part number T219-H4CL-503X).

The piezos can be cut with a wire saw, which uses a tungsten wire to rub alumina grit on the part to be cut for several hours.
The usual way to secure the piezos to the sample holder is using crystal bond. The Corning PC-35 hot plate in the lab needs to be heated to about halfway between “low” and “2” settings to melt crystal bond. It is not a good idea to work at too high a hot plate temperature since the piezo material depolarizes at 230 °C.

F.4 Sample Handling

1. You should almost always keep the sample in a “secure location” where it cannot be blown, scratched, dropped or otherwise damaged, such as a sealed sample box or sample holder.

2. You should never have to move a sample more than 30 cm from secure location to secure location.

3. You should never leave a sample unsecured.

4. You should never walk around with an unsecured sample.

5. Try not to drop wafers. They tend to drop buttered side down.

6. When picking up chips with tweezers, note the direction in which the cleaved edges slope. When the chip slopes down and outwards on both edges that you want to grip, you may find it hard to pick up the chip without damaging it.

F.5 Indium Contacts

The goal is to connect a copper or gold wire to bismuth contact pads using indium.

First, I press indium pads onto the bismuth contact pads. I do this by cutting off slightly less than 1 mm length of 1 mm diameter indium wire. Then I cut that tiny piece in half and mash one of the halves into a bismuth pad with a wooden stick cut to a cross-section about the size of the desired indium pad.

The next step is to prepare the copper or gold wire for the indium contact. I tin indium onto the wire, careful to eventually leave a blob of indium on the wire. This will help form a contact between the wire and indium contact. Then I use the same wooden stick to press the wire and indium blob onto the indium pad.
As an optional final step, I sometimes tried mashing another half of a cut indium wire on top of the copper or gold wire and the indium pad. It is not clear to me that this helps maintain contact well.

F.6 Static Electricity

Static electricity zaps devices. Some precautions:

- Ground yourself before touching a Hall probe.
- Work on a work table with a high-resistance connection to ground, possibly covered in aluminum foil.
- Use a grounding strap to remove any static charge you may gather.

(a) Hall probe from pattern written June 22, 2009. (b) Hall probe after destruction by static discharge.

Figure F.1: This 80–100 nm Hall probe (measured from the smallest constrictions leading to the active region) was destroyed by a static discharge. The gaps that appear post-discharge may well have been single crystals of bismuth that vaporized.
Appendix G

Hall Probe Production Recipes

G.1 Conventional Hall Probe Fabrication

G.1.1 Spin-Coating Resist onto Wafer

This is to be done in the 4D LABS cleanroom using their machines. Detailed training on the machines is available with the cleanroom staff.

1. Clean substrate.
   i) Acetone rinse.
   ii) IPA rinse.
   iii) Blow-dry with N\textsubscript{2}.

2. Bake dry.
   i) Hot plate, 180 °C, 5 minutes.
   ii) Cool, 3 minutes.

3. Spin-coat \sim 2000 Å of PMMA (950) using Cee #2\textsuperscript{1}.
   i) Mount wafer in Cee #2 with vacuum chuck. Centre the wafer.
   ii) Turn on vacuum.

\textsuperscript{1}This is one of the cleanroom’s spin coaters.
iii) Pipette PMMA solution onto wafer.

iv) Spin recipe: 500 RPM spread of resist for 5 s, 4000 RPM spin coat of 4% PMMA solution in anisole\(^2\) for 45 s.

v) Pre-bake on hot plate\(^3\): 180 °C, ∼90 s.

vi) Cool, 3 minutes.

vii) Turn off vacuum and clean spin coater.

4. Dice wafer\(^4\).

5. For each probe chip, make tiny scratch on a part that will not be patterned.

**Additional Notes**

I used (100) silicon wafers for the Hall probes. These wafers naturally cleave square edges. The natural cleave is important for getting a perfect corner for fabricating a Hall probe there. You should have a strategy to maximize the number of perfect corners you have, which is not trivial because you need to scratch some part to make the wafer want to cleave.

I liked to dice things into approximately 1 cm × 1 cm squares to accommodate a spot to make a scratch for SEM focussing, but after fabrication, I cleaved the chips to about 0.5 cm × 0.5 cm to mount nicely on carrier chips for the Hall probe microscope.

**G.1.2 Spin-Coating Resist onto Chip**

Procedure for spin-coating small samples (courtesy of Grace Li from the 4D LABS cleanroom staff). I never had to use this, but it seems like a useful thing to know.

1. Put tiny amount of PMMA (6% in anisole is suitable) in middle of wafer to be used as sample carrier.

2. Place sample on top of PMMA spot. There should be just enough PMMA to coat the bottom with none squirting out if you press down on sample. Any more will make life difficult.

---

\(^2\)If 3% PMMA in anisole is used, 1000 RPM should be used for the second phase.

\(^3\)Pre-heat before spin-coating.

\(^4\)This part can be done outside the cleanroom to save money, though if you are paranoid about dust, you can do it inside the cleanroom too.
3. Bake big wafer on 90 °C hot plate for a few seconds until sample is secure to a light push.

4. Spin coat sample as normal.

5. For sample removal, push on sample sideways using metal tweezers. Use little nudges. Pretend it is a dance where tweezers advance slowly on the sample.

6. Pre-bake sample as normal.

G.1.3 E-Beam Lithography

1. Load sample\(^5\), pump down chamber.

2. Set origin at clean corner of chip. Do angle alignment to (hopefully) clean bottom edge of chip.

3. Focus and correct for stigmatism using particles near scratch.

4. Pick particle on which to align write field.

5. Measure e-beam current.

6. Set up pattern write, exposure settings\(^6\).

7. Write pattern.

8. Repeat steps 2–7 for different patterns as needed\(^7\).

9. Vent chamber and remove samples.

---

\(^5\)The resist-coated chips should have a perfect 90° corner. The imperfect ones can be used for exposure or pattern tests. I usually put in three or four onto the sample holder just in case a couple of them are bad.

\(^6\)You should know what exposure to use from a previous exposure test to find the optimal e-beam dose.

\(^7\)I usually do the active region of the probe first, then the leads as a separate exposure, making records of exposure settings as the first pattern writes. When setting the origin for the exposure of the leads, I only do it at low magnification to avoid unnecessary exposures to the high beam current used for the leads.
G.1.4 Bismuth and Germanium Evaporation

1. Load desired evaporation material and substrates.

2. Close all valves. Open foreline valve.

3. Rough out rear of the diffusion pump to 20-30 microns of Hg. This should not take long.


5. Rough out chamber for 5-10 minutes (to 30 microns of Hg).

6. While chamber pumps down, turn on diffusion pump cooling water and the diffusion pump itself. Fill diffusion pump’s cold trap with liquid nitrogen. Make sure water is actually cooling pump.

7. Wait about 20 minutes for diffusion pump to warm up.

8. Close roughing valve. Open foreline valve and slowly open main valve, making sure pressure stays below 100 microns of Hg. When pressure stabilizes, slowly open main valve wide.

9. When chamber is sufficiently pumped down, turn on ion gauge. Wait for chamber to reach desired pressure.


12. Vent main chamber at your leisure and remove samples.

13. Continue pumping on back of the diffusion pump until bottom of the diffusion pump is merely warm or cooler to the touch.

---

8 There is only one working pressure gauge at low vacuum, so you may need to plug the gauge into different spots as you set up the evaporator.

9 Try to keep the bell jar centred over the sealing surface as it pumps down.

10 The waste water and the water pipes should both be cold.

11 If starting with the diffusion pump cold, you can ignore steps 2–8 and pump with the chamber through the backing valve while adding liquid nitrogen to the cold trap and warming up the diffusion pump.

12 The chamber should go down to $2 \times 10^{-6}$ Torr or lower.

13 You may want to wait a bit for the chamber’s contents to cool off.

14 This may take an hour or two.
14. (Optional) Open roughing valve and pump out main chamber to about 30 microns of Hg.

15. Close all valves and turn off cooling water.

16. Vent roughing pump by disconnecting the quickconnect joint.

Additional Notes

The evaporator shields and bell jar should be regularly cleaned of evaporated material to prevent contamination.

Flaking of films in the chamber can occur. I am told by cleanroom staff (Chris Balicki) that venting the chamber exacerbates the flaking problem, so after finishing evaporation, it is best to leave the carousel in a position so that flakes do not fall onto the evaporation boats. Also, keep the chamber under vacuum when not in use.

For evaporating a germanium capping layer, I made a sample holder with an adjustable aluminum mask for blocking the leads so that the germanium only covers the tip of the probe. The sample holder fits in the evaporator carousel.

The germanium evaporation gets everything in the chamber very hot, including the leads and feedthroughs. I do not know how well it would stand up to repeated evaporations of germanium.

G.1.5 Lift-Off

1. Go to fume hood.

2. Put about 20 mL of acetone\(^\text{15}\) in 50 mL beaker.

3. Place bismuth-covered probe chip in beaker.

4. Carefully swirl acetone around, making sure chip always remains immersed in acetone. Most of unattached bismuth should lift off.

5. Remove bismuth films freely floating in acetone.

6. (Optional) Transfer chip to new 50 mL beaker with 20 mL of fresh acetone.

\(^{15}\)High-purity acetone might be helpful to prevent residues from forming on the chip.
7. Put beaker in ultrasonic bath\textsuperscript{16} and operate for a second or two while holding onto the chip with tweezers\textsuperscript{17}.

8. Check using microscope that Hall probe is properly lifted off\textsuperscript{18}. If not, then repeat ultrasonic treatment and check again.

9. Transfer chip to new 50 mL beaker with 20 mL of fresh acetone, carefully avoiding any bismuth flakes that may be floating in the beaker. Swirl acetone in new beaker a bit to remove any flakes that may remain on chip.

10. Take chip out of beaker and onto Kimwipe on flat piece of aluminum\textsuperscript{19}. While holding onto chip with tweezers, blow-dry using N\textsubscript{2} gas, aiming from probe active region out towards leads.

G.1.6 Bismuth Doping

For this thesis, I used two methods of doping bismuth. The first is simply putting the desired amounts of lead and bismuth in an evaporation boat and then heating the boat for some time to allow the elements to mix. Without a clear indication of successful mixing, the precise mixing conditions I tried were not extremely consistent.

Table G.1: Doping parameters for mixing bismuth with dopants in evaporation boats.

<table>
<thead>
<tr>
<th>Doping</th>
<th>Bi (mg)</th>
<th>Pb (mg)</th>
<th>Evaporation rate (Å/s)</th>
<th>Time (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>“1%”</td>
<td>1967.29</td>
<td>22.71</td>
<td>0.17</td>
<td>60</td>
</tr>
<tr>
<td>“4%”</td>
<td>1915.299</td>
<td>80.925</td>
<td>0.06</td>
<td>10</td>
</tr>
<tr>
<td>“0.5%”</td>
<td>1700</td>
<td>8.660</td>
<td>0.17</td>
<td>45</td>
</tr>
</tbody>
</table>

The second method is to put the desired amounts of bismuth and lead in a round-bottom flask that is being evacuated by a turbo pump. The technique can make large quantities of doped bismuth, which can be divided into several ingots for many, more uniform evaporations. The bismuth and lead is annealed with a heat gun to remove any adsorbed gases. While still being pumped, the flask is blow-torched at the neck. When the glass at the neck softens enough to be manipulated, the flask is twisted and manipulated.

\textsuperscript{16}I tended to hold the beaker in the bath with my hand.
\textsuperscript{17}This prevents the chip from slamming into the side of the beaker and possibly destroying the probe.
\textsuperscript{18}This can be hard to see.
\textsuperscript{19}I tended to do this part in the lab, despite the acetone smell. The aluminum is to avoid taking the paint off the tables in the lab.
pulled so that the bismuth and lead are sealed inside under vacuum. Then the flask
is placed in a furnace at about 400 °C for an hour. This was done with the “0.1%”
doping with 9430 mg of bismuth and 9.547 mg of lead.

G.1.7 Atomic Absorption Spectroscopy

To check the lead concentration in a 0.1% nominal acceptor concentration ingot, I
went to use the atomic absorption spectroscopy (AAS) apparatus in the analytical
chemistry lab in SFU’s Department of Chemistry. Paul Mulyk was kind enough to
run the analysis for me. In brief, one form of AAS works by atomizing a substance
via flame and examining the absorption spectrum of the atomized material.

The procedure for my specific case is as follows:

1. Dissolve doped bismuth in concentrated nitric acid.

2. Dilute solution to 100 mL using 2% nitric acid solution.

3. Dilute further if required to reach sensitivity range of atomic absorption (about
   0.1 to 12 µg/mL; note: if you end up doing this, the chemists will report results
   in units of ppm = µg/mL).

4. Prepare calibration standards using Pb(NO₃)₂ powder.

The atomic absorption measurements approximately confirmed the lead concen-
tration, though the measured lead concentration was 5–10% higher than expected.

G.2 Novel Probe Production

Most of this is to be done in the 4D LABS cleanroom.

G.2.1 Bottom Gold Pads

1. Clean substrate.

   i) Acetone rinse.
ii) IPA rinse.

iii) Blow dry with N₂.

2. Bake dry.

i) Hot plate, 180 °C, 5 minutes.

ii) Cool, 3 minutes.

3. Spin coat and pre-bake ~650 nm of LOR (5A).

i) Mount wafer in Cee #1 with vacuum chuck. Centre the wafer.

ii) Turn on vacuum.

iii) Pipette LOR (5A) onto wafer.

iv) Spin recipe: 500 RPM spread of resist for 5 s, 2200 RPM spin coat for 45 s.

v) Pre-bake: 180 °C, 5 minutes on hot plate.

vi) Cool, 3 minutes.

vii) Turn off vacuum and clean spin coater.

4. Spin coat and pre-bake ~1000 nm of AZ MiR 703 19 cps.

i) Mount wafer in Cee #1 with vacuum chuck. Centre the wafer.

ii) Turn on vacuum.

iii) Pipette AZ 703 resist onto wafer.

iv) Spin recipe: static dispense for a few seconds, 4000 RPM spin coat for 55 s.

v) Pre-bake: 90 °C, 60 seconds on hot plate.

vi) Cool, 3 minutes.

vii) Turn off vacuum and clean spin coater.

5. Align mask using OAI aligner.

6. Expose films for 5 seconds to near UV.

7. Post-bake.

\[\text{In my work, I followed this with another pre-bake of 110 °C for 60 seconds, then cooling for 3 minutes, but this appears to come from a misreading of the instructions. That should be done only after exposure.}\]
i) Post-bake: 110 °C, 60 seconds on hot plate.

ii) Cool, 3 minutes.

8. Develop.

   i) Swirl AZ 300 developer around dish containing wafer.

   ii) Once developed, place wafer in dish of deionized water.

   iii) Rinse with deionized H₂O.

   iv) Blow dry with N₂.

9. Check resist pattern under microscope.

10. Descum on PEII-A plasma stripper.

    • Pressure: 275 mTorr (O₂)

    • Power: 50 W

    • Etch time: 30 s

11. Metal deposition using e-beam evaporator.

   i) Ti adhesion layer: 5 nm.

   ii) Au electrode layer: 200 nm.

12. Lift-off.

   i) Place wafer in dish of acetone in ultrasonic bath.

   ii) Sonicate until unwanted gold removed (~5–10 minutes).

   iii) Change dishes of acetone to remove Au particles.

   iv) Immerse wafer in Microposit 1165 remover to dissolve LOR.

   v) Spray wafer with acetone and water to disperse Au particles from lift-off.

   vi) Rinse in dish of water.

13. Check Au pattern under microscope.
G.2.2 Trilayer Deposition

The Al/SiO$_2$/Al trilayer is to be deposited on 4D LABS’ PVD4 system. I did not include instructions for using the PVD4 system because they are rather lengthy, detailed, and not particularly important to the processing.

1. If substrate is not clean, rinse with acetone and IPA, blow dry with N$_2$, then bake dry at 180 °C on hot plate for 5 minutes.

2. Load into PVD chamber. Make sure before beginning deposition that chamber gets close to base pressure.

3. Deposit first Al layer.
   - DC power: 350 W
   - Pressure: 3 mTorr (Ar)
   - Pre-sputter: 3 min
   - Deposition time$^{21}$: 220 s
   - Substrate rotation: 20 RPM

4. Deposit SiO$_2$ layer.
   - RF power: 350 W
   - Pressure: 3 mTorr (97% Ar, 3% O$_2$)
   - Pre-sputter: 10 min
   - Deposition time: 3900 s
   - Substrate rotation: 20 RPM

5. Deposit second Al layer.
   - DC power: 350 W
   - Pressure: 3 mTorr (Ar)
   - Pre-sputter: 3 min
   - Deposition time: 220 s
   - Substrate rotation: 20 RPM

$^{21}$This is based on a deposition rate measured by the tool owner in 4D LABS.
G.2.3 Trilayer Resist Mask

1. Spin coat and pre-bake AZ MiR 703 19 cps.
   i) Mount wafer in Cee #1 with vacuum chuck. Centre the wafer.
   ii) Turn on vacuum.
   iii) Pipette AZ 703 resist onto wafer.
   iv) Spin recipe: static dispense for a few seconds, 4000 RPM spin coat for 55 s.
   v) Pre-bake: 90 °C, 90 seconds on hot plate.
   vi) Cool, 3 minutes.
   vii) Turn off vacuum and clean spin coater.

2. Align mask using OAI aligner. Use left side alignment marks.

3. Expose films for 4 s to near-UV.

4. Post-bake.
   i) Hot plate: 110 °C, 90 s.
   ii) Cool: 3 minutes.

5. Develop.
   i) Put wafer in dish of AZ Diluted Developer 1:1, swirl for about 60 s.
   ii) Rinse thoroughly with deionized water.
   iii) Blow dry with N2.

6. Check resist pattern under microscope.

G.2.4 Trilayer Etching

The details here should be optimized further. Specific etch recipes follow the first list.

1. Al etch on RIE122.

---

22RIE1 and RIE2 are specific reactive ion etch machines in the 4D LABS cleanroom.
2. (Optional) Use probe station to check top layer of etched parts does not conduct.

3. SiO₂ etch on RIE2.

4. (Optional) Use probe station to check top layer of etched parts conducts.

5. Al etch on RIE1.

6. Check pattern under microscope. Use probe station to check for conducting parts on the etched areas.

7. Etch again as required.

8. Flood expose the remaining resist to near UV for 5 s using the OAI aligner.

9. Dissolve the remaining resist with AZ Diluted Developer 1:1 for as short a time as possible, watching for when no more resist is dissolving.

Al RIE recipe

1. Pump down chamber.

2. Set chamber to 500 mTorr pressure with 50 sccm of N₂ for 3 minutes.

3. Pump down chamber, stop N₂ gas.

4. Start gases (35 sccm BCl₃, 10 sccm Cl₂), wait 30 s.

5. Set pressure to 50 mTorr. Wait 1 minute.

6. Turn on RF power to 100 W for 90 s.

7. Turn off RF power, stop gas flow. Pump to high vacuum.

8. Set flow to 50 sccm N₂, 500 mTorr pressure for 1 minute.

9. Stop flow, pump to high vacuum.

10. Repeat step 8.

11. Stop gas.
SiO$_2$ RIE recipe

1. Pump down chamber.
2. Start CF$_4$ gas, 30 sccm, wait 30 s.
3. Set pressure to 30 mTorr. Wait 1 minute.
4. Turn on RF power to 100 W for 4.5 minutes.
5. Turn off RF power, stop gas flow. Pump for 2 minutes.
6. Set flow to 50 sccm Ar, 500 mTorr pressure for 1 minute.
7. Stop flow, pump for 1 minute.

G.2.5 Top Gold Pads

This procedure is substantially the same as that for the bottom gold pads, but without the LOR step because the top pads do not need a gentle taper and without the plasma stripping step because I am not sure how well the trilayer holds up to that kind of treatment.

1. Clean substrate.
   
   i) Acetone rinse.
   
   ii) IPA rinse.
   
   iii) Blow dry with N$_2$.
2. Bake dry.
   
   i) Hot plate, 180 °C, 5 minutes.
   
   ii) Cool, 3 minutes.
3. Spin coat and pre-bake ~1000 nm of AZ MiR 703 19 cps.
   
   i) Mount wafer in Cee #1 with vacuum chuck. Centre the wafer.
ii) Turn on vacuum.

iii) Pipette AZ 703 resist onto wafer.

iv) Spin recipe: static dispense for a few seconds, 4000 RPM spin coat for 55 s.

v) Pre-bake: 90 °C, 90 seconds on hot plate.

vi) Cool, 3 minutes23.

vii) Turn off vacuum and clean spin coater.

4. Align mask using OAI aligner. Use right side alignment marks.

5. Expose films for 4 seconds to near UV.

6. Post-bake.

   i) Post-bake: 110 °C, 90 seconds on hot plate.

   ii) Cool, 3 minutes.

7. Develop.

   i) Swirl AZ 300 developer around dish containing wafer.

   ii) Once developed, place wafer in dish of water.

   iii) Rinse with deionized H₂O.

   iv) Blow dry with N₂.

8. Check resist pattern under microscope.

9. Flood expose the remaining resist to near UV for 5 s using OAI aligner.

10. Metal deposition using e-beam evaporator.

    i) Ti adhesion layer: 5 nm.

    ii) Au electrode layer: 200 nm.

11. Lift-off.

    i) Place wafer in dish of acetone in ultrasonic bath.

    ii) Sonicate until unwanted gold removed (~5–10 minutes).

---

23In my work, I followed this with another pre-bake of 110 °C for 60 seconds, then cooling for 3 minutes, but this appears to come from a misreading of the instructions. That should be done only after exposure.
iii) Change dishes of acetone to remove Au particles.
iv) Immerse wafer in Microposit 1165 remover to dissolve LOR.
v) Spray wafer with acetone and water to disperse Au particles from lift-off.
vi) Rinse in dish of water.

12. Check Au pattern under microscope.

13. Use probe station to check connections between contacts.

G.2.6 Dicing

The 4D LABS wafer scriber (Tempress 1713-10C) should be fairly straightforward to use with training from cleanroom staff. The only note I will make here is to use 100 g of “force” and then cleave over some convenient edge.

It appears 4D LABS now also has a dicing saw.

G.2.7 Active Region Fabrication

This active region fabrication procedure was done on a wafer that underwent most of the standard procedure except the trilayer etching, which was done with wet chemicals, specifically Transene aluminum etchant A and a 10:1 HF-based buffered oxide etch. It is possible that the procedure should be different for RIE-based etching.

1. Spin coat wafer in $\sim 4000$ Å of PMMA (950) using Cee#2.
   i) Mount wafer in Cee #2 with vacuum chuck. Centre the wafer.
   ii) Turn on vacuum.
   iii) Pipette PMMA solution onto wafer.
   iv) Spin recipe: 500 RPM spread of resist for 5 s, 3000 RPM spin coat of 6% PMMA solution in anisole for 45 s.
   v) Pre-bake on hot plate$^{24}$: 180 °C, $\sim 90$ s.
   vi) Cool, 3 minutes.

$^{24}$Pre-heat before spin-coating.
vii) Turn off vacuum and clean spin coater.

2. Carbon coating: when under vacuum, spark carbon electrodes for about 45 s, sharpening electrodes if necessary.

3. Trilayer milling with FEI Strata Dualbeam 235. Throughout, try to minimize possible exposure of PMMA by taking single images, using low magnification, and by working quickly.
   
i) Load sample and pump down chamber.
   
ii) Using 10 or 15 kV electrons and SED detector, locate particles on the surface to set the focus and correct for astigmatism. If necessary, locate the gold electrodes first.
   
iii) Align sample at eucentric height.
   
iv) Find trilayer features with electron imaging. A 52° tilt may help. Set key locations to memory: active region, test milling pad.
   
v) Test milling using small ion beam aperture: 10 pA or smaller.
   
vi) Use EDX to check elemental composition of the milled hole. Mill again if required.
   
    vii) Mill the active region.
   
    viii) Vent chamber and unload sample.
   
4. Evaporate bismuth and lift off film as in the conventional probe case.

5. Check electrical connections with cleanroom probe station.

G.2.8 Finishing

1. Glue probe to carrier chip.

2. Wire bond the probe to carrier chip pads.

3. Polish away excess silicon near Hall probe tip.

---

25 The carbon coater is not a precision instrument by any stretch of the imagination. Expect variations from run to run, whether from sample placement changes the inherent reproducibility problem in the crude controls with rapidly eroding carbon source pieces. I suspect I overdid it on the carbon, but I wanted to compensate for edge effects of milling and milling due to imaging.

26 It is worth trying to electrically connect the carbon layer with the SEM stub using carbon paint. This should help reduce sample charging somewhat.

27 This can be done by scratching the surface near, but not at, the active region.
G.3 Carrier Chip Fabrication Recipe

The recipe is the same as that for making the bottom gold pads from section G.2.1 except without the LOR step because making a gentle taper is not a concern here.

As with the novel probes, the carrier chip wafer can then be diced using the wafer scribe in the 4D LABS cleanroom, as noted in section G.2.6.
Appendix H

LabVIEW and the Prologix GPIB-USB Controller

At the beginning of the project, I used a Prologix GPIB-USB Controller 3.12, an early model controller made to undercut National Instruments’ own expensive GPIB-USB controllers. It is annoyingly close to compatible. VIs typically will not work with our Prologix controller without significant modification. We later got a much better official National Instruments GPIB-USB interface. If you must use the Prologix one, here are some tips on how to get it to work.

H.1 Mac Installation

This is an addition to step 5 of the installation procedure for the GPIB USB Controller 3.x Manual.

Not noted in the Mac installation instructions for version 3.12 of the controller is where you need to configure the settings for your virtual COM port to use the GPIB-USB controller. You need to access Applications → National Instruments → NI-VISA → NI-VISA Configuration.

You will be asked whether you want to query your system for dynamic resources. Click “yes.” In the Resource Editor tab, you will find a list of serial ports. Find out which one is the serial port for the GPIB-USB controller by choosing an available port and
The GPIB-USB controller does not accept native LabVIEW commands to do simple things like polling for device errors or triggering data collection. Instead, the controller accepts separate text commands to do such tasks. VIs must be coded specifically to send the controller the required text strings. Standard VI drivers must then be modified to replace the unacceptable native LabVIEW commands with the user-made VIs.

You also must read the manual to figure out how your lab instruments signal the end of communication or end of a “sentence.” The EOI and EOS commands in the GPIB-USB controller change the communication standard recognized by the controller. I found it useful to make special “setup” VIs to initialize the GPIB-USB controller with the correct EOI and EOS settings for proper communication.
Appendix I

Additional References

The Australian Microscopy and Microanalysis Research Facility web site [82] has good EDX learning resources.

Two papers authored by Williams et al. [83, 84] are an excellent starting point for developing etch recipes for device fabrication.