Performance Improvements for Particle Tracking Detectors in Extreme Rate and Radiation Environments

Veronica Wallängen
Performance Improvements for Particle Tracking Detectors in Extreme Rate and Radiation Environments

Veronica Wallängen

Abstract
In order to increase its discovery potential, the Large Hadron Collider (LHC) at CERN is being transformed into a higher luminosity machine expected to be operational around 2026. The number of particle collisions will increase by a factor of 10 beyond the current design value, which means that the detectors installed around the LHC are facing various new challenges. The most demanding challenges include handling the enormous data quantities that will be transferred from the front-end readout modules at significantly higher rates than previously, as well as the radiation effects that arise as a consequence of the intense particle flow and that cause damage to sensor elements and electronics.

At the ATLAS experiment, a multipurpose detector operating at the LHC, the impact of the luminosity increase is especially severe for the silicon pixel tracking detector, being the central subsystem located closest to the particle interaction point and therefore exposed to the highest radiation dose and hit density. The extreme radiation doses that the pixel modules will be subject to will cause deformation of the sensor material structure and thus loss of the signals, which after subsequent digitization by the pixel readout chip must be transferred over relatively long distances through a low-mass data link, causing further signal distortion.

The work presented here addresses both major challenges described and outlines solutions for the upcoming upgrade of the ATLAS pixel detector system with regards to these. Firstly, it is demonstrated how improved accuracy of detector simulations and reconstruction of particle trajectories through the detector can be achieved as higher particle fluences are approached, by modeling radiation damage effects that occur in the pixel sensors. Secondly, it is shown how a receiver integrated circuit utilizing an industry standard technique novel within high-energy physics applications has been designed as an integral part of a high-speed transmission link to efficiently restore the signal quality in order to achieve adequate data readout rates.

Stockholm 2019

http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-175161

ISBN 978-91-7797-910-4

Department of Physics

Stockholm University, 106 91 Stockholm
PERFORMANCE IMPROVEMENTS FOR PARTICLE TRACKING DETECTORS IN EXTREME RATE AND RADIATION ENVIRONMENTS

Veronica Wallängen
Performance Improvements for Particle Tracking Detectors in Extreme Rate and Radiation Environments

Veronica Wallängen
Sammanfattning

För att öka potentialen hos Large Hadron Collider (LHC) vid CERN att göra nya upptäckter genomgår den en uppradering till en maskin med högre luminositet, vilken väntas vara klar för drift runt år 2026. Antalet partikelkollisioner kommer att öka med en faktor 10 bortom vad den nuvarande acceleratorn är designad för, vilket innebär att detektorerna som är installerade runt LHC står inför nya utmaningar och i sin tur behöver upprgraderas. De främsta av dessa utmaningar inkluderar att hantera de enorma mängder data som måste överföras från detektormodulerna i betydligt högre hastigheter än tidigare, samt de strålningseffekter som uppstår som en konsekvens av det intensiva partikelflödet och som orsakar skador på sensorelement och elektronik.

För ATLAS-experimentet, en multifunktionell detektor vid LHC, är inverkan av luminositetsökningen särskilt påtaglig för pixeldetektorn, vilket är det centrala sub-systemet som är placerat närmast kollisionspunkten och därför utsatt för den högsta stråldosen med flest genomkorsande partiklar per ytenhet. De extrema stråldoser som pixeldetektorn kommer att utsättas för orsakar deformeringshossensormaterials struktur, något som degraderar signalerna, vilka efter påföljande digitalisering av utläsningschippet måste skickas över relativt långa avstånd genom en tunn datalänk, vilket orsakar ytterligare signalstörningar.

I den här avhandlingen presenteras lösningar för båda de stora utmaningar inför den kommande högluminositetsuppraderingen av ATLAS pixeldetektorsystem. Dels presenteras en noggrann modell av de stålskador som sker i pixeldetektorn, vilket förbättrar precisionen hos rekonstruktionen av partikelbanorna genom detektorn vid högre partikelflöden. Vidare presenteras designen av en integrerad krets som ska agera mottagarenhet för pixelutläsningslänken och som använder en nyutvecklad teknik, standardiserad inom industriapplikationer men hittills inte använd inom högenergifysik, i syfte att effektivt återställa signalkvaliteten för att möjliggöra dataöverföring i höga hastigheter.
Acknowledgments

Many people have contributed to the research efforts presented in this thesis, both directly and indirectly, and I am deeply grateful for all these contributions. I would like to start by thanking my supervisor at Lawrence Berkeley National Laboratory (LBNL), Maurice Garcia-Sciveres, who is always inspirational and helpful in providing guidance, feedback and new ideas. I would also like to thank my supervisors at Stockholm University (SU), Samuel Silverstein at the instrumentation physics division and Sara Strandberg at the elementary particle physics division, for all their kind help and support during my first year of PhD studies at SU as well as after my transition to LBNL for my research exchange. Many thanks to all my supervisors for their flexibility in tailoring my PhD studies according to my interests, wishes, strengths and weaknesses, to make it both a challenging as well as an enjoyable experience.

Further, my sincerest thanks are dedicated to all my colleagues in the instrumentation physics division and the ATLAS group at SU for a happy and inspiring welcome into my PhD studies. I would like to extend special thanks to the persons previously and presently working on the TileCal digitizer system, Steffen Muschter, Henrik Åkerstedt, Eduardo Valdes Santurio and Christian Bohm who have all been of great assistance in helping me understand the system better to be able to carry out my ATLAS authorship qualification task. Additional thanks to the TileCal test bench team at CERN and participating institutions, especially the team leaders at the time Carlos Solans Sanchez and Giulio Usai, for regularly providing constructive feedback and helping out with practicalities to enable remote testing, something that was crucial to complete the project successfully.

I would also like to thank the entire LBNL ATLAS group of interns, undergraduate students, graduate students, visiting researchers, post-docs and senior researchers for your constant curiosity and valuable feedback during our meetings, always steering the research along an interesting path. Special thanks to Ben Nachman for guiding me in the work on radiation damage modeling for 3D pixel sensors and for efficiently leading the work of the group forward alongside Marco Bomben. Many thanks also to Gilberto Giugliarelli at University of Udine, who collaborated with me on the 3D digitizer model development, without your excellent work on producing TCAD simulation data, the creation of the digitizer model would not have been possible. Thanks to Alex Zeng Wang at University of Wisconsin-Madison for further development and validation of the 3D digitizer model, and for working on integrating it into the Athena framework.

I would like to thank the office of Visiting Scholar and Postdoc Affairs at University of California Berkeley (UCB) for your help in practical arrangements related to my stay as a visiting research scholar. In addition, I would like to thank everyone at Berkeley Wireless Research Center (BWRC) at UCB who has been involved in the design of the discrete equalizer stage which became part of the GBCR chip, in particular Elad Alon, Jaeduk Han and Eric Chang for generously spending time teaching me about high-speed data transmission in general and about the tools developed by BWRC in particular. Without your help, the fabricated ASIC including the discrete equalizer would not have been realized. The same goes for the IC designers at LBNL, Carl Grace who provided invaluable help in completing the layout of the discrete equalizer as well as the essential configuration DACs, and Amanda Krieger who helped teaching me about the transistor simulation models. With regards to the
Acknowledgments

equalizer project, I would also like to extend my warmest thanks to everyone working on data transmission for the ATLAS Phase-II Upgrade at the University of Glasgow, University of New Mexico, University of California Santa Cruz and SLAC National Accelerator Laboratory who have provided me with cable assemblies and connectors to carry out the cable characterization for the high-speed data link, as well as interesting questions, comments and suggestions to take my work forward in the right direction and deepen my knowledge about this topic. Further, I would like to thank my colleagues at Southern Methodist University, who were responsible for the design and fabrication of the GBCR ASIC, in particular Datao Gong, Chufeng Chen, Jingo Ye, Tiankuan Liu and Peilong Wang for their hard work on integrating my part of the design into the chip and patiently providing help and support.

Special thanks go to the Hans Werthén Fund through the Royal Swedish Academy of Engineering Sciences as well as to the Ingegerd & Viking Olov Björks Scholarship Trust through the Sweden-America Foundation for their generous donations, which have made financing my research exchange at LBNL possible. Greatest of thanks also to the Grace Hopper Celebration committee, VERTEX 2017 committee, as well as the Knut and Alice Wallenberg Foundation, G & E Kobbs, Kyltekniska Fonden, C A Öbergs, L & E Kinanders and Rhodins Minne for believing enough in my abilities to provide travel stipends as contributions to financing of my attendance to conferences and physics schools, which have been incredibly valuable to me. Many thanks also to the Japan Society for the Promotion of Science (JSPS) for giving me the opportunity to be a research fellow in Japan and an unforgettable experience, on a professional as well as personal level.

I am also deeply thankful that, in addition to my main field of research, I have been able to pursue my interest in medical physics during two summer internships where I got to work on the development of Positron Emission Tomography (PET) detector systems. For this, I am very grateful to Professor Taiga Yamaya and the Imaging Physics Team at National Institute of Radiological Sciences, and to Professor Juan José Vaquero Lopez and his team in the bioengineering department at Universidad Carlos III de Madrid, who kindly welcomed me into their respective research groups for my summer internships during which I learned more than I could have imagined.

And finally, last but certainly not least, endless thanks to all my friends and family for your constant love and support, whether local or from a distance, throughout my entire journey of studies and research experiences. I am especially grateful to my wonderful loving mother Elisabet, my supportive brother Patrik, and my incredible fiancé Anand, you have made all the difference.
# Contents

Sammanfattning .................................................. iii  
Acknowledgments ............................................... v

## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
<td>xvii</td>
</tr>
<tr>
<td>1 High-Energy Physics Instrumentation</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Radiation Detection</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Particle Accelerators</td>
<td>1</td>
</tr>
<tr>
<td>1.3 Tracking Detectors</td>
<td>3</td>
</tr>
<tr>
<td>2 The ATLAS Experiment at the LHC</td>
<td>5</td>
</tr>
<tr>
<td>2.1 The Large Hadron Collider (LHC)</td>
<td>5</td>
</tr>
<tr>
<td>2.2 The ATLAS Experiment</td>
<td>7</td>
</tr>
<tr>
<td>2.3 The ATLAS Pixel Detector</td>
<td>9</td>
</tr>
<tr>
<td>2.3.1 Tracking Principle</td>
<td>9</td>
</tr>
<tr>
<td>2.3.2 Semiconductor Detectors</td>
<td>11</td>
</tr>
<tr>
<td>2.3.3 Particle Sensing and Data Readout</td>
<td>14</td>
</tr>
<tr>
<td>3 The High-Luminosity LHC and the ATLAS Inner Tracker Upgrade</td>
<td>17</td>
</tr>
<tr>
<td>3.1 The High-Luminosity LHC</td>
<td>17</td>
</tr>
<tr>
<td>3.1.1 Increasing Accelerator Luminosity</td>
<td>17</td>
</tr>
<tr>
<td>3.1.2 Physics Motivation</td>
<td>18</td>
</tr>
<tr>
<td>3.1.3 Detector Challenges</td>
<td>18</td>
</tr>
<tr>
<td>3.2 The ATLAS Inner Tracker (ITk) Upgrade</td>
<td>19</td>
</tr>
<tr>
<td>3.2.1 Tracker Structure</td>
<td>19</td>
</tr>
<tr>
<td>3.2.2 Pixel Modules of the Next Generation</td>
<td>20</td>
</tr>
<tr>
<td>3.2.3 Major ITk Challenges</td>
<td>22</td>
</tr>
<tr>
<td>4 ITk Challenge 1: Radiation Damage</td>
<td>25</td>
</tr>
<tr>
<td>4.1 Radiation Damage in Silicon Devices</td>
<td>25</td>
</tr>
<tr>
<td>4.1.1 Radiation Damage Effects</td>
<td>25</td>
</tr>
<tr>
<td>4.1.2 Radiation Damage in Pixel Sensors</td>
<td>26</td>
</tr>
<tr>
<td>4.1.3 3D Silicon Sensor Technologies</td>
<td>29</td>
</tr>
<tr>
<td>4.1.4 Radiation Hardening of Readout Components</td>
<td>32</td>
</tr>
<tr>
<td>4.2 Pixel Sensor Radiation Damage Simulation</td>
<td>33</td>
</tr>
<tr>
<td>4.2.1 Modeling of Radiation Effects</td>
<td>33</td>
</tr>
</tbody>
</table>
5 ITk Challenge 2: Data Readout Speed

5.1 ATLAS ITk Readout System Architecture
5.1.1 Readout Bandwidth
5.1.2 Transmission Link System
5.1.3 Data Cable Prototypes
5.1.4 Active Cable Concept

5.2 Signal Recovery with Equalization
5.2.1 Signal Distortion
5.2.2 Equalization Techniques
5.2.3 High-Speed Link Equalizer Components

5.3 Equalizer Architecture and Circuit Implementation
5.3.1 Design Description
5.3.2 Equalizer Tuning
5.3.3 Simulations and Design Verification

5.4 Equalizer Prototype Performance
5.4.1 The GBCR Prototype ASIC and Test Board Setup
5.4.2 Summary of Test Results
5.4.3 Test Result Analysis
5.4.4 Discrete Equalizer Prototype Improvements
5.4.5 Future GBCR Prototype

5.5 Pulse Amplitude Modulation Studies

6 Summary and Conclusions
6.1 ITk Challenge 1: Radiation Damage
6.2 ITk Challenge 2: Data Readout Speed

A Readout and DAQ Systems for PET Detectors
A.1 PET Basics
A.2 Development of PET-DOI detector with TOF capability
A.2.1 Background
A.2.2 Result Summary
A.3 Readout and DAQ System for PET Scintillation Crystals
A.3.1 Background
A.3.2 Result Summary

References
## List of Figures

1.1 Fundamental particle discoveries ........................................ 2
1.2 Tracking detector block diagram ........................................ 3
1.3 Particle tracking .......................................................... 4

2.1 CERN’s accelerator complex .............................................. 6
2.2 ATLAS coordinate systems ................................................ 7
2.3 ATLAS detector ............................................................. 8
2.4 ATLAS Inner Detector ....................................................... 9
2.5 Pixelated detector vs. microstrip detector .............................. 9
2.6 Particle deflection in magnetic field .................................... 10
2.7 Semiconductor doping ..................................................... 11
2.8 p-n junction ................................................................. 12
2.9 Hybrid pixel module architecture ....................................... 14

3.1 LHC timeline ................................................................. 18
3.2 ATLAS ITk layout ........................................................... 19
3.3 Pixel modules ............................................................... 21
3.4 Simulated particle density inside ATLAS tracker ...................... 23

4.1 Simulated fluence for ITk .................................................. 26
4.2 Silicon radiation damage illustration .................................... 28
4.3 ITk sensors technologies ................................................. 29
4.4 3D pixel sensors ............................................................ 30
4.5 Lorentz force in pixel sensors ........................................... 31
4.6 EUDET telescope setup .................................................... 34
4.7 Drift and diffusion acting on charge carriers .......................... 34
4.8 Electric field in 3D pixel sensors and charge collection time ....... 35
4.9 Ramo potential ............................................................... 36
4.10 Carrier drift in electric field ............................................. 39
4.11 Time-over-Threshold (ToT) distributions ............................. 40
4.12 Charge collection efficiency vs. fluence ............................... 40

5.1 Illustration of bandwidth estimation quantities ....................... 44
5.2 Pixel hit density simulations .............................................. 45
5.3 ITk Pixel Detector readout architecture ............................... 46
5.4 Nyquist theorem ........................................................... 48
5.5 Active cable illustration .................................................. 49
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.6</td>
<td>Inter-symbol interference (ISI)</td>
<td>50</td>
</tr>
<tr>
<td>5.7</td>
<td>Effect of equalization</td>
<td>51</td>
</tr>
<tr>
<td>5.8</td>
<td>Impulse response with pre- and post-cursors</td>
<td>52</td>
</tr>
<tr>
<td>5.9</td>
<td>CTLE schematic</td>
<td>54</td>
</tr>
<tr>
<td>5.10</td>
<td>FFE block diagram</td>
<td>55</td>
</tr>
<tr>
<td>5.11</td>
<td>Description of FFE combined with DFE</td>
<td>56</td>
</tr>
<tr>
<td>5.12</td>
<td>Frequency response for channel, FFE and DFE</td>
<td>57</td>
</tr>
<tr>
<td>5.13</td>
<td>Block diagram of GBCR chip</td>
<td>58</td>
</tr>
<tr>
<td>5.14</td>
<td>Discrete equalizer block diagram</td>
<td>58</td>
</tr>
<tr>
<td>5.15</td>
<td>Equalizer latches</td>
<td>59</td>
</tr>
<tr>
<td>5.16</td>
<td>Discrete equalizer taps</td>
<td>60</td>
</tr>
<tr>
<td>5.17</td>
<td>Transfer functions of ATLAS ITk cable prototypes</td>
<td>61</td>
</tr>
<tr>
<td>5.18</td>
<td>eGUllizer online modeling tool</td>
<td>61</td>
</tr>
<tr>
<td>5.19</td>
<td>Process corner simulations</td>
<td>63</td>
</tr>
<tr>
<td>5.20</td>
<td>Equalizer parasitic extraction simulations</td>
<td>63</td>
</tr>
<tr>
<td>5.21</td>
<td>Temperature simulations</td>
<td>63</td>
</tr>
<tr>
<td>5.22</td>
<td>Supply voltage simulations</td>
<td>64</td>
</tr>
<tr>
<td>5.23</td>
<td>BER simulations</td>
<td>65</td>
</tr>
<tr>
<td>5.24</td>
<td>GBCR block diagram</td>
<td>65</td>
</tr>
<tr>
<td>5.25</td>
<td>GBCR prototype test board</td>
<td>66</td>
</tr>
<tr>
<td>5.26</td>
<td>GBCR test setup</td>
<td>66</td>
</tr>
<tr>
<td>5.27</td>
<td>GBCR prototype eye diagrams</td>
<td>67</td>
</tr>
<tr>
<td>5.28</td>
<td>Channel 1 data recovery</td>
<td>68</td>
</tr>
<tr>
<td>5.29</td>
<td>GBCR layout and illustration of FIB modification</td>
<td>69</td>
</tr>
<tr>
<td>5.30</td>
<td>Discrete equalizer simulation results</td>
<td>69</td>
</tr>
<tr>
<td>5.31</td>
<td>Discrete equalizer prototype test results</td>
<td>70</td>
</tr>
<tr>
<td>5.32</td>
<td>External clock source vs. recovered clock</td>
<td>70</td>
</tr>
<tr>
<td>5.33</td>
<td>Discrete equalizer simulation results</td>
<td>71</td>
</tr>
<tr>
<td>5.34</td>
<td>Dynamic logic</td>
<td>71</td>
</tr>
<tr>
<td>5.35</td>
<td>Supply voltage simulations</td>
<td>72</td>
</tr>
<tr>
<td>5.36</td>
<td>External amplifier simulations</td>
<td>72</td>
</tr>
<tr>
<td>5.37</td>
<td>Effect of temperature variations</td>
<td>73</td>
</tr>
<tr>
<td>5.38</td>
<td>Effect of supply voltage variations in simulation</td>
<td>73</td>
</tr>
<tr>
<td>5.39</td>
<td>Effect of supply voltage variations in hardware</td>
<td>74</td>
</tr>
<tr>
<td>5.40</td>
<td>Data rate limits</td>
<td>74</td>
</tr>
<tr>
<td>5.41</td>
<td>Hardware test vs. simulation results</td>
<td>75</td>
</tr>
<tr>
<td>5.42</td>
<td>Improved equalizer design</td>
<td>75</td>
</tr>
<tr>
<td>5.43</td>
<td>Updated equalizer simulation results</td>
<td>76</td>
</tr>
<tr>
<td>5.44</td>
<td>Updated baseline readout scheme</td>
<td>77</td>
</tr>
<tr>
<td>5.45</td>
<td>Cable loss vs. data rate</td>
<td>78</td>
</tr>
<tr>
<td>5.46</td>
<td>PAM4 illustration</td>
<td>78</td>
</tr>
<tr>
<td>5.47</td>
<td>PAM4 simulations</td>
<td>79</td>
</tr>
</tbody>
</table>

### A.1
- HEP vs. medical imaging detector systems | 83 |

### A.2
- Positron Emission Tomography (PET) | 84 |

### A.3
- PET flood histogram | 85 |

### A.4
- PET module crystal look-up table | 86 |

### A.5
- OpenPET | 87 |
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.6</td>
<td>DOI-PET detector module</td>
<td>88</td>
</tr>
<tr>
<td>A.7</td>
<td>TOF-PET detector module</td>
<td>89</td>
</tr>
<tr>
<td>A.8</td>
<td>TOF-PET test setup</td>
<td>89</td>
</tr>
<tr>
<td>A.9</td>
<td>Correction results</td>
<td>90</td>
</tr>
<tr>
<td>A.10</td>
<td>Corrected flood histogram</td>
<td>91</td>
</tr>
<tr>
<td>A.11</td>
<td>PET crystals</td>
<td>92</td>
</tr>
<tr>
<td>A.12</td>
<td>PicoDigitizer system setup</td>
<td>92</td>
</tr>
<tr>
<td>A.13</td>
<td>PicoAnalyzer GUI</td>
<td>93</td>
</tr>
</tbody>
</table>
# List of Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADC</td>
<td>Analog-to-Digital Converter</td>
</tr>
<tr>
<td>AFC</td>
<td>Automatic Frequency Calibration</td>
</tr>
<tr>
<td>AFP</td>
<td>ATLAS Forward Proton Detector</td>
</tr>
<tr>
<td>ASIC</td>
<td>Application Specific Integrated Circuit</td>
</tr>
<tr>
<td>AWG</td>
<td>American Wire Gauge</td>
</tr>
<tr>
<td>BER</td>
<td>Bit Error Ratio</td>
</tr>
<tr>
<td>BWRC</td>
<td>Berkeley Wireless Research Center</td>
</tr>
<tr>
<td>CCE</td>
<td>Charge Collection Efficiency</td>
</tr>
<tr>
<td>CDR</td>
<td>Clock Data Recovery</td>
</tr>
<tr>
<td>CERN</td>
<td>Conseil Européen pour la Recherche Nucléaire/European Laboratory for Particle Physics</td>
</tr>
<tr>
<td>CMOS</td>
<td>Complementary Metal Oxide Semiconductor</td>
</tr>
<tr>
<td>DAC</td>
<td>Digital-to-Analog Converter</td>
</tr>
<tr>
<td>DAQ</td>
<td>Data Acquisition</td>
</tr>
<tr>
<td>DEQ</td>
<td>Discrete Equalizer</td>
</tr>
<tr>
<td>DFE</td>
<td>Decision Feedback Equalizer</td>
</tr>
<tr>
<td>DOI</td>
<td>Depth-Of-Interaction</td>
</tr>
<tr>
<td>e-h</td>
<td>electron-hole</td>
</tr>
<tr>
<td>FE</td>
<td>Front-End</td>
</tr>
<tr>
<td>FEOL</td>
<td>Front End Of Line</td>
</tr>
<tr>
<td>FFE</td>
<td>Feed-Forward Equalizer</td>
</tr>
<tr>
<td>FIB</td>
<td>Focused Ion Beam</td>
</tr>
<tr>
<td>GBCR</td>
<td>Gigabit Cable Receiver</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>GUI</td>
<td>Graphical User Interface</td>
</tr>
<tr>
<td>HEP</td>
<td>High Energy Physics</td>
</tr>
<tr>
<td>HL-LHC</td>
<td>High Luminosity Large Hadron Collider</td>
</tr>
<tr>
<td>I/O</td>
<td>Input/Output</td>
</tr>
<tr>
<td>IBL</td>
<td>Insertable B-Layer</td>
</tr>
<tr>
<td>IC</td>
<td>Integrated Circuit</td>
</tr>
<tr>
<td>ITk</td>
<td>Inner Tracker</td>
</tr>
<tr>
<td>LBNL</td>
<td>Lawrence Berkeley National Laboratory</td>
</tr>
<tr>
<td>LFS</td>
<td>Lutetium Fine Silicate</td>
</tr>
<tr>
<td>LHC</td>
<td>Large Hadron Collider</td>
</tr>
<tr>
<td>LOR</td>
<td>Line-Of-Response</td>
</tr>
<tr>
<td>LYSO</td>
<td>Lutetium Yttrium Orthosilicate</td>
</tr>
<tr>
<td>MPPC</td>
<td>Multi-Pixel Photon Counter</td>
</tr>
<tr>
<td>MPV</td>
<td>Most Probable Value</td>
</tr>
<tr>
<td>NIEL</td>
<td>Non-Ionizing Energy Loss</td>
</tr>
<tr>
<td>NIRS</td>
<td>National Institute of Radiological Sciences</td>
</tr>
<tr>
<td>NMOS</td>
<td>N-type Metal Oxide Semiconductor</td>
</tr>
<tr>
<td>NRZ</td>
<td>Non-Return-to-Zero</td>
</tr>
<tr>
<td>PAM</td>
<td>Pulse Amplitude Modulation</td>
</tr>
<tr>
<td>PET</td>
<td>Positron Emission Tomography</td>
</tr>
<tr>
<td>PMOS</td>
<td>P-type Metal Oxide Semiconductor</td>
</tr>
<tr>
<td>PRBS</td>
<td>Pseudo-Random Binary Sequence</td>
</tr>
<tr>
<td>SiPM</td>
<td>Silicon Photo-Multiplier</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal-to-Noise Ratio</td>
</tr>
<tr>
<td>SSLE</td>
<td>Sub-Surface Laser Engraving</td>
</tr>
<tr>
<td>TDAQ</td>
<td>Trigger and Data Acquisition</td>
</tr>
<tr>
<td>TID</td>
<td>Total Ionizing Dose</td>
</tr>
<tr>
<td>TOF</td>
<td>Time-Of-Flight</td>
</tr>
<tr>
<td>ToT</td>
<td>Time-over-Threshold</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>------------------------------------</td>
</tr>
<tr>
<td>TSMC</td>
<td>Taiwan Semiconductor Manufacturing Company</td>
</tr>
<tr>
<td>TTC</td>
<td>Timing Trigger and Control</td>
</tr>
<tr>
<td>TWP</td>
<td>Twisted-Wire Pairs</td>
</tr>
<tr>
<td>UI</td>
<td>Unit Interval</td>
</tr>
<tr>
<td>VCO</td>
<td>Voltage Controlled Oscillator</td>
</tr>
</tbody>
</table>
Preface

The work underlying this thesis has been carried out as part of a joint effort between Lawrence Berkeley National Laboratory (LBNL) and Stockholm University through the Stockholm-Berkeley Physics Instrumentation PhD Exchange Program. The thesis addresses the two most crucial challenges for the design of the ATLAS Pixel Detector to overcome in preparation for the Phase-II Upgrade, during which it is being completely remodelled for compatibility with a higher luminosity operation of the LHC; namely radiation damage considerations and increased data readout speed. The thesis structure is of compilation style type, where the first part serves to cover the necessary background information needed for the four publications contained in the later part. Chapter 1-3 will introduce the reader to the field of instrumentation for particle physics, the LHC and the ATLAS Pixel Detector in particular, as well as describe the motivation behind the high-luminosity upgrade of the LHC and the Phase-II Upgrade of the ATLAS detector. Chapter 4 and 5 will outline the basics behind radiation damage modeling for pixel sensors and high-speed signaling within high-energy physics respectively. Chapter 6 contains a summary and conclusions about the main topics covered. Following the main matter are four published articles based on these topics titled Paper I-IV.

In addition to the main research topics presented in this thesis and related to the ATLAS ITk Phase-II Upgrade, a compact description of two smaller projects carried out by the author as internships in collaboration with nuclear medicine imaging research teams within the field of positron emission tomography (PET) can be found in Appendix A.

The research presented in this thesis is a result of work conducted as part of a collective effort and has been made possible by various contributions from within the ATLAS collaboration as well as involvement from other institutes not affiliated with CERN. The author’s original contributions to the different areas covered in the thesis and to each of the articles I-IV found in the back of this document are described below. For clarity, separation into the two main topics addressed has been made.

**Topic 1: Radiation Damage to Pixel Sensors**

The author has worked on the radiation damage model for 3D pixel sensors made from silicon, and has written the algorithm for how charge carriers released through energy deposits from traversing particles are propagated through the silicon and charge induced on the electrodes as a result. The author has tested the validity of the model in simulation at different levels of radiation exposure. The model has been extended from an existing model originally developed for planar silicon sensors within the same
working group. Simulations of the electric field and Ramo potential, which are inputs to the model, have been supplied to the author by other group members. The author has created the model inside a framework dedicated for development, and it has subsequently undergone validation and migration into the main ATLAS simulation model by other group members.

Radiation damage to pixel sensors in the ATLAS detector and the digitizer model development including radiation induced effects are described in general in Paper I, while Paper II is specifically dedicated to the 3D pixel sensor radiation damage model, which has been the main work of the author. Chapter 4 gives an overview of radiation damage in pixel silicon sensors, with special focus on the 3D sensor geometry to serve as background to these papers.

**Topic 2: High-Speed Data Transmission**

The author is responsible for part of the equalizer design included in the test channel inside the receiver integrated circuit chip called the Gigabit Cable Receiver (GBCR), in terms of schematic creation and verification as well as parasitic simulations. The architecture of this discrete equalizer block is based on a design by Berkeley Wireless Research Center (BWRC), but includes many modifications and simplifications as well as transistor resizing made by the author. The layout and parasitic extraction as well as design of the configuration digital-to-analog converters (DACs) have been done by an experienced integrated circuit designer. Integration of the discrete equalizer circuit into the packaged chip has been made by a separate institute, which was also responsible for the design of other main receiver components as well as the chip fabrication. Testing and verification of the functionality of the discrete equalizer block in hardware has been done by the author. An equalizer modeling tool in the form of a graphical user interface has been developed by the author. Paper III describes the general architecture of the discrete equalizer block, while Paper IV covers the complete design of the GBCR chip.
Chapter 1

High-Energy Physics Instrumentation

1.1 Radiation Detection

Human curiosity of nature has always driven us to investigate our surroundings, and observations which require going beyond what we can directly measure using our senses have eventually driven developments of more advanced methods for studying the world around us. Today, instrumentation developments are essential to push scientific progress forward. For processes which are too brief in time or too small in space for direct observations to be possible, radiation, i.e. the emission of energy in the form of waves or particles, can be used as the observable quantity to infer information about the process using detection instruments. Radiation detectors have been utilized for major advancements within science and technology spanning a wide range of fields, like cancer diagnosis and therapy within medical applications (such as described in Appendix A) as well as for high-energy physics experiments where the fundamental laws and constituents of our universe are uncovered.

1.2 Particle Accelerators

In early theories of the structure of matter atoms were long thought to be the smallest possible constituents. Advancements in instrumentation enabled the confirmation of existence of subatomic structures and more elaborate models started to emerge. Figure 1.1 shows the mass of most massive subatomic particles known to date in relation to their respective time of confirmation of existence. As can be seen in this illustrative graph, most of the heavy particles out of this group have been discovered later throughout history\(^1\). The reason for such a trend is not a coincidence.

Many of the heavier particles in Figure 1.1 have been confirmed to exist due to the invention of accelerators for producing beams of particles with high kinetic energy, which transformed the field of particle physics into high-energy physics (HEP). The principle behind accelerator physics relies on Einstein’s theory of relativity [1] relating

\(^1\)The u, s and b quarks deviate from the trend in this graph as their experimental discoveries are not related to their masses, but mainly to quark confinement due to the strong force, i.e. the fact that quarks cannot exist as isolated entities but only be observed within composite particle structures.
Figure 1.1: Fundamental particle and antiparticle masses as function of their respective year of confirmation of existence. Massless or very low mass particles (photon, gluon, neutrino) as well as the muon are not included in this illustration. The masses are given in natural units.

mass to energy such that for a particle of mass \( m \) the total energy \( E \) in the rest frame of the particle (where it has zero momentum) can be expressed as

\[
E = mc^2
\]

where \( c \) is the speed of light. This implies that any particle we would like to study can be created from energy and also makes it evident that very high energy is needed to produce large masses and search for new physics we have not yet been able to probe with our current instruments. Additionally, since momentum energy is inversely proportional to distance, the length scale of phenomena being studied in a collider is related to the energy of the collisions. This means that the smaller the object to be viewed is, the higher is the required energy (and momentum) of the accelerated particle. Consequently, large accelerators can act as powerful microscopes to resolve small structures. However, since building such enormous accelerators is a challenging task from an engineering point of view, the existence of the Higgs boson, one of the heaviest of all known fundamental particles, was not experimentally validated until its long anticipated discovery at the European Laboratory for Particle Physics (CERN) in 2012 [4], decades after it had been theoretically predicted to exist [5][6]. As more refined technologies evolve, better instruments can be constructed and larger masses and smaller distance scales probed by colliding high energy particles inside accelerators to develop a deeper understanding of the atomic- and subatomic structures.

\[\text{Figure 1.1: Fundamental particle and antiparticle masses as function of their respective year of confirmation of existence. Massless or very low mass particles (photon, gluon, neutrino) as well as the muon are not included in this illustration. The masses are given in natural units.}\]

\[\text{mass to energy such that for a particle of mass } m \text{ the total energy } E \text{ in the rest frame of the particle (where it has zero momentum) can be expressed as}\]

\[
E = mc^2
\]

where \( c \) is the speed of light. This implies that any particle we would like to study can be created from energy and also makes it evident that very high energy is needed to produce large masses and search for new physics we have not yet been able to probe with our current instruments. Additionally, since momentum energy is inversely proportional to distance, the length scale of phenomena being studied in a collider is related to the energy of the collisions. This means that the smaller the object to be viewed is, the higher is the required energy (and momentum) of the accelerated particle. Consequently, large accelerators can act as powerful microscopes to resolve small structures. However, since building such enormous accelerators is a challenging task from an engineering point of view, the existence of the Higgs boson, one of the heaviest of all known fundamental particles, was not experimentally validated until its long anticipated discovery at the European Laboratory for Particle Physics (CERN) in 2012 [4], decades after it had been theoretically predicted to exist [5][6]. As more refined technologies evolve, better instruments can be constructed and larger masses and smaller distance scales probed by colliding high energy particles inside accelerators to develop a deeper understanding of the atomic- and subatomic structures.

\[\text{2 The discovery of the muon/antimuon is too complex to fit into this illustration [2].}\]

\[\text{3 As postulated by de Broglie [3], any particle with momentum } p \text{ is a so-called de Broglie wave of associated wavelength } \lambda \sim \frac{\hbar}{p}, \text{ which thus constitutes the minimum length scale that can be probed by accelerating particles to a given momentum.}\]
1.3 Tracking Detectors

In order to study what is taking place during a collision inside a particle accelerator some form of detector is needed to let us infer information about physics processes, despite appearing invisible to the human eye. Subatomic particles can be identified from traces they leave in detectors to discover new particles and the forces acting on them.

A common type of radiation detector within HEP is one which utilizes the phenomenon of ionization, where incident charged particles impart sufficient energy to atoms in a sensor medium to create mobile charge carriers as the excess energy introduced releases electrons from the atoms they are bound to. These charge carriers can then be collected to form a tiny pulse of electric charge, such that the signal is proportional to the particle energy lost and absorbed in the sensor medium, which provides information about where the particle was present when its motion triggered a signal.

![Diagram of a tracking detector system](image)

Figure 1.2: High level overview of a tracking detector system, from signal triggered by incident radiation to track reconstruction. A reconstructed ATLAS event is used for illustration [7].

Figure 1.2 shows a block diagram of the main components of a HEP tracking detector system. The detector structure commonly comprises several detector layers with semiconductor sensors (usually made from silicon) and readout electronics modules. The signals registered by the modules in the different detector layers constitutes a coordinates time series, or trajectory, of each particle’s path through the detector. Such a trajectory is also called a track, from where the name tracking detector, or simply tracker, is derived. To translate the electric pulse from the sensor into a format that can be interpreted by a computer, the recorded signal is treated by a readout system which can amplify, shape and digitize the individual sensor outputs associated with each event before they can be stored to memory in binary form as a series of so-called "hits", which can be computationally reconstructed to form particle tracks. Such reconstruction involves pattern recognition to identify which detector hits belong to which track, as well as track fitting to approximate the path of a particle. This tracking information in combination with signals from other types of detectors are then used for subsequent physics analysis to help unravel what took place during a collision event inside an accelerator and to identify new particles that were created, at which location they were created and from which parent particles they originated. Figure 1.3 illustrates reconstructed particle tracks in a cylindrical tracking detector.
This thesis focuses on research related to enhancing the capabilities of particle tracking detectors situated in extremely dense and high dose radiation environments, where radiation is not only the measurable quantity but also constitutes a source of damage to the detector elements.
Chapter 2

The ATLAS Experiment at the LHC

2.1 The Large Hadron Collider (LHC)

Since it started operation in 2008, the LHC at CERN has served as the largest and most powerful particle accelerator in the world and has produced data that in 2012 proved the existence of the Higgs boson [4]. The LHC is a circular accelerator with close to 27 km circumference, constructed to deliver proton-proton collisions at a maximum center-of-mass energy of 14 TeV with the purpose to study aspects of the known fundamental particles and forces, as well as search for new physics. Because it enables investigation of fundamental structures on the smallest distance scales ever probed, the LHC can be seen as a powerful microscope. For instance, the proton possesses a physical size on the order of $10^{-15}$ m, making it distinct from point-like particles such as the ones mentioned in Figure 1.1. Instead, protons are part of a particle family classified as hadrons, which are composite particles made of quarks, hence the name "Hadron Collider". Is a quark a fundamental particle or is it composite like protons? Measurements at the LHC indicate that if quarks would be composite their size must be less than $10^{-20}$ m or evidence for their physical size would have been collected at the LHC. Because they do not consist of smaller building blocks and are as far as current research indicates indivisible, quarks belong to the collection of fundamental, or elementary, particles of matter.

So how are particle accelerators used to investigate the nature of matter? Circular accelerators like the LHC first boost particles to high energies using radio-frequency cavities and powerful electromagnets generate magnetic fields to force the two opposite facing particle beams in orbit. CERN’s accelerator complex is a succession of particle accelerators, where each accelerator boosts the speed of a beam of particles before injecting it into the next one in the sequence, such that increasingly higher energies are reached. Protons are obtained by ionizing hydrogen atoms to remove electrons, and then accelerated and focused before being injected into a linear accelerator. From the linear accelerator, the protons are injected into a first stage synchrotron booster, followed by two subsequent synchrotron stages until finally being injected into the LHC, where they will circulate for around 20 minutes before reaching their maximum speed and energy (see Figure 2.1). Now traveling close to the speed of light, the proton beams are made to collide in order to form new particles from the energy of the collisions (c.f. Eq. 1.1). Surrounding each of four beam intersection...
points around the LHC ring is an advanced detector installment, which provides clues about these new particles and the complex ways in which they decay by registering details such as their mass, speed and electric charge, to help establish their identities. The two counter-rotating proton beams are forced to cross head on at regular time intervals of 25 ns, corresponding to a bunch spacing of about 7.5 m. Each beam contains 2808 bunches and each bunch is composed of 120 billion protons, resulting in approximately 600 million collisions every second [9]. Since the conservation of momentum before and after collision holds, the momentum in the transverse plane (i.e. perpendicular to the beam line, see Figure 2.2) is zero prior to collision and must so remain after the collision. Thus, transverse momentum can be used to infer the presence of particles produced in the collisions. The probability that one single collision event (bunch crossing) contains multiple proton-proton interactions is referred to as event pile-up and constitutes a challenge in the reconstruction process where each interaction must be mapped to a specific vertex. Each detector complex at the LHC is engineered and run by collaborations of scientists from institutes all over the world, and each detector and all activities related to it is referred to as an *experiment*.

![Diagram of LHC accelerator complex and injection circuits](image)

Figure 2.1: The LHC accelerator complex and injection circuits, showing the circumference of each system in parenthesis below their abbreviated names. This illustration is not drawn to scale.
The ATLAS Experiment [10] is the largest out of the four main experiments at the LHC and, like a typical advanced particle detector, it comprises several cylindrical layers of complex subdetectors, each measuring different properties of the traversing particles (see Figure 2.2). Inner detector layers form the tracker, where as little as possible of an incident particle’s energy is absorbed in order to reveal its presence without affecting its trajectory through the sensitive material. Therefore, the medium will be chosen to minimize energy loss and particle scattering. Surrounding the innermost subsystem of ATLAS are calorimeters, where in contrast to the tracker, a maximum amount of the particle’s energy should be deposited within the detector in order to stop the particle and measure its energy. In addition to these layers of subdetectors, the ATLAS detector also includes an outer system referred to as the muon spectrometer, with the purpose to identify and measure the momenta and trajectories of muons, which are particles that pass though the inner detector and calorimeters without being stopped. By combining the information from the different subsystems, a digital summary of each collision event can be constructed and used for subsequent data analysis. The main subdetector systems of ATLAS are depicted in Figure 2.2, and Figure 2.3 shows particle interactions in each subsystem.

![Figure 2.2: Illustration of the ATLAS detector and its subsystems. The ATLAS right-handed Cartesian coordinate system is shown in red. Original image source: CERN [11].](image-url)

---

Muons do not interact via the strong force and therefore very rarely undergo hard collisions with matter where they could lose significant energy.
Figure 2.3: (a) ATLAS detector cross section showing particle interactions in the different detector layers. Original image source: HYPATIA [12]. (b) Event display from ATLAS dataset showing cross-sectional view of tracks formed in the inner detector and energy deposits in the calorimeters, as well as muon tracks passing through to the muon chambers. (c) Side-view of the full ATLAS detector during the same event. Original image source: CERN [13].

Constituting the innermost subsystem, the ATLAS tracking detector itself in turn consists of three subsystems, combining gaseous detectors at the outer radii with semiconductor elements at the inner radii to provide high granularity around the vertex region. The high-precision measurements of the two inner subsystems are achieved using silicon pixel detectors followed by a silicon microstrip detector, with modules arranged in concentric cylinders around the beam axis in the barrel region and mounted on disks perpendicular to the beam axis in the end-cap region (see Figure 2.4).
The Strip Detector can provide two-dimensional position measurements by microstrips arranged in a grid pattern; however, several simultaneous particle hits will result in ambiguous readout. Therefore, better spatial resolution is required in the highest hit density regions closest to the particle interaction point, something which is achieved by the Pixel Detector through more finely segmented, or pixelated, regions resulting in unambiguous position measurements, as illustrated in Figure 2.5.

Figure 2.4: ATLAS Inner Detector. Original image source: CERN [14].

Figure 2.5: Two traversing particles causing true hits and so-called "ghost" hits in a double-sided strip detector (a) in contrast to unambiguous hits in a pixelated detector (b).

2.3 The ATLAS Pixel Detector

2.3.1 Tracking Principle

Charged particles, like electrons and positrons, transfer energy by ionization as they interact with matter along their paths inside the detector material, and the ATLAS tracking system measures the trajectories of these particles. The Pixel Detector registers the total energy deposited, $dE$, from the collected charge and by using the path length, $dx$, through the material it is possible to determine the mean rate of energy loss $dE/dx$ due to the retarding force acting on charged particles as they interact with matter. This quantity depends on the velocity of the particle and together with momentum
measurements it can be used to discern different particle species. The inner tracking system of the ATLAS detector is surrounded by magnets providing a nominal field of 2 T, hence each particle’s momentum in the transverse plane, perpendicular to the beamline and the direction of the magnetic field $B$, can be expressed as

$$p_T = qrB$$  \hspace{1cm} (2.1)$$

where $q$ is the particle charge and $r$ the radius of curvature of the path. By the information retrieved from the detectors, indicating where the particle was present, its trajectory through the magnetic field can be reconstructed. Such data will then help physicists determine its momentum and electric charge from the path curvature (see Figure 2.6).

![Figure 2.6: Illustration of deflection of positively charged (red), negatively charged (blue) and neutral (green) charges $q$ traveling with velocity $v$ in the transverse plane to a magnetic field $B$.](image)

As charge carriers are liberated by particles traversing the sensor, they experience different types of motion: drift towards the electrodes as a consequence of the electric field, with additional deviation from this path due to the effect of the magnetic field, as well as diffusion which is a random process that causes the spatial distribution of the charge cloud to spread. Depending on these effects as well as the incident angle of a traversing particle with respect to the pixel sensor, energy is typically deposited in more than one pixel, so-called charge sharing. By grouping neighboring pixels in which a hit was produced into clusters and using measurement data from all these pixels of the magnitude of the collected charge, an estimate of where the particle intersects the module can be computed using a charge interpolation technique. Thus, charge sharing significantly improves the position determination precision and supersedes the geometrical limit set by the size of the pixel, making typical space point resolutions smaller than 10 $\mu$m. In this way, the planar detector extension combined with a layered structure results in one 3D space point precision measurement per particle transversing the sensor along with an arrival time stamp sufficiently narrow to associate the point with a given bunch crossing at the 40 MHz collision rate of the LHC. Arranging many thousands or millions of these detectors around a collision point in a particle accelerator can then produce an accurate picture of the particles paths. [16]
2.3.2 Semiconductor Detectors

The sensor material being ionized inside the pixel detectors of ATLAS is silicon, which is a type of semiconductor. From a slightly simplistic viewpoint, a semiconductor can be considered to have an atomic structure where electrons surrounding the nucleus are confined to certain bands of energy. The outermost shell of an atom that can be occupied by electrons at absolute zero temperature is called the *valence band*, while the lowest energy shell with vacant states that electrons can be excited to is called the *conduction band*. The *band gap* (see Figure 2.8) is the energy difference between the valence band and the conduction band, which is an energy range where normally no electron state can exist. While an insulating material has a large band gap and the valence and conduction bands may overlap inside a material of conductor type, a semiconductor has an intermediately sized but non-zero band gap. As indicated by the name (*semi*conductor) it allows electron excitation into its conduction band at temperatures below its melting point but behaves as an insulator when approaching absolute zero.

Figure 2.7: Illustration of doped silicon, only displaying outer atomic shells containing valence electrons. A phosphorus atom acting as a donor is showed in green, and with its 5 valence electrons, one is left so weakly attached to the nucleus the thermal energy within the crystal at room temperature is sufficient to free it. A boron atom acting as as acceptor is illustrated in orange, with one electron vacancy (hole) due to having only 3 valence electrons available to form covalent bonds with the silicon atoms.

Silicon is the most widely used semiconductor in solid state electronics, mainly due to its availability and cost efficiency, as well as electrical, chemical and mechanical properties. A single silicon atom has 4 valence electrons in its outermost shell that are available for chemical bonding and are therefore shared equally among the atoms in a crystal, as each atom bonds to four other atoms through so-called *covalent bonding*.

The electrical characteristics of a silicon crystal can be altered in a process called *doping*, where impurity atoms are purposely introduced in pure silicon. The impurity atom replacing some of the silicon atoms during this process can be a *donor*, with more than 4 valence electrons, so that the 4 covalent bonds that will be formed with the neighboring silicon atoms will result in excess free electrons and the resulting crystal material is then called *n-type* (negative type) silicon. Alternatively, the silicon can be of *p-type* (positive type) if the dopant atom is instead an *acceptor*, with less than 4 valence electrons, which results in excess free electron vacancies or *holes*². Figure 2.7

---

² An electron hole can be treated as a positive charge carrier and is the absence of an electron at a position within an atom or atomic lattice where it could exist, leaving a net positive charge at that particular location.
illustrates n-type silicon doped with phosphorus as a donor and p-type silicon doped with boron as an acceptor. While the dopant atoms are stable at temperatures near absolute zero, at room temperature sufficient thermal energy is introduced to break the bonds of the donor and free electrons, as well as push electrons into the vacancies of the acceptor. The atom that supplied the electron then has an electron vacancy that can be filled by an electron from another atom in the crystal and in this way the hole can move from atom to atom, which can be viewed as positive charges moving through the crystal structure.

Figure 2.8: Illustration of a reversed bias p-n junction.

By doping one side of a piece of silicon with boron (p-type) and the other side with phosphorus (n-type), a p-n junction can be formed in the interface region where the n-type and p-type sides join. The p-n junction is the fundamental building block of semiconductor devices and is illustrated in Figure 2.8. The mobile electrons and holes will experience random thermal motion through diffusion to even out the charge carrier distribution throughout the crystal volume. As the free electrons in the n-type silicon diffuse towards the p-type side, they leave behind phosphorus ions which form a local positive electric charge in the n-type silicon. Conversely, the free holes in the p-type region leave behind negatively charged boron ions as they diffuse towards the n-type side so that a local negative electric charge is formed in the p-type silicon. These charges regions of ions will thus cause a "built-in" electric field to form across the p-n junction, which influences the direction of the charge carrier drift, with electrons being attracted towards the positive phosphorus ions and the holes being attracted towards the negative boron ions. This causes some of the electrons and holes to flow in the opposite direction to the flow caused by diffusion, as illustrated in Figure 2.8. These opposing flows eventually reach a stable equilibrium where the number of diffusing electrons exactly balances the number of electrons drifting in the opposite direction to the electric field, so that there is no net flow of current across the junction.
The established electric field is the basis of the operation of semiconductor devices. The p-n junction is depleted of free charge carriers, leaving only the fixed charges associated with the dopant atoms, and hence form a non-conductive layer where electron-hole (e-h) pairs will not recombine (i.e. annihilate so that both charge carriers are eliminated), but rather drift along the electric field lines. As a result, this depletion layer is highly resistive and behaves as a nearly perfect insulator, but the resistance can be manipulated by changing the voltage across the p-n junction, the bias voltage. Forward bias applies an external electric field in the same direction as the built-in electric field and will allow electric charge to easily flow due to reduced resistance as the junction barrier becomes smaller. In contrast, reverse bias applies an external electric field in the opposite direction of the built-in electric field by applying negative voltage to the p-type side and a positive voltage to the n-type side, as seen in Figure 2.8, to instead create minimal current flow and can therefore be used to artificially extend the depletion zone.

In semiconductor detectors, the depletion zone constitutes the sensitive volume where e-h pairs will be created as particles traverse it and hence sufficient bias voltage is required in order to enable full depletion of the sensor and thereby charge collection from a larger volume. The ionizing radiation striking the detector is measured by the amount of charge carriers released in this sensitive volume placed between electrodes, where the number of charge carriers released is proportional to the energy of the incident radiation. Under the influence of an electric field, electrons and holes travel to the electrodes, where their motion results in a measurable pulse. While particle tracking systems can also be constructed from gaseous detectors, semiconductor detectors offer better position resolution, which is crucial to achieve good track separation in order to resolve tracks inside environments with high hit density. The position resolution of a detector material depends on the ionization statistics, i.e. how many e-h pairs that are produced as a result of particles traversing the sensor. This depends on the density of the material, which determines the energy loss per unit length, as well as the band gap, which determines the charge per unit energy loss. The amount of energy required to create an e-h pair varies depending on the semiconductor material, but is significantly lower than the ionization energy for gases, which are also considerably less dense than semiconductors, making it difficult to produce thin gaseous detectors with sufficient position resolution.

Silicon is the dominant semiconductor material used for position sensitive detectors within particle physics. The ionization energy required to produce one e-h pair in silicon is \( \sim 3.65 \text{ eV} \) and the energy deposited by a particle as it passes through a silicon sensor causes approximately 80 e-h pairs per \( \mu \text{m} \) to be released. This ionization energy is high enough to ensure low dark current (charges generated in the detector without external radiation) which ensures good signal-to-noise ratio (SNR) and, in addition, the high electron mobility in silicon makes the charge collection fast and gives good time resolution. Silicon is also a material which can easily be segmented and assembled into complex geometries. Further, detectors made from silicon can be operated at room temperature, in contrast to detectors made from germanium, another popular semiconductor material with 4 electrons in its outer atomic shell and therefore possessing similar chemical properties as silicon.\(^3\)

\(^3\)Because germanium has a relatively low band gap, detectors made from this material must be cooled in order to reduce the thermal generation of charge carriers so that leakage current induced noise can be avoided.
2.3.3 Particle Sensing and Data Readout

The construction of the pixel systems currently in use at the LHC experiments pushed the development of the successful concept of so-called hybrid pixel detectors, which currently constitute state-of-the-art technology for HEP detectors. Consisting of an electronically passive silicon sensor element mechanically and electrically coupled to a separate front-end (FE) readout chip through a technique referred to as bump-bonding⁴ (see Figure 2.9), a hybrid pixel detector allows for decoupled development and optimization of the sensors and readout integrated circuits (ICs), thus resulting in a better overall performance of the detector module.

![Figure 2.9: Graphic illustration of hybrid pixel module architecture and readout.](image)

The pixel sensor elements are made from silicon and doped to function like diodes (where electricity is allowed to flow in one direction only), which are then reverse biased. When charged particles pass through the depleted silicon junction, mobile charges are released as the energy imparted will excite electrons to the conduction band, leaving vacant states in terms of holes in the valance band. The charge carriers will drift in the electric field towards the electrodes and the small ionization currents produced as this happens can be detected and measured by a readout application-specific integrated circuit (ASIC) which integrates the current pulse, amplifies and digitizes the signal, and sends the resulting hit information to the data acquisition (DAQ) system, as seen in Figure 2.9. To accommodate this, each chip has a dedicated amplification and discrimination channel per pixel front-end and performs on-chip digitization using the Time-over-Threshold technique, which digitizes analog signals by counting the number of clock cycles for which the pulse remained above a certain threshold. The hit information contains the deposited charge and crossing time, and

---

⁴Bump bonding is an attachment technology in which microscopic solder bumps are used to connect electronic chips to external circuitry, in contrast to wire bonding, where wires are used as interconnects between the chip pads and external circuitry.
the chip must hold this information until a trigger\textsuperscript{5} decision is received upon which the triggered information must be read out sufficiently fast and with minimal loss.

With an increased amount of particles colliding as the LHC will be upgraded to the High Luminosity LHC (HL-LHC) as described in Section 3.1, the high radiation doses will cause damage to the silicon sensors and readout electronics. This issue is covered in Chapter 4 as well as in Paper I and Paper II. In addition, significantly increased data readout speed must be achieved with maintained signal integrity, which imposes new challenges on the pixel system previously not encountered within the HEP community. Further details related to this will be outlined in Chapter 5 as well as in Paper III and Paper IV.

\textsuperscript{5}The role of the trigger system is to select which event data that contain interesting enough information to be sent to subsequent readout stages in order to reduce the size of the total stored data set. Further details can be found in Section 5.1.1.
Chapter 3

The High-Luminosity LHC and the
ATLAS Inner Tracker Upgrade

3.1 The High-Luminosity LHC

3.1.1 Increasing Accelerator Luminosity

The LHC will undergo a major upgrade which is expected to be finished in 2026. An overview of the timeline for the LHC upgrades since Run 1 in 2011 can be seen in Figure 3.1. After Run 3, the statistical gain in running the LHC without a considerable luminosity increase beyond its current design value will become marginal and very long run periods will be necessary to reduce statistical errors in the measurements. For this reason, the LHC will be upgraded to a higher luminosity machine, the so-called High Luminosity LHC (HL-LHC), with a goal instantaneous luminosity of \(7 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}\). The expected average integrated luminosity is 250 inverse femtobarn (fb\(^{-1}\)) per year of operation, summing up to 3000 fb\(^{-1}\) throughout the planned 12-year period of operation of the HL-LHC. Meanwhile, the upper target for the total integrated luminosity is as high as 4000 fb\(^{-1}\). As luminosity is proportional to the number of collisions occurring during a given amount of time, an increased luminosity means an increased amount of particle collisions and consequently also data collected. Specifically, the targeted highest achievable peak luminosity of the HL-LHC is computed to correspond to approximately 200 proton-proton interactions per bunch crossing. This upper pile-up limit has been the baseline requirement for the ATLAS upgrade projects. [17]

\(^1\)The inverse femtobarn (fb\(^{-1}\)) is a unit used to measure the amount of of particle-collision events per femtobarn cross-sectional area, which is in direct proportion to the time-integrated luminosity as well as the amount of data collected. As time-integrated luminosity is related to the ratio of the number of detected events to interaction cross-sectional area, where the latter is measured in units of barn (1 b = 10\(^{-24}\) cm\(^2\)). X fb\(^{-1}\) of integrated luminosity corresponds to X number of events per femtobarn of cross-section within these data accumulated by the detector.
Chapter 3. The High-Luminosity LHC and the ATLAS Inner Tracker Upgrade

3.1.2 Physics Motivation

With the unprecedented amount of integrated luminosity, the aim of the HL-LHC is to support a broader physics program under an extended period of operation. Increasing the LHC’s luminosity by close to an order of magnitude means the HL-LHC will collect around 10 times more data during its period of operation than has been recorded by the machine at present. The rigorously sized dataset accumulated is thus expected to contain information that can enable a wide range of physics to be explored and lead to great advancements in measurements of processes belonging to the Standard Model of Particle Physics, and observations of rare processes not possible at the present sensitivity level. In particular, it will enable more precise probing of the Higgs boson properties with reduced statistical uncertainties and better sensitivity, as well as provide experimental exploration of less frequent production modes and decay channels, such as the rare $H \rightarrow \mu^+ \mu^-$ decay, never before made possible by any experiment. For the couplings of the Higgs boson to other particles in their dominant decay modes, a tenfold improvement is expected with respect to the experimental precision allowed by the present dataset, and the Higgs boson self-coupling via double Higgs production can also be studied. Furthermore, the discovery potential will be increased by up to 30% in terms of extended mass reach and improved sensitivity for heavier particles compared with the baseline LHC programme. This allows several beyond the Standard Model searches, such as theories with extra dimensions, which would provide a possible explanation for why gravity is so weak compared to the other fundamental forces, and supersymmetry particles, which are expected to be produced with relatively low probability and which could explain the light mass of the Higgs boson as well as reveal a candidate dark matter particle. [19]

3.1.3 Detector Challenges

The challenging high pile-up conditions and hostile tracking environment of the HL-LHC requires extensive modifications of the detectors as well as the beam line around them inside the accelerator itself; for instance, the beam will be more intense with an increased number of protons per bunch by a factor of two above nominal value. Increased luminosity is also achieved by beam collimation using more powerful focusing magnets and cavities with the ability to "tilt" the proton bunches in each beam in order to reduce the angle at which the bunches cross to maximize their geometrical overlap at the collision point and hence also increase the collision probability. The detectors installed around the accelerator’s circumference must be upgraded to...
comply with the unique performance demands, and this upgrade will introduce several innovative technologies and concepts that are likely to be a stepping stone for the next generation of colliders. All new detector installments required to address the challenges of the HL-LHC operation will be conducted during a period called long shutdown 3 (LS3) of the accelerator and is referred to as the Phase-II Upgrade (see Figure 3.1).

### 3.2 The ATLAS Inner Tracker (ITk) Upgrade

#### 3.2.1 Tracker Structure

The ATLAS inner detector will be completely replaced during the Phase-II upgrade and transformed into a central tracking system referred to as the Inner Tracker (ITk). The ITk is being architected to achieve at minimum equivalent performance as the current detector but in the uniquely challenging environment expected at the HL-LHC, both in terms of efficiency and position resolution. Consisting of a Pixel Detector at a small radius surrounded by a Strip Detector for lower granularity measurements in the outer layers (see Figure 3.2), the ITk will have a new all-silicon design with improved momentum resolution and extended coverage in the dense tracking environment, while the material inside the tracking volume will be kept to a minimum. As both particle density and received radiation dose increase with proximity to the particle interaction point, the ITk will be subject to the most extreme conditions in terms of high pile-up and radiation damage due to its location. The innermost pixel barrel layer will be situated at a 39 mm radial distance from the beam line, while the outermost pixel layer will be located 271 mm from the beam line. [20] [21]

![Illustration of the planned ATLAS ITk layout.](image)

Figure 3.2: Illustration of the planned ATLAS ITk layout. [21]
The ITk has been designed for extended coverage in terms of pseudorapidity ($\eta$), which is a commonly used spatial coordinate to describe the angle of a moving particle relative to the beam axis.\footnote{More specifically, the definition of the pseudorapidity parameter is $\eta \equiv -\ln \left| \tan \left( \frac{\theta}{2} \right) \right|$, where $\theta$ is the angle between the particle momentum and the beam axis (positive direction), see Figure 2.2.} Along the beam axis, where the polar angle $\theta = 0$, the pseudorapidity goes towards infinity, and thus particles with high pseudorapidity values are escaping through the space in the detector undetected. To prevent this from happening and thereby improve the detection efficiency, the ITk layout is designed for reconstruction of particle trajectories up to $|\eta| < 4$, as illustrated in Figure 3.2. This will increase the angular acceptance of the detector, which becomes increasingly important the higher the luminosity.

Position measurements in the Pixel Detector will be carried out in 5 cylindrical barrel layers centered around the beam pipe, and additional end-cap ring disks. In order to optimize for hermetic coverage, the short barrel located in low $\eta$ regions will have flat modules parallel to the beam axis, while the intermediate $\eta$ regions of the 3 outer layers will be covered by inclined modules, tilted with respect to the beam axis. High $\eta$ regions are covered by end-cap modules, positioned perpendicular to the beam axis. The module positioning for the full Pixel Detector is shown in Figure 3.2.

The outer pixel barrel and end-cap detector will be operated to collect a total integrated luminosity of up to 4000 fb$^{-1}$, while the inner barrel and end-caps are designed for a maximum dose of 2000 fb$^{-1}$ and foreseen to be replaced half-way through the HL-LHC program in order to maintain good tracking performance.

### 3.2.2 Pixel Modules of the Next Generation

Like the current ATLAS Pixel Detector, the upgraded system will also make use of hybrid pixel modules with a separate sensor element and readout chip. With regards to the sensors, certain geometry goals have been targeted in order to optimize radiation hardness, granularity, material budget and processing costs. High intrinsic spatial resolution and track separation is achieved by small pixel sizes, with two alternative aspect ratios of 50×50 $\mu$m$^2$ or 25×100 $\mu$m$^2$. The rectangular pitch is an interesting option due to the improved resolution in the $x$-$y$ plane, which is important for tracks to be accurately associated with vertices. On the other hand, the larger pitch along $z$ worsens the pile-up rejection of the tracker, i.e. the ability to isolate tracks belonging to individual proton-proton collision during a given bunch crossing. While the 50×50 $\mu$m$^2$ pixel size currently constitutes the baseline design option, due to a square pitch being an easier choice for production of hybrid modules as the pixel sensors are matched to the readout channels on the pixel chip, the process of making a final decision about the location of each design in the ATLAS ITk is still on-going.

Further, the readout ASIC has also been specifically developed by the dedicated CERN collaboration RD53 \cite{15} to fulfill the postulated HL-LHC demands. Compared to previous generations of readout chips, the radiation tolerance has been increased and the read-out architecture has been improved to comply with the higher hit density and event rate. As unprecedented data quantities will be produced at the HL-LHC, the data rate and memory capability of the chip to ensure high enough latency are pushed to their limits. Therefore, the next generation readout ICs have been designed with the capability of handling hit densities above 100 hits per 25 ns bunch crossing and
the output bandwidth per front-end has been increased from 160 Mbps in the present tracker to to 5.12 Gbps in the ITk, to handle the hit rates in the innermost detector layers. In order to optimize the readout system and take maximum advantage of the available bandwidth, two types of pixel modules have been planned for different parts of the detector:

- **Quad-modules** consisting of four chips bump-bonded to a single sensor to be used in the outer flat barrel layers and in the outer end-cap rings

- **Single-chip modules** consisting of one front-end chip bump-bonded to a sensor to be used in the innermost barrel layer and the inclined part of the outer barrel

In order to accommodate the estimated data rate requirements as efficiently as possible, the digital data output stage of each front-end pixel chip has four serial lanes that can transmit data at 1.28 Gbps each for a maximum total output bandwidth of 5.12 Gbps. However, the number of enabled lanes is programmable such that the chip can be configured to use only one or two output lanes should less output bandwidth be required. A sketch of the different types of pixel modules is shown in Figure 3.3, which also illustrates how the module is mounted on the local support structure for the pixel modules, referred to as a *stave*, and how a flex is glued on the sensor backside. The flex acts as an electrical interface for hosting passive components and auxiliary chips, for instance for data aggregation and transmission, and has wire bond connections to the sensor and FE chip.

![Figure 3.3: Illustration of baseline design of ATLAS ITk pixel modules.](image-url)

The successful front-end chip development for the ITk pixel system to accommodate the challenging Phase-II upgrade demands is a consequence of the fact that the number of transistors per area in commercial electronics follows Moore’s Law, i.e. doubles every two years as feature sizes decrease. This has resulted in sufficient logic density of transistor processing technologies to meet the requirements on readout speed and radiation tolerance. The hit rate will depend on the flux of particles incident on the detector, and for the innermost pixel layer the required hit data storage density is approximately 40 Gbps/cm² with a projected hit rate close to 3 GHz/cm² [16]. The total memory needed depends on the trigger latency. The storage capacity of the chip, i.e. the maximum memory available per unit area, is limited by the logic density of the
CMOS process, which is thereby also directly related to the rate capability. Therefore, in order to handle higher hit rates and/or extend latencies, high logic density is a crucial requirement. For this reason, state-of-the-art pixel readout chip technologies within HEP make use of increasingly deeper submicron CMOS processes. The pixel front-ends currently in use in the Pixel Detector are fabricated in the 130 nm TSMC technology node, whereas a 65 nm process is used in the next generation readout chips currently under development. This node allows for a sufficient amount of logic per area and is therefore dense enough to fit the digital architecture, while having verified radiation hardness capabilities. In order to comply with the radiation tolerance requirements, a chip with radiation tolerance beyond 500 Mrad [22] was developed. This radiation tolerance is computed to be sufficient for the ITk Pixel Detector assuming the two inner layers to be replaced at least once during the HL-LHC operation. [21]

With the increased track density following the HL-LHC upgrade, the ITk must be designed to keep the detector occupancy sufficiently low for efficient high position resolution tracking performance. The Pixel Detector will cover a total active area of 12.7 m² and contain close to 5 billion readout channels distributed over around 10 000 hybrid pixel modules with 1-4 front-end ASIC(s). Each FE chip hosts a 400×384 pixel matrix on a 2 mm thick chip bottom and measures 21×20 mm². Layer 0 (innermost pixel layer) will consist of single chip modules with 150 µm thick 3D sensors to accommodate an estimated occupancy of 250 hits/event, while Layer 1 will consist of quad modules with 100 µm thick planar sensors handling approximately 35 hit pixels/event. Layer 2-4 will house quad modules with 150 µm thick planar sensors, since the hit occupancy is lower. The amount of energy deposited will increase with sensor thickness and therefore greater thickness is needed to maintain high enough SNR. [23]

3.2.3 Major ITk Challenges

Figure 3.4 shows simulations of the ATLAS tracker at a previous level of luminosity as well as the increased luminosity at the HL-LHC, where a visual comparison should make it evident that the particle density will be significantly increased. The demands facing the ATLAS ITk due to the change in conditions when the luminosity is increased for the HL-LHC operation are many, but the major challenges the tracker design will be subject to can be broken down into two essential areas:

1. **Improved radiation hardness** to avoid radiation damage to sensors and readout electronics.

2. **Increased data readout speed** to cope with the significantly higher hit density following as a consequence of the high pile-up conditions, and to avoid bandwidth saturation.

The following two chapters will treat each of these areas; more specifically, research related to modeling radiation damage effects in pixel sensors, pertaining to challenge 1 above, and designing a data transmission link for maintained signal integrity at high data rates, applicable to challenge 2.

---

3Complementary Metal-Oxide Semiconductor (CMOS) is a transistor fabrication process that uses complementary and symmetrical pairs of p-type and n-type field-effect transistors for logic functions.

4Taiwan Semiconductor Manufacturing Company (TSMC) is a fabricator of silicon integrated circuits.
3.2. *The ATLAS Inner Tracker (ITk) Upgrade*

Figure 3.4: Simulated collision event in the baseline layout of the ATLAS ITk (shown in blue) with a pile-up of 23, corresponding to previous run conditions (left), and a pile-up of 230, corresponding to HL-LHC operation (right) [24].
Chapter 4

ITk Challenge 1: Radiation Damage

4.1 Radiation Damage in Silicon Devices

4.1.1 Radiation Damage Effects

The radiation environment inside the ATLAS detector originates from the energy emission by subatomic particles arising from the collisions. While radiation is the quantity the detector is constructed to register, the extreme radiation doses it will be subject to after the high-luminosity upgrade will also cause damage to sensors and electronics. Such damage is especially severe in the innermost detector layers, making the ITk the primary target for radiation-induced defects. Identifying and characterizing the radiation effects responsible for deteriorating the silicon device properties is of utter importance in order to be able to understand, simulate, and mitigate radiation damage effects.

As a particle strikes the detector it can transfer energy both by ionization and by atomic displacement. The damage caused by these types of interactions is characterized by two\(^1\) main quantities giving macroscopic cumulative effects in the detector from the total dose received:

- **Surface damage due to Total Ionizing Dose (TID)**, defined as the ionization energy deposited in a material divided by the mass of the material. This type of radiation effect does not cause any structural damage but highly impacts the transistor functionality and is thus mostly significant for silicon integrated circuit elements. The absorbed radiation dose is commonly measured in units of rad, where 1 rad = 0.01 J/kg.

- **Bulk damage due to Non-Ionizing Energy Loss (NIEL)**, a quantity that describes the rate of energy loss due to lattice displacements in the silicon as a particle transfers kinetic energy to nuclei while traversing the detector medium. In this case, ionization occurs as a secondary process caused by resulting lattice vibrations and constitutes only a small fraction of the total energy transfer. This type of damage mainly adheres to the charge collection process, and is therefore

\(^1\)A third quantity responsible for transient effects as a consequence of high energy (\(> 20\) MeV) hadrons releasing enough energy to produce a stochastically occurring malfunction in electronic components, a so-called Single Event Effect (SEE), is also used but is not covered here.
mostly significant for the sensors elements. In order to compare damage caused by different types of particles with different energies, NIEL is commonly measured in units of 1 MeV equivalent neutron fluence, or $n_{eq}$/cm$^2$, i.e. number of 1 MeV neutrons per unit area that would cause the same amount of displacement damage in silicon as the actual spectrum of mixed particles contained in the received radiation dose.

In order to account for the effects following the extreme dose levels the ITk will be subject to, radiation background simulations have been performed using the *PYTHIA8* [25] event generator and the *FLUKA* [26] particle transport code. The simulated 1 MeV neutron equivalent fluence per 4000 fb$^{-1}$ of integrated luminosity in the ITk can be seen in Figure 4.1, where the innermost layer (lowest $r$) will be exposed to the highest dose. Assuming a replacement halfway through the expected period of operation of the HL-LHC, i.e. after a maximum integrated luminosity of 2000 fb$^{-1}$, the highest fluence this inner layer of the ITk will receive has been computed to $1.3 \times 10^{16} n_{eq}$/cm$^2$ [27]. During the same period the ionizing dose will be 720 Mrad. To give an idea about the magnitude of this number, it is equivalent to the average radiation absorbed in 72 billion chest X-ray scans [28].

![Figure 4.1: Results from simulations of the total non-ionizing energy loss in the ATLAS ITk for proton-proton collisions at 14 TeV centre-of-mass energy [29].](image)

### 4.1.2 Radiation Damage in Pixel Sensors

Non-ionizing energy losses will cause displacement damage in the silicon bulk. This occurs when impinging particles impart an energy higher than the displacement threshold energy ($> 15$ eV [30]), causing a single silicon atom to be knocked out from its crystal lattice site, which results in the creation of a vacancy (vacant lattice site which would be occupied in a perfect crystal) and the atom becomes an interstitial (atom occupying site in crystal structure at which there is usually not an atom). This so-called *Frenkel defect* is illustrated in Figure 4.2 (a). If the imparted energy on the displaced silicon atom is high enough, a cascade effect may be triggered where several silicon atoms continue to be knocked out of their lattice sites, thereby generating further defects and causing larger disordered regions to form in the lattice. Most of the interstitial-vacancy pairs separated by short distances (smaller than the lattice constant) recombine, resulting in no net damage to the crystal, but the remaining defects
4.1. Radiation Damage in Silicon Devices

will diffuse through the crystal. As this happens, the defects will react with other defects or impurity atoms and thereby cause damage to the silicon bulk, mainly by introducing additional energy levels in the silicon band gap. Such electrically active defects will affect the device performance in various ways depending on concentration, energy level and cross-section. [31]

Three main NIEL effects can be identified on the device performance level [32]:

i) Increased leakage current. Deep energy levels, i.e. energy states close to the middle of the bandgap, act as efficient charge carrier generation (or recombination) centers that lead to an increased leakage current. This translates to increased power consumption and increased amplifier noise, which critically affects the performance of the analog part of the front-end pixel chip. The leakage current increases exponentially with temperature, making cooling an effective method to combat the detrimental effects.

ii) Change of effective doping concentration. Radiation-induced defects that act as donors or acceptors can change the effective doping concentration by introducing excess donors-like defects in the upper half of the band gap or excess acceptors-like defects in the lower half. This deforms the electric field distribution within the device and shifts the required depletion voltage so that higher operation voltages have to be applied to establish an electric field throughout the full sensor in order to avoid loss of active volume and thereby signal. The modified effective doping concentration risks further complication by substrate type inversion, which occurs when the n-type bulk material becomes p-type as the particle fluence increases. This is an effect of negative space charge forming, which compensates the initial positive space charge of the n-type material, and eventually the net space charge decreases to a point where it changes sign from positive to negative. The applied voltage required for depletion will thus rise accordingly, eventually going beyond the detector breakdown voltage and forcing operation in underdepletion. As the initial space charge in p-type sensors is already negative before irradiation, they are not affected by type inversion.

iii) Increased carrier trapping. Defect energy levels can capture (trap) mobile charge carriers and while thermal excitation can successively release (de-trap) the trapped carriers, this effect increases the charge collection time. If the de-trapping time is longer than the timescale of the measurement or if the concentration of defects (trapping centers) is very high, it will result in incomplete charge collection and the overall sensor signal will be reduced. The probability for the trapping centers to form increases with the amount of high-energy radiation received by the sensors, and since the effect of trapping affects the charge collection efficiency of silicon devices the formation of trapping centers constitutes the ultimate limiting factor for detector design, operation and lifetime in high-fluence applications.

The three radiation-induced effects described above are illustrated in Figure 4.2 (b).

The charge carrier motion is directed by the electric field, which can be increased to speed up the carrier drift velocities and thereby reduce the risk of trapping. Nevertheless, holes will always move slower than free electrons as their mobility is restricted due to hole transport involving sequential transition probabilities depending on vacant
states as results of electrons being excited to higher energy states. In this way, holes are constrained to the valence band and subjected to the stronger atomic force pulled by the nucleus than the electrons in higher energy states, and thus holes have a lower mobility. As a consequence of their higher mobility making the charge collection faster and thus the trapping probability lower, electrons will make a greater contribution to the signal than holes. In order to improve radiation hardness, it can therefore be advantageous to collect electrons instead of holes at the sensing electrodes. For this reason, the pixel sensors to be used in the ITk have their readout electrodes connected to the n-type implant, which is collecting electrons.

Despite the fact that radiation defects decay over time through the process of annealing\(^2\), the overall net effect of radiation exposure will progressively cause damage to the sensors, resulting in reduction of the charge collected and thereby signal loss and decreased SNR. While this effect can be mitigated by increasing the bias voltage, such an adjustment will also further increase the leakage current. Reduction of trapping effects are possible through careful device engineering and design choices. Consequently, when designing a system like the ITk, sensor technology choices for each detector region have been made such that stable thermal operation and high charge collection efficiency can be guaranteed over the active lifetime of the sensor. The Pixel Detector will use a combination of sensor technologies, where high-resistivity planar and so-called 3D silicon sensors are included. [21]

\(^2\)Since both the interstitial and vacancy are mobile depending on temperature, new defects could form or existing defects dissociate due to their thermal motion in the silicon lattice. This is referred to as annealing and depends on the thermal history of the sensor. Annealing can be used as a heat treatment using thermal energy to re-arrange the atomic defects, which can structurally recover damaged sensors. More information about annealing can be found in Paper I.
4.1 Radiation Damage in Silicon Devices

A more detailed description of how radiation damage affects the performance of silicon sensors can be found in Paper I.

4.1.3 3D Silicon Sensor Technologies

3D silicon sensors were developed in the 1990’s [33] as a more radiation-hard alternative to traditional planar sensors. The name originates from the fact that the charge collecting electrodes are columnar implants etched to penetrate perpendicular into the sensor bulk rather than being implanted on the wafer surface such as in standard sensors of planar geometry. With electrodes oriented in this way, the charges will drift in a two-dimensional plane ($x$-$y$) around the columns as opposed to a one-dimensional line ($z$) towards the planar electrode, as illustrated in Figure 4.3.

![Diagram of 3D and planar sensors](image)

Figure 4.3: Side view illustration of sensors technologies used in the ITk. $n^+$ denotes the n-type electrodes and $p^+$ denotes the p-type electrodes, while $n^-$ and $p^-$ denotes the lightly doped n-type and p-type silicon bulk respectively.

What is then the reason 3D sensors are intrinsically able to withstand radiation induced bulk damage better than other available sensor technologies, such as their planar counterpart? In fact, conventional planar sensors have demonstrated sufficient radiation hardness after their thickness has been reduced [16], but in addition to radiation tolerance requirements the choice of sensor thickness is based on minimization of the material budget as well as multiple scattering in the detector weighted against the cost implications due to the complexity of fabricating thinner sensors. The primary reason 3D sensors are interesting candidates is simply a better ratio of ionization thickness to charge collection distance than for planar sensors. Due to the inter-electrode distance between 3D columns being decoupled from the sensitive substrate thickness, it is possible to position the electrodes significantly closer together than in a planar sensor, where the electrodes are confined to the detector surface and a certain thickness is required for sufficient SNR as a thicker sensor generates more charge when a particle is traversing. The decreased inter-electrode distances result in faster charge collection times, lowering the risk of carriers being trapped from radiation induced defects as they drift towards the electrodes. More specifically, collection distances and collection times are about one order of magnitude lower than for sensors of planar technology and in addition, another advantage of the 3D technology is that the operational voltage of the sensors is about two orders of magnitude lower compared to what would be required to generate an equivalent drift field in planar sensors [33].
On the other hand, the main disadvantage of the 3D technology has historically been a limited production yield. However, lately the technology has matured as fabrication procedures have undergone several process improvements, making 3D sensors increasingly more affordable and a standard choice for applications requiring extreme radiation hardness. For this reason, along with their low power dissipation, 3D sensors are chosen to be the baseline option for the innermost layer of the ITk Pixel Detector (Layer 0 and Ring 0), as well as a possible alternative for Layer 1, while for the outer layers traditional planar sensors are considered due to a higher fabrication yield and cost efficiency.

Meanwhile, this is not the first time 3D sensors are used within the ATLAS experiment. In fact, both the ATLAS Forward Proton detector (AFP) [34] and part of the innermost barrel layer in the current ATLAS Pixel Detector, the insertable B-Layer (IBL) [35], use 3D pixel sensors [21]. This first generation of 230 µm thick 3D sensors, with a pixel size of 50 × 250 µm and an inter-electrode distance of 62 µm (see Figure 4.4 (a)), already demonstrated a radiation tolerance up to $9 \times 10^{15} \text{n}_{\text{eq}}/\text{cm}^2$ [36], superseding the required tolerance for both AFP and IBL. Apart from even higher radiation hardness, the ITk Pixel Detector also requires 3D sensors with smaller pixel sizes and thinner active areas in order to achieve a higher intrinsic position resolution and reduce detector occupancy in the innermost layer. Low pixel occupancies are desired in order to reduce pattern recognition mistakes due to incorrect hit associations which can be caused by in-pixel pile-up, i.e. multiple simultaneous hits within the same pixel. Therefore, 3D devices of 130 µm thickness of the two alternative pixel aspect ratios 50 × 50 µm$^2$ and 25 × 100 µm$^2$, compatible with the newly developed read-out ASIC, have been produced (see Figure 4.4 (b) and (c)).

```
Figure 4.4: Top view illustration of 3D pixel sensors versions compatible with the IBL geometry (a), and the ITk geometry of square pitch (b) and rectangular pitch (c). n$^+$ electrode columns are marked in blue and p$^+$ electrode columns are marked in red.
```

The ultimate spatial dimensions for hybrid pixel detectors are limited by the fact that the pixel area of the FE chip must match the pixel area of the sensor. Thus, the minimal pixel size is limited by the amount of electronics needed to amplify, discriminate, and process the hit information in the area occupied by the pixel cell, which is determined by the number of transistors per unit area in the CMOS process. In addition, since the signal of a traversing particle diffuses 4-8 µm for a typical sensor thickness of 150-200 µm [16], this sets a lower limit on the pixel pitch as too small pixel sizes would cause excessive charge sharing.
4.1. Radiation Damage in Silicon Devices

\[ F = q(E + v \times B) \]

\[ q(\mathbf{v} \times \mathbf{B}) \]

\[ \mathbf{v} \times \mathbf{B} \]

Signal size

Particle

\( \mathbf{B} = 0 \)

\( \mathbf{B} > 0 \)

Planar sensor

3D sensor

Figure 4.5: (a) Illustration of the Lorentz force \( \mathbf{F} \) acting on a point charge \( q \) moving with velocity \( \mathbf{v} \) and subject to an electric field \( \mathbf{E} \) and a magnetic field \( \mathbf{B} \). (b) Sensor orientation in the solenoid field of the ATLAS detector (not to scale). (c) Illustration of carrier motion (blue arrows) in planar and 3D silicon sensors with and without an applied magnetic field, demonstrating that the effect of the magnetic field is negligible on charge carrier collection in 3D sensors as the Lorentz angle is very small, irrespective of the incident angle of the impinging particles. This illustration is inspired by [37].

Apart from the difference in charge collection time, and thereby radiation hardness, another main difference between planar and 3D sensors is their signal response.
to magnetic fields. The solenoid magnet surrounding the Pixel Detector produces an axial magnetic field of 2 T at the centre of the ATLAS tracking volume, orthogonal to the electric field inside the sensor bulk of planar modules. The combination of these fields will cause an electromagnetic force, the Lorentz force to act on the charge carriers differently depending on the incident angle of the particle, see Figure 4.5 (a). This is causing them to deviate from the path they would have moved along had they not been subject to a magnetic field, as illustrated in Figure 4.5 (c). Meanwhile, carrier motion has low sensitivity to the Lorentz force in 3D sensors, as a result from the electric and magnetic fields being co-planar, as shown in Figure 4.5 (b). This has been confirmed by measurements in [38] and therefore only minor effects from the magnetic field on the signal size are expected.

4.1.4 Radiation Hardening of Readout Components

Radiation-induced damage also impacts the pixel readout chip which, like the sensor, is made of silicon. However, excessive damage can be prevented and the lifetime of the chip extended by careful design of the circuit at the transistor level. Conventional electronics design tools offer several different models to simulate the same transistor, i.e. in addition to a typical model for simulating nominal operation, so-called corner models are also provided. The use of corner models allows simulating operation with different temperatures, voltages, and/or variation of fabrication process parameters within some defined range in order to facilitate design validation prior to device fabrication and ensure production of circuits functioning under all possible extreme conditions. For ATLAS integrated circuit designs, such conditions also include extensive radiation exposure since this could alter the transistor behavior; for instance, the speed of a transistor may vary with received radiation dose.

<table>
<thead>
<tr>
<th>Device type</th>
<th>Minimum length (L)</th>
<th>Minimum width (W)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NMOS</td>
<td>-</td>
<td>120 nm</td>
</tr>
<tr>
<td>PMOS</td>
<td>300 nm</td>
<td>120 nm</td>
</tr>
</tbody>
</table>

Table 4.1: Recommended gate dimensions for transistors fabricated in 65 nm CMOS process to ensure 500 Mrad radiation hardness. The minimum gate length for NMOS devices is not constrained by radiation tolerance.[21]

When upgrading the Pixel Detector, the 130 nm CMOS technology node used in the current system is being scaled down to 65 nm CMOS in order to benefit from the increased logic density offered by smaller feature sizes. Therefore, the effects of radiation exposure to silicon devices fabricated in 65 nm technology have been studied by several CERN collaborations, in terms of understanding and validating the radiation damage mechanisms and how these impact transistor performance[22][39]. From these studies, additional corner models representing radiation damaged devices have been implemented and customized by parameterizing measurement data. These simulation models have been developed to emulate the behavior of the transistor after a received dose of 200 Mrad and 500 Mrad respectively, and have been verified by measuring the performance degradation during experimental irradiation tests. Further, the transistors must be sized to provide sufficiently large gate dimensions to ensure
radiation hardness. Based on this reasoning, design guidelines for transistor sizing using the 65 nm technology, as stated in Table 4.1, have been developed by dedicated CERN collaborations. These guidelines have been used to design analog circuits, which have been simulated with the 500 Mrad model, prototyped and successfully radiation tested in silicon to doses above 500 Mrad. [21]

4.2 Pixel Sensor Radiation Damage Simulation

4.2.1 Modeling of Radiation Effects

A simulation model for bulk radiation damage effects in the pixel sensors is in the process of being developed and incorporated into the simulation framework of the ATLAS detector, where previously no such effects have been accounted for. The simulation of the ATLAS detector operation is performed using a dedicated offline software chain where the production of simulated samples follows certain steps:

1. **Event generation using Monte Carlo methods** to simulate the particle collisions.

2. **Detector simulation using Geant4** to simulate particle interactions with the detector material.

3. **Digitization of simulated energy deposits into detector read-out data format** to translate the detector response into digital form.

4. **Event reconstruction based on digitized data** to study what took place during the collisions (identical treatment of simulation and collision data).

The radiation damage model is being implemented as part of the digitization step of the software chain (number 3 above), where the detector response is translated into a digital format. The digitizer model is fed particle hit data from Monte Carlo simulations (1) that have been converted into a detector response in terms of \((x, y, z)\) location coordinates and magnitude of energy deposited into the material (2). The functionality of the digitizer part of the simulation is to emulate the collection process of charge carriers released through these energy deposits by propagating the carriers towards the sensor electrodes and translate the collected charge into a digital ToT value. As charge carrier trapping will cause signal loss in this part of the readout process it is of crucial importance to include radiation damage effects here in order to enable accurate track reconstruction, especially as the luminosity is increased.

Monte Carlo data is obtained from simulations where the sensor is placed inside a geometry mimicking the EUDET-type beam telescope setup [40], which was established to provide test-beam infrastructure for research and development within HEP. An illustration of this setup is shown in Figure 4.6.
The digitizer effectively models the motion of charge carriers, i.e. electron-hole pairs, released through ionization of the silicon. The propagation of charge carriers will consist of two types of simultaneous\(^3\) motion: drift and diffusion. Figure 4.7 illustrates stochastic thermal diffusion acting on a collection of charge carriers and causing a spatial spread among them, while experiencing drift, which is deterministically dictated by the electric field within the sensor.

\[ x = 0 \]

\[ x = 5 \]

\[ x = 10 \]

\[ x = 15 \]

\[ x = 20 \]

Since the charge collection process differs depending on the sensor geometry, separate models for the planar sensors and 3D sensors have been developed. This study focuses on the 3D digitizer model only.

### 4.2.2 3D Sensor Digitizer

#### TCAD Simulations

The amount of charge induced on the electrodes as a consequence of the carrier motion in the electric field can be computed using a geometry-dependent weighting potential called the *Ramo potential* (for a more detailed description of the Ramo potential refer to Paper I). Therefore, both the Ramo potential and the electric field within the sensor

\[^3\text{Even though drift and diffusion are simultaneous processes they are modeled sequentially as investigations have shown that the effects on the accuracy of modeling the carrier motion in this way are negligible.}\]
after exposure to a certain level of irradiation are important inputs to the digitizer simulation. Since these quantities are deterministic and fixed for a given detector geometry and set of conditions (temperature, bias voltage, fluence), they are computed prior to runtime in device simulations using Technology Computer Aided Design (TCAD) tools, and stored as lookup tables that map each position within the sensor to a corresponding value. These maps are used by the digitizer to model the propagation of charge carriers within the sensor in order to predict the total charge induced at various levels of radiation exposure.

Figure 4.8: Simulated electric field in 3D pixel sensors and corresponding charge collection time map after irradiation at a temperature of -15°C, displayed on a cross-section of the x-y plane in the middle of the sensor with the columnar electrodes penetrating inwards in the z-direction. The n⁺ and p⁺ implants are regions of zero field due to their high doping concentration. (a) 230 µm thick n⁺-in-p modules of IBL-compatible sensor geometry (250×50 µm²) after a received fluence of \(10^{15}\) n_{eq}/cm² (maximum expected). (b) 130 µm thick n⁺-in-p modules of ITk-compatible sensor geometry (50×50 µm²) after a received fluence of \(10^{16}\) n_{eq}/cm² (maximum expected).

The TCAD simulations of the electric field profile are performed using the so-called Perugia model (see [41], [42]) which models the displacement of silicon atoms from their lattice sites for a given received fluence by effective energy levels in the
silicon band gap. The spatial distribution of space-charge density becomes more non-uniform with increased fluence, thus causing non-uniformity also of the electric field. This effect remains even when the bias voltage is adjusted to compensate for the radiation-induced damage. TCAD data has been extracted for pixel dimensions compatible with the IBL geometry to be used as input to the digitizer model when performing Run 2 simulations of the existing inner detector system, as well as for pixel dimensions compatible with the baseline ITk geometry for simulations of operation following the HL-LHC upgrade.

In order to optimize the speed and efficiency of the digitizer, as many deterministic position dependent quantities as possible are stored in lookup tables and retrieved as inputs during runtime for a given \((x, y)\) coordinate within the pixel. Apart from the Ramo potential and electric field being such quantities, the *charge collection time*, i.e. the time until the arrival at the electrode of a charge carrier released inside the silicon, is another important input parameter. The charge collection time depends on the electric field as well as the sensor thickness, but is typically on the order of a few nanoseconds, and can easily be computed from the electric field data and stored as a separate lookup table, mapping each position within the pixel to a corresponding collection time. The data stored in these maps is important as it is used to determine whether a carrier is trapped or not, as well as to determine the location where each carrier gets trapped in order to compute the charge induced on the electrodes as a consequence of the carrier motion in the electric field. Figure 4.8 shows the position dependent TCAD maps with a spatial resolution of 1 µm for the electric fields and charge collection times after the expected maximum received fluence for the IBL and the innermost layer of the ITk respectively. Visually, these maps illustrate a cross-section of the pixel sensor, with the IBL-compatible geometry having two n\(^{+}\) electrodes for electron collection and the ITk-compatible geometry having only one collecting electrode per pixel.

---

Figure 4.9: (a) Ramo potential displayed over cross-sectional view of \(x\)-\(y\) sensor plane covering 3x3 (full) neighboring pixels with pixel boundaries marked with dark grey solid lines. The number overlaid in the center of each pixel states the maximum normalized charge that can be induced on the primary pixel from which the electrons and holes originate (middle pixel marked with blue dashed lines) by a carrier moving in the given pixel. Contributions from pixels outside of the yellow dashed square are negligibly small and therefore not accounted for in the computation of the total amount on induced charge from neighbors of the primary pixel. (b) Corresponding three-dimensional view of the Ramo potential in a 3D pixel sensor matrix.
Because the current pulse induced at the electrode as a consequence of the motion of carriers consists of the sum of induced pulses from each carrier’s path towards (electrons) and away from (holes) the collecting electrodes (n+), the amount of induced charge must be determined regardless of whether a carrier physically reaches the collecting electrode or gets trapped along the way. In order to compute the induced charge, the individual carrier contributions are weighted by the Ramo potential, resulting in carrier paths close to the pixel implant having the strongest contributions to the signal. This is because of the nature of this weighting potential, as seen in Figure 4.9, which shows the Ramo potential over a 3D pixel sensor matrix. Consequently, the closer a carrier travels to the collecting electrode the more charge is induced as a consequence, which should make it evident why radiation induced charge carrier trapping reduces the recorded signal. Moreover, even though the potential decays quickly with distance to the electrode, the motion of each carrier will result in charge induced not only on the closest but additionally, albeit significantly smaller in magnitude, on adjacent electrodes. This effect is shown in Figure 4.9 (a).

Charge Collection Model

The 3D digitizer simulates the signal formation within the 3D sensors in the presence of radiation damage effects using a charge collection model that receives information about a series of “hits” from particles traversing the sensor and translates this to a digitized response in the form of a ToT value representing the amount of induced charge collected at the pixel electrodes. The charge induced by a carrier moving in the pixel sensor can be determined from the Ramo potential of the 3D sensor geometry (as displayed in Figure 4.9), given the initial starting position \((x_i, y_i)\) and the final position \((x_f, y_f)\), where it either gets trapped or reaches the electrode. The input to the digitizer is a hits collection produced in Monte Carlo simulations, consisting of energy deposited, \(E_{\text{dep}}\), at position \((x_i, y_i)\), and from this the model computes the final position of the carrier \((x_f, y_f)\) and the charge induced \(Q\). The digitizer receives information about detector geometry (pixel size, type, tilt angle), bias voltage, operating temperature and fluence upon initialization. Modeling of charge collection as energy is deposited in the detector is then structured as shown by the broken down flowchart that follows:

1. **Charge carrier grouping.** Each energy deposit, \(E_{\text{dep}}\), is converted to a number of charge "chunks", where each chunk consists of several electrons/holes grouped together. Propagating groups of fundamental charges through the digitizer simultaneously instead of as individual carriers allows faster digitizer performance to be achieved without loss of accuracy, provided the number of carriers per chunk is sufficiently small \((\sim O(10))\) so that overestimations of fluctuations are negligible.

\(^4\)Because the movement of charge carriers in \(z\) is negligibly small due to the essentially uniform electric field across the 3D pixel sensor depth, it is not accounted for in the 3D digitizer model as it does not significantly affect the amount of induced charge.
2. Thermal diffusion. The effect of thermal diffusion is modeled by spreading the chunks through random one-step displacements in the \(x\)-\(y\) plane, such that the new position of each carrier will be \((x_i + \Delta x, y_i + \Delta y)\). \(\Delta x\) and \(\Delta y\) are drawn from a Gaussian distribution where the standard deviation \(\sigma(E, t_{\text{drift}})\) depends on the carrier mobility \(\mu\), which in turn is dependent on the electric field \(E\) and thereby also on fluence as the electric field is affected by the received radiation dose. The standard deviation also depends on the time the carrier spends drifting \(t_{\text{drift}}\), either until getting trapped or until reaching the electrode, which also has a statistical dependence on fluence.

3. Drift and trapping. The so-called trapping time, i.e. the time a chunk spends drifting before getting trapped, is drawn from a probability distribution that scales with received fluence \(\Phi\). This time is then compared to the charge collection time for the same carrier based on its starting position, which is computed using the TCAD simulation data stored as lookup tables, as seen in Figure 4.8. If the charge collection time is greater than the trapping time, the carrier is considered to be trapped at the location to where it has drifted during the trapping time.

4. Induced charge computation. The charge induced, \(Q\), as a consequence of each chunk moving from the initial hit position to the final position is computed using the Ramo potential \(\phi\) according to \(Q = \phi(x_f, y_f) - \phi(x_i, y_i)\). Charge contributions from all chunks are summed and converted to a digitized ToT value. Figure 4.10 shows the amount of charge induced by carriers being trapped at
different locations within the sensor when drifting in the electric field while simultaneously experiencing diffusion.

A more detailed description of the 3D digitizer structure can be found in Paper II.

4.2.3 Model Performance

The most important quantity for assessing the impact of radiation damage on tracking performance is the charge collection efficiency (CCE). The CCE expresses the fraction of charge initially deposited by particles interacting with the sensor material to the charge collected by the electrodes and registered by the readout system. Thus, the CCE can be used to evaluate the impact of radiation damage on signal loss. Decreased CCE along with increased electronic noise due to larger leakage currents results in a worse SNR in irradiated detectors, which can lead to the inability of detectors to discriminate the signal from noise.

Due to the statistical nature of the ionization process occurring as particles traverse matter, the energy deposited in a single detector cluster exhibit large fluctuations. The distribution probability for an incident particle to lose a certain amount of energy when passing through a sensitive detector layer of a certain thickness will resemble a Landau distribution. Therefore, the average charge collected per event can be computed at a given fluence by performing a Landau-Gaussian fit to the ToT distribution produced by the 3D digitizer, where a Landau distribution is convolved with a Gaussian distribution to also account for experimental errors and signal contributions from electronic noise. The fit is used to find the peak and width of the distribution, in cases where the most probable value (MPV) is a a more reliable measure than the mean of the signal.
distribution. Further, by normalizing this MPV to the MPV obtained by an identical fit to a ToT distribution of a non-irradiated sensor, the CCE can be calculated. ToT distributions from 3D digitizer data are displayed in Figure 4.11 for a non-irradiated and irradiated sensor respectively. The simulations were performed with the sensors tilted by 15° in φ (cf. Figure 4.6) in order to increase charge sharing for better tracking resolution as the track position can be more precisely determined, as well as to optimize tracking efficiency. Because of the way the electrodes are positioned, reducing the area where particle hits can be registered, 3D devices exhibit some loss of hit efficiency for normal incident tracks but recover full efficiency with track tilt angles > 10° [37]. Results of CCE computations using the 3D digitizer model for various fluences up to \(10^{16}\) n_{eq}/cm² can be seen in Figure 4.12, which displays a decaying trend in CCE with fluence, as well as the importance of increased bias voltage to improve CCE.

![Figure 4.11: ToT distributions for non-irradiated sensor (pink) and sensor irradiated to \(10^{16}\) n_{eq}/cm² (green) from all hits in individual 50x50 \(\mu\text{m}²\) pixels positioned with 15° tilt with respect to the beamline.](image)

![Figure 4.12: Charge collection efficiency vs. fluence at different levels of bias voltage for pixels positioned with 15° tilt with respect to the beamline.](image)
The digitization model for 3D sensors presented here is written in C++ using the ROOT [43] libraries and has been developed on a stand-alone platform called Allpix [44], which is a wrapper around Geant4 [45], a toolkit for simulations of particle-matter interactions. The model has thereafter undergone validation and migration to the Athena framework [46] (a production workflow software created by the ATLAS collaboration to manage event generation, simulation and reconstruction), to be incorporated into the ATLAS offline software chain. The current focus of the 3D digitizer model development is to compare simulations to test beam data, and the observed trend in CCE with respect to fluence for the IBL geometry is close to what has been found in previous studies [47]. Simulation results are also being compared to Run 2 collision data for the IBL geometry, as well as to test beam data for the ITk geometry to evaluate the performance of the 50×50 µm² and the 25×100 µm² pixel pitches to help inform the decision about which one to use.
Chapter 5

ITk Challenge 2: Data Readout Speed

5.1 ATLAS ITk Readout System Architecture

5.1.1 Readout Bandwidth

The unprecedented data volumes that will be produced at the HL-LHC makes the design of the data acquisition and readout system for the ITk Pixel Detector considerably more challenging than for the existing ATLAS inner detector system and imposes new demands on the data extraction process. Because of the increased pile-up and high-rate operation, the raw data volume produced by the detector will be too high for read-all operation to be possible, especially while keeping the mass of the data transmission cables sufficiently low to fit into the tracker volume and minimize material interference with the measurements. Further, reading out all events is not efficient, since the particle production during the collisions is a statistical process where on average only 0.1\% of the total number of proton-proton interactions contains characteristics that might lead to new discoveries and are interesting enough to be kept for data analysis. In order to reduce the event rate to a manageable level for data recording and offline analysis, a trigger system can ensure that only collision data containing useful information is filtered out for permanent storage according to given criteria that correspond to signatures for different physics processes of interest. In such a triggered system, the readout chip will store hits temporarily for a pre-defined latency period, typically on the order of a few µs, until a trigger decision determines which samples to pass on to the subsequent readout stage, while all other hits are discarded. The ATLAS common read-out and control architecture embeds the read-out and control of the ITk Pixel Detector, where the baseline read-out scheme features a single hardware trigger level for complete readout of the Pixel Detector at a foreseen latency of 10 µs and trigger rate of 1 MHz [48]. The system is developed to allow for an alternative hierarchical trigger mode featuring two hardware trigger levels, where the trigger rate in the outer layers would be 4 MHz in order to trigger on possible new physics during future tracker operation, see [27] for details.

The trigger system must operate at a rate matching the incoming particle flux and pile-up, and the readout bandwidth $B$ must then be scaled accordingly. The required bandwidth depends on the number of bits to be transferred per trigger $N_{\text{trig}} = \mu b$, where $\mu$ denotes the number of interaction per bunch crossing (pile-up factor) and $b$ the
number of bits required for digital storage of each event (containing information about position, arrival time and collected charge of each hit). The bandwidth also scales with trigger rate \( R_{\text{trig}} = \mu pr \), where \( p \) denotes the probability of an interesting proton-proton interaction taking place during a given bunch crossing and \( r \) the collision rate (40 MHz at the LHC). Thus, the required readout bandwidth can be expressed as

\[
B = N_{\text{trig}} R_{\text{trig}} = (\mu b) \cdot (\mu pr) \sim \mu^2
\]  

(5.1)

The quantities of Equation 5.1 are clarified in Figure 5.1

From Equation 5.1 it is evident that the readout bandwidth \( B \) scales quadratically with the pile-up factor \( \mu \). This means that even though the average number of proton-proton interactions per bunch crossing will only be raised from 40 at the LHC [49] to 200 at the HL-LHC [17], this factor of 5 escalation in pile-up will correspond to a much greater increase in readout bandwidth to ensure sufficient readout speed without compromising data quality.

![Figure 5.1: Illustration of quantities needed to estimate the required readout bandwidth \( B \), demonstrating how collision events occurring with frequency \( r \) with a pile-up factor \( \mu \) and probability of interesting interactions taking place \( p \) sets the trigger rate \( R_{\text{trig}} \), and how the digital storage size of each event \( b \) together with pileup determines the number of required bits to read out per trigger \( N_{\text{trig}} \).](image)

The required number of bits per trigger, \( N_{\text{trig}} \), can be predicted from the estimated pixel chip hit occupancies. Average hit rates per chip per event are derived from Monte Carlo simulation studies of physics events in the Pixel Detector layout using Geant4, which are then passed through the ATLAS reconstruction (for further details see [21]). Hit occupancy is defined as the number of simultaneously hit pixels during a given collision event, i.e. number of pixels for which the amplified signal exceeds the discriminator threshold. Results from simulations of the hit occupancy per chip in the various layers of the ITk Pixel Detector can be seen in Figure 5.2.

Because of the high hit occupancies imposing requirements on the readout bandwidth, one major advancement during the Phase-II Upgrade of the ITk Pixel Detector is the replacement of previous generation data links operating at 160 Mbps with a new link system to provide multi-Gbps readout, in order to accommodate a data rate increase by a factor close to 40 [16]¹. As the density of particles traversing the pixel

¹The use of a 64b/66b based encoding scheme in the new system instead of 8b/10b in the present adds approximately 23% of additional usable data bandwidth on top of the increased transmission frequency due to the decrease in relative overhead, making the effective data rate increase by a factor of \((5.12 \times 64/66)/(160 \times 8/10)) \approx 39\).
5.1. ATLAS ITk Readout System Architecture

4.3 Pixel Module

The advantage in data compression becomes higher as the data rate increases. The encoding scheme used is 66b encoding, which will be applied as the final step before transmission. 64b is a digital data representation that transforms 64-bit payload data to 66-bit data sequences by adding 2 coding bits so that enough state changes are provided to allow clock recovery and data stream alignment at the receiver. This encoding scheme also provides DC-balance, i.e. assures that roughly equal amounts of high and low bits are being transmitted in order to avoid a DC offset.

modules is decreasing with radial distance to the interaction point, the average number of hits per front-end chip per event varies by pixel layer, as the computations underlying Figure 5.2 confirm. This results in a corresponding need for variation in output readout rate in the different layers. Based on the data format\(^2\) implemented in the front-end chip along with occupancy estimates, it is possible to determine the average link occupancy in the various layers of the Pixel Detector. Estimates of the required data rates for single chips can be found in Table 5.1 (for a clarification of the different Pixel Detector regions refer to Figure 3.2).

Table 5.1: Results from computations of required data rate per single chip in different layers of the ATLAS ITk Pixel Detector (also accounting for different trigger rates). To allow for event-to-event fluctuations in the data rate and leave sufficient overhead in relation to the simulation results it is required that the data rate does not exceed 70% of the estimated bandwidth available on the link.

<table>
<thead>
<tr>
<th>Layer</th>
<th>Trigger rate [MHz]</th>
<th>Data rate/chip [Gbps]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Flat barrel</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>2.53</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.43</td>
</tr>
<tr>
<td>2</td>
<td>1 (4)</td>
<td>0.30</td>
</tr>
<tr>
<td>3</td>
<td>1 (4 end-cap)</td>
<td>0.19</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>0.59</td>
</tr>
</tbody>
</table>

\(^2\)The pixel readout chip will use an open source commercial protocol implementation of the industry standard 64b/66b encoding, which will be applied as the final step before transmission. 64b/66b is a digital data representation that transforms 64-bit payload data to 66-bit data sequences by adding 2 coding bits so that enough state changes are provided to allow clock recovery and data stream alignment at the receiver. This encoding scheme also provides DC-balance, i.e. assures that roughly equal amounts of high and low bits are being transmitted in order to avoid a DC offset.

5.1.2 Transmission Link System

The solution of a programmable number of enabled data lanes per pixel readout chip (see Section 3.2.2) is implemented to reduce the number of data links in the system.

Figure 5.2: Simulated average number of hit pixels per chip per event in barrel layers and end-cap rings in the Phase-II ITk Pixel Detector baseline layout. [21]
since the inner layer chips will use three lanes\(^3\), while outer layer chips will use a single lane only. Apart from accommodating bandwidth demands, the concept of reducing the number of links also aligns with the essential requirement to minimize mass inside the tracker. This is especially important because the ATLAS experiment is designed to cover as large an area around the interaction point as possible and such hermetic coverage means most data cables must pass through the active detector volume. In order to further reduce the number of cables and the amount of inactive material in the tracker, a serial scheme will be adopted to supply power to the modules where the module reference voltage will vary along the serial power line. As a consequence of not having a common ground reference shared between the modules, the data transmission lines must be AC-coupled, and to reduce the risk of potential grounding issues it would be beneficial to use optical transmission as soon as possible along the data line. Optical transmission is in general often preferred as a more attractive solution over electrical transmission due to its low mass, low-loss transmission capability and therefore high bandwidth over long distances, as well as immunity to electromagnetic interference. However, optical transmission is not a viable option for the first part of the Pixel Detector readout chain due to the proximity to the beam interaction point, as optical elements are less radiation tolerant and no available optical solutions with sufficient radiation tolerance to be able to operate in environments exceeding 100 Mrad have been identified [16]. Another reason optical solutions are not suitable in the innermost detector volume is that fragile fibers with low bend radius do not offer the flexibility required to achieve hermetic coverage in the geometrically compact Pixel Detector.

![Diagram of initial baseline readout architecture for the ITk Pixel Detector.](image)

Figure 5.3: Initial baseline readout architecture for the ITk Pixel Detector.

An overview of the initial\(^4\) baseline ITk Pixel Detector readout architecture is

\(^3\)Three lanes are sufficient for the current baseline innermost radius of 39 mm, but a smaller radius is still under consideration where all four lanes would be used to reduce link occupancy.

\(^4\)An updated baseline design for the ITK Pixel Detector data transmission system was recently established, see Section 5.4.5 for details.
5.1. ATLAS ITk Readout System Architecture

depicted in Figure 5.3. Due to the above mentioned lack of radiation hardness and flexibility of optical solutions, the link system is split into an inner electrical and an outer optical part. This allows all optical components along with the opto-electrical conversion interface, referred to as the opto-board, to be situated in a less challenging radiation environment far from the interaction point, where existing radiation tolerant solutions can be deployed. Moreover, for maintenance reasons, it is preferred to place the opto-board in a region that can be accessed during a long LHC shutdown. The baseline opto-board technology is realized through the Versatile Link+ project [50], which is a CERN LHC-wide effort to develop a radiation-hard optical link system between the on-detector and off-detector electronics, and the output data driver of the pixel system will transfer data to the Versatile Transceiver (VTRx+) module. The subsequent main interface between the detector and all off-detector systems is the FELIX system [51]. The DAQ chain also consists of a trigger interface between the ITk systems and the ATLAS-global Timing Trigger and Control (TTC) system. The backend systems will be comprised by several PCs for Detector Control System (DCS), ITk monitoring, data handling and control.

The serial outputs from the pixel readout ICs will be coupled to an electrical data transmission line, connecting the front-end chip with the opto-board. This communication between on- and off-detector electronics is realized through a dedicated link system, separated into command (TTC) links to carry clock, trigger and command signals from off-detector electronics to on-detector modules, so-called downlinks, and data links to carry data from the on-detector modules to the off-detector electronics, so-called uplinks. As downlinks operate at significantly lower speed, solutions fulfilling requirements for data transfer are expected to qualify also as command links without issues. In what follows, focus will be on the more challenging implementation of the uplink system as initially planned, referred to as the active cable and described in Section 5.1.4. The active cable uplink system is highlighted in blue in Figure 5.3 above, and the main idea behind it is to minimize mass by transitioning as early as possible to high-speed links by multiplexing data from several chip output lanes instead of preserving the 1.28 Gbps parallel outputs. [21]

5.1.3 Data Cable Prototypes

Cables suitable to act as transmission media in the data link are scarcities, since commercial data cables are not developed with low-mass requirements and contain dielectric materials that are not radiation hard. For this reason, several low-mass, radiation hard, high-speed prototype cables have been developed and characterized to determine their suitability for the data link. One of the most important aspects when selecting a cable is the bandwidth it provides, since this highly impacts the signal quality. Any data sequence containing a maximum number of transitions between '0' and '1', translating to symbols constantly shifting between high and low in analog domain, corresponds to a frequency of half the data rate. Therefore, the minimum required analog bandwidth (without added safety margin) to transfer a digital signal of \( X \) bits per second, is \( X/2 \) Hz. Accordingly, the minimum required bandwidth for the pixel data link transmitting data at a rate of 5.12 Gbps is 2.56 GHz, as illustrated in Figure 5.4. With added safety margin, the targeted system bandwidth has become 3 GHz. A list of the cable prototypes and their corresponding loss at this maximum transmission
frequency can be found in Table 5.2, where various lengths are explored to be used in different detector regions.

<table>
<thead>
<tr>
<th>Cable type</th>
<th>Length [m]</th>
<th>Wire material</th>
<th>Shield</th>
<th>Dielectric</th>
<th>Insulation</th>
<th>dB loss at 3 GHz (per meter)</th>
</tr>
</thead>
<tbody>
<tr>
<td>28 AWG twinax</td>
<td>6</td>
<td>Cu</td>
<td>Al</td>
<td>PE</td>
<td>PES</td>
<td>14.0 (2.33)</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>Cu</td>
<td>Al</td>
<td>PE</td>
<td>PES</td>
<td>16.6 (2.08)</td>
</tr>
<tr>
<td>30 AWG twinax</td>
<td>6</td>
<td>Cu</td>
<td>Al</td>
<td>PE</td>
<td>PES</td>
<td>17.0 (2.83)</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>Cu</td>
<td>Al</td>
<td>PE</td>
<td>PES</td>
<td>19.0 (2.38)</td>
</tr>
<tr>
<td>34 AWG twinax</td>
<td>4</td>
<td>Cu</td>
<td>Al</td>
<td>PE</td>
<td>PES</td>
<td>20.0 (5.00)</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>Cu</td>
<td>Al</td>
<td>PE</td>
<td>PES</td>
<td>32.00 (4.00)</td>
</tr>
<tr>
<td>36 AWG TWP</td>
<td>1</td>
<td>Cu</td>
<td>PE</td>
<td>PEEK</td>
<td>Kapton</td>
<td>5.0 (5.0)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Cu</td>
<td>Al</td>
<td>PE</td>
<td>PES</td>
<td>12.0 (4.00)</td>
</tr>
</tbody>
</table>

Table 5.2: Cable prototypes for pixel data transmission link. The conductor diameter is measured in American wire gauge (AWG).\(^5\)

![Figure 5.4](image)

**Figure 5.4:** Illustration of minimum sampling rate (Nyquist frequency) to accurately sample analog signal. For the analog signal (blue) with \( f_{\text{Nyquist}} = \frac{1}{T} = 2.56 \text{ GHz} \) the corresponding digital signal (red) will be transmitted at a rate of \( N \cdot f_{\text{Nyquist}} = 5.12 \text{ Gbps} \), where \( N \) is the number of bits transferred during period \( T \) so that the sample rate is \( T/N \). In this case \( N = 2 \) and the bit period is \( T/2 \).

The use of longer transmission lines combined with minimized data cable mass means a high signal attenuation tolerance is required by the ATLAS data transmission system. Meanwhile, in order to allow sufficient margin for secure transmission, the maximum tolerated loss across the full data chain\(^6\) has been set to 20 dB at 3 GHz over 5 m, motivated by the fact that current state-of-the-art FPGA receivers can cope with a loss of 28 dB to guarantee reliable transmission when applying the best equalization and signal integrity techniques available [21]. For this to be possible, the number of interconnections along the line must also be limited. Assuming a loss of 1 dB per circuit board (module flex or boards in the opto-board) and 2 dB loss for each connector, the 20 dB maximum loss target translates into 14 dB for the cable alone over the desired length of the electrical links. While the fabricated cable prototypes include both flexible printed circuits and twisted-wire pairs (TWPs), the current solution being explored is a dedicated custom development twin-axial (twinax) cable, which is a

\(^5\)American wire gauge (AWG) is a logarithmic stepped gauge system for electrically conducting wire in which increasing gauge numbers denote decreasing wire diameters.

\(^6\)The total loss of the link is estimated by adding the losses in the full system: circuit boards (traces), connectors and cables/flexes.
micro-coaxial cable with two inner conductors to ensure equal laylenghts for matched signals. This prototype is considered the primary candidate due to its proven radiation hardness capabilities, as well as the fact that it fulfills the specifications of sufficiently low attenuation at 3 GHz.

5.1.4 Active Cable Concept

The uplink architecture should make optimal use of the available bandwidth in all detector layers while minimizing the material contribution from the data cables in the detector volume by keeping the number of links to a minimum. Therefore, the initial baseline design for the 5.12 Gbps electrical transmission solution to cover the distance between the pixel front-end module on one side and the opto-board on the other, was to combine the active components needed to multiplex the four front-end chip serial outputs with a passive low-mass cable into a custom active cable. This system builds on a commercial connector and cable technology in order to benefit from the reliability of proven industrial solutions; meanwhile, two custom application specific integrated circuits (ASICs) on the transmitting and receiving ends will be supplied by ATLAS to ensure sufficient radiation tolerance, and will be designed as part of a combined system taking into account the characteristics of the cable used.

![Diagram of Active Cable](image)

Figure 5.5: Active cable illustration.

Figure 5.5 illustrates the pixel active cable, which is a concept with several advantages. Firstly, it makes integration of the detector less complicated because the twinax cable to connector termination will be done to make the transmission line ideally coupled to both driver and receiver ends, while being completely decoupled from the rest of the system and the FE chip, thereby enabling independent development and qualification of the link. Secondly, the active cable approach offers significant system design flexibility as the cable does not have to be connected directly to the module, but can be placed at the end of the stave. Such a configuration enables data transmission along the staves at 1.28 Gbps, which can be achieved with a wide range of low-mass cables, like twisted pairs or flex cables. Further, since the twinax cable is integrated between two custom ICs without any intervening connectors as opposed to commercial ICs with a fixed impedance, the cable impedance can be chosen as needed and thus allowing for smaller diameter cables than would be the case for standard 100 Ω impedance.

The transmitter ASIC is referred to as the Aggregator chip and has four 1.28 Gbps inputs and one 5.12 Gbps output and is able to serve a single FE chip for the inner layer and up to four FE chips at higher radius. The receiver ASIC is called the Equalizer chip and needs to be designed to compensate for any distortion caused during
transmission in order to restore the original signal and enable clock and data recovery. The design of this ASIC is covered in Section 5.3, while the distortion it is compensating for and the technique used to achieve signal recovery are discussed in Section 5.2.

5.2 Signal Recovery with Equalization

5.2.1 Signal Distortion

In its simplest form, a general data transmission system typically consists of a transmitter transferring a serial signal over some distance through a transmission channel, for instance a cable, to a receiver. The most essential requirement on such a system is that the receiver measures the same pattern that was sent by the transmitter. Theoretically, each bit in a data sequence can be transmitted as a square pulse but practically these pulses will be rounded to some extent due to limitations in system bandwidth. In some cases, the pulse shapes of the bits will be so severely modified as they traverse the channel medium that it is impossible for clock/data to be extracted at the receiving end.

![Figure 5.6: (a) Illustration of general data transmission link suffering from ISI and resulting eye closure at cable output. (b) Illustration of distortion of bits contained in transmitted signal (blue) and fractions \( c_k \) of previously received bits \( r[n-k] \), where \( k = 1, 2, 3 \), remaining in current bit \( r[n] \) (green).](image)

In addition to being a result of random noise and crosstalk, the signal degradation arising during data transfer is mainly due to the spectral characteristics of the transmission medium being similar to those of a low-pass filter, which causes a non-uniform frequency response. This will result in elongation of the pulses, where the degree of distortion will depend on the frequency components of the signal, intensifying as the data rate increases. Such smearing of the signal can cause the magnitude of each bit to have some overlap with adjacent bits when the signal does not have time to reach full strength within its designated time frame, thus impacting the voltage levels of neighboring bits. As a result, the receiver may measure different bit voltage levels than those that were transmitted. That is, even though a bit is transmitted as a logical '0', signal degradation may cause the logic level of the same bit to approach close enough to the decision threshold as it reaches the receiver to be misidentified as a logical '1', thereby producing a bit error due to inaccurate decoding. This is called inter-symbol...
interference (ISI) and will affect both voltage and timing components of the signal. Figure 5.6 illustrates the problem of ISI in a generic data transmission system.

ISI constitutes a main issue in high-speed signaling systems. As high frequencies exhibit more loss than low frequencies, the shape of each logic transition in a data sequence will depend on the number of consecutive identical bits preceding the transition. A greater number of data transitions, and thus greater dominant frequency, means that the time between logic voltage swings decreases. This implies that the risk of ISI increases at higher data rates. For instance, a data sequence like '000111000', having a dominant frequency of only 1/3 of the data rate, may pass through the transmission line without being affected by ISI, while a sequence like '010101010' will have the same dominant frequency as the data rate and will thus be subject to a much greater risk of ISI. When plotting an eye diagram\textsuperscript{7} of a serial signal with high enough frequency components to make ISI is a significant issue, low-frequency trajectories of the waveform would be accurate, while high frequency components would be attenuated if the frequency response of the channel is such that it filters out the higher frequency content of the waveform. This will result in either the complete or partial closing of the eye diagram, as seen in Figure 5.6.

### 5.2.2 Equalization Techniques

By applying certain corrections to compensate for degrading effects of transmission lines, a distorted signal can be restored to its original form and the data rate at which the system can operate can be extended. When such corrections are applied at the transmitting end of the link it is referred to as pre-emphasis, which is a technique used to modify the transmitted waveform to send purposely distorted signals by generating a voltage swing that is higher for a bit following a transition than for a bit that does not, thus having the effect of counteracting the expected cable distortion that suppresses high frequency signals. When signal corrections are instead applied at the receiver it is referred to as equalization, a technique which will be more elaborately explained in what follows. The effect of equalization is shown in Figure 5.7.

![Equalization Effect](image)

Figure 5.7: Effect of equalization on signal recovery, showing how each bit in a distorted signal (purple) is corrected by ISI elimination to yield the equalized signal (yellow).

\textsuperscript{7}An eye diagram is the synchronized superposition of all possible realizations of a bit sequence viewed within a particular time interval, and is an efficient way to assess combined effects of noise, jitter and ISI present in the system.
While remaining a relatively unexplored area within the HEP community, many industry applications already have sophisticated performance optimizations to compensate for high signal loss. Currently, particle physics designs are also moving closer to commercial protocols and solutions, and equalization will be an essential part of these future systems. The receiver ASIC of the Pixel Detector data transmission link is referred to as the Equalizer chip because it includes equalization matched to the twinax cable prototype in order to ensure reliable transmission. In addition to equalization being deployed at the receiving end of the link, the chip drivers will also have pre-emphasis capabilities to boost high frequencies at the transmitter end. Compared to pre-emphasis, equalization has the advantage that it can more easily include adaptive mechanisms due to its location at the receiver, which provides a shorter feedback path. With adaptive equalization features, manual tuning of the circuit can be avoided for each link or every time environmental conditions change, something that suits a time varying channel, for instance one whose transfer function will be affected by radiation. Adaptive equalization is also beneficial to be able to make the same circuit compatible with a variety of different cable lengths that will be used in the ITk, without individual matching of the equalizer corrections to each cable prior to operation.

Signal degradation in the form of ISI is normally caused by a combination of the geometry of the circuit, the medium from which it is composed (conductor and dielectric) and the voltage swing of the signal. Since these parameters can be determined prior to signal transmission and are subject only to small random fluctuations, ISI can be corrected. Due to this deterministic nature of ISI, the information of the original signal emerging from the transmitter is still present in the received signal despite the eye diagram appearing closed. Equalization provides means to discern the original signal given a distorted received signal. By correcting for the high-frequency component voltage levels and thus the trajectories of these components in the corresponding eye diagram, equalization can assist in re-opening the eye. However, it should be noted that while the equalizer can correct for the degradation caused by the channel to a great extent, the eye will rarely be as clean as for the original signal even for a perfectly working equalizer. This is because in addition to ISI there will also be random noise present in the system, which due to its non-deterministic nature cannot be corrected for by conventional equalization techniques.

![Figure 5.8: Illustration of impulse input (Kronecker delta) and sampled impulse response with the main cursor (bit of interest) surrounded by pre-cursors (green) and post-cursors (blue) constituting ISI and separated by the sampling period T.](image-url)
5.2. Signal Recovery with Equalization

Ideally, a signal should be transmitted with a flat frequency spectrum over an infinite bandwidth such that every frequency component is equally represented. For a continuous analog signal in time domain, this translates to transmitting an infinitely narrow, infinitely high amplitude "impulse": a Dirac-delta function $\delta(t)$. The equivalent function in discrete form, making the number of components of the signal finite, would be a Kronecker delta\(^8\) function $\delta_{ij}$. When such an impulse is used as input, the resulting output will be the system impulse response. The purpose of equalization is to process the received signal in such a way that the impulse response of the channel is inverted and converted back to $\delta_{ij}$, in order to eliminate ISI and restore the initially transmitted signal. This means that for the equalizer to recover the original signal, its transfer function must be the inverse of the channel transfer function. The system impulse response can be used in order to derive a suitable equalizer with appropriate corrections applied to obtain this transfer function. Figure 5.8 illustrates a sampled version of a continuous impulse response, consisting of several cursors. A cursor is the voltage at the center of a bit. Following this terminology, a pre-cursor is the voltage at the center of any bit prior to the one of interest, while a post-cursor is the voltage at the center of any bit after. As the initially transmitted discrete impulse response is the Kronecker delta, which contained only the main cursor, both pre-cursors and post-cursors constitute ISI and should be eliminated by an equalizer.

5.2.3 High-Speed Link Equalizer Components

An ideal equalizer would remove all ISI so that only the signal and filtered noise remain. In order to achieve this, it may be necessary to apply different types of equalization techniques to a signal. Typically, a complete equalizer circuit for high-speed applications is comprised of three complimentary sub-circuits: a Continuous-Time Linear Equalizer to remove continuous long-tailed ISI, a Feed-Forward Equalizer for pre-cursor removal and a Decision Feedback Equalizer for post-cursor removal. Each of these equalizer types are described below.

Continuous-Time Linear Equalizer (CTLE)

A CTLE is an analog linear filter that attenuates low-frequency components of a signal, amplifies components around the Nyquist frequency, and filters out high-frequency components. The schematic of a simple active CTLE circuit can be seen in Figure 5.9 (a). By adding parallel resistors and capacitors in the source of a differential amplifier, the CTLE introduces resistive and capacitive negative feedback, thus achieving high-pass filter characteristics to counteract the low-pass nature of the channel. Figure 5.9 (b) shows the frequency response of this CTLE and of one of the ATLAS ITk pixel cable prototypes that it has been matched to, as well as how the overall system frequency response is flattened out by using the CTLE. Figure 5.9 (c) demonstrates how the eye is re-opened as the CTLE acts on it. The input signal is a pseudorandom binary sequence (PRBS), a test pattern that is generated by a deterministic algorithm but that exhibits statistical behavior similar to a truly random sequence and is able to stress the device under test by exploring its response to a wide variety of bit combinations.

\(^8\)The Kronecker delta function is defined as $\delta_{ij} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$
Feed-Forward Equalizer (FFE)

While the CTLE is a purely analog circuit, an equalizer can also be a mixed signal system acting in a discrete manner and yielding an extended flat frequency spectrum by constructing a peaking filter using the voltage levels of adjacent bits to correct the voltage level of a given bit of interest. In its simplest form, a discrete equalizer obtains the corrected voltage level at the location of interest by forming a correction sum of the voltage levels of the waveforms received at tap-delayed locations prior to the waveform associated with the bit of interest multiplied by correction factors, such that the equalized voltage level of bit $n$ can be compactly expressed as

$$e[n] = \sum_{k=0}^{N} c_k r[n - k]$$  \hspace{1cm} (5.2)$$

where $r[n]$ is the voltage level at the center of the received bit of interest, $r[n - k]$ is the voltage level of the bit received $k$ samples before the bit of interest and $N$ the number of tap-delayed locations. The voltage level of each bit will be multiplied by a coefficient $c_k$ specific for that bit in order to invert the distortion caused by the channel. The equalizer is essentially a tapped delay line filter, thus each such coefficient is referred to as a tap value or weight. Taps are dimensionless correction factors applied to the voltage levels that can be seen as the ratio of the voltage that the receiver should have seen to the voltage the receiver did see, and represents the ISI attributable to the corresponding symbol. The sampling period corresponds to one unit interval (UI), which is the symbol period (or the inverse of the data rate if one bit per symbol is transmitted) so that the taps are separated by the tap delay $T$ (see Figure 5.8). As a rule of thumb for the number of taps needed, the length of the impulse response can be used. In principle this length extends infinitely, while in practice it is the length until the amplitude falls below the maximum tolerated noise level. Thus, in Eq. 5.2 $N$ should be the number of pre-cursor canceling taps (or pre-taps) used to eliminate cursors appearing before the main cursor. Each equalizer tap adds to the complexity of the system, which increases costs, thus only taps which cancel cursors of appreciable magnitude should be used. More details on setting correct tap coefficients can be found in Section 5.3.2.
5.2. Signal Recovery with Equalization

The equalization model described by Equation 5.2 acts in a linear manner and only uses information from previously received bits, and is therefore a linear feed-forward equalizer (FFE). Like the CTLE, an FFE combats ISI by inverting the channel’s low-pass characteristics, but because of its discrete character it acts as a finite impulse response (FIR) filter. Meanwhile, due to the FFE being finite and thus not long enough to completely correct the impulse response containing continuous ISI, the long-tailed ISI is typically handled by a complementary CTLE. A block diagram of an FFE is displayed in Figure 5.10.

![Block diagram of a 2-tap FFE.](image)

Figure 5.10: Block diagram of a 2-tap FFE. \( r[n] \) is the received signal level of bit \( n \) and \( e[n] \) is the equalized signal level of bit \( n \).

**Decision Feedback Equalizer (DFE)**

The FFE uses information only from the current and previous bits and its operation may result in amplification also of the noise. Another set of corrections can therefore be included based on the best guess of the current and previous bits to further cancel ISI by including post-cursor canceling taps (or post-taps) used to cancel out cursors occurring after the main cursor.

The additional correction term uses a sum of logic decisions (i.e. whether a bit is a ’1’ or a ’0’) made on previous bits as feedback to the symbol decoder to help determine whether the current bit is a ’1’ or a ’0’. In contrast to the FFE, which corrects the received waveform with information about the waveform itself, this type of equalizer instead uses information about the logical decisions made on the waveform and is therefore called a decision feedback equalizer (DFE). A DFE introduces additional taps that are applied to a decoded signal and adds a feedback loop component to cancel the post-cursor ISI that is not eliminated by the FFE by matching the taps to the combined response of the FFE and the circuit. An \( M \)-tap DFE multiplies each of \( M \) recently received symbols by respective tap coefficients. The sum of these products is then subtracted from the received signal prior to sampling, thereby eliminating ISI associated with prior data. The correction value computed for each bit \( n \) can be mathematically described as

\[
 s[n] = \sum_{m=1}^{M} a_m d[n - m] \tag{5.3}
\]
where \( a_m \) is the correction coefficient for the bit \( m \) bits prior to the bit of interest and \( d[n - m] \) is the logical value of the data bit located \( m \) bits prior to bit \( n \). Referring to Equation 5.2, the equalized voltage level for bit \( n \) in a combined FFE and DFE circuit will become

\[
e[n] = \sum_{k=0}^{N} c_k r[n - k] - \sum_{m=1}^{M} a_{m} d[n - m]
\]

(5.4)

A block diagram describing such a combined circuit is shown in Figure 5.11. To be noted is also that the bit of interest is not used in the summation for the DFE, whereas the FFE uses previous bits and the current bit in its sum.

Figure 5.11: (a) Top level conceptual view of an equalizer system comprising a CTLE and an FFE combined with a DFE, and opening of the eye diagram at each equalizer stage. (b) Shift register description of an FFE combined with a DFE. \( r[n] \) is the received signal level of bit \( n \), \( d[n] \) is the logical decision of bit \( n \), \( s[n] \) is the DFE correction of bit \( n \) and \( e[n] \) is the equalized signal level of bit \( n \).

Generally, the voltage correction value computed by the DFE is added to the logical decision threshold of the data slicer, i.e. the threshold above which the waveform is considered a logical high and below which it is considered a logical low. However, the correction can be used either to shift the decision threshold up or down so that new logical decisions can be made on the waveform based upon this new equalized threshold level or, alternatively, to shift the waveform of the current bit relative to a constant threshold, thus having an equivalent effect. Shifting the voltage level of the bit of interest up by the same amount as determined by equation 5.3 as opposed to shifting the threshold down by this amount will have the effect on the eye of raising the corresponding trajectory, thus opening up the eye.

As the DFE uses previous logical decisions made on the waveform to obtain the corrected voltage offset for the threshold level at the bit of interest, for an \( M \)-tap DFE the \( M \) bits prior to the current bit need to have their bit levels determined. Thus, the
5.3. Equalizer Architecture and Circuit Implementation

DFE needs to be initialized by seeding a sequence of bit values for the first several bits. Assuming the logical decisions made on these initial bits are correct, the algorithm can then utilize that information to determine the logical value of the current bit. Unlike an FFE, the DFE will not amplify noise; however, a negative aspect of the nonlinear nature of the DFE is the error propagation that may follow as a consequence of a wrong decision being made. Both the FFE and DFE adds high-pass features to the circuit to flatten the overall frequency response, but the DFE requires an applied gain factor in order to preserve the initial signal amplitude as illustrated in Figure 5.12.

![Figure 5.12: Illustration of frequency response for transmission channel (cable), FFE and DFE.](image)

5.3 Equalizer Architecture and Circuit Implementation

5.3.1 Design Description

A first version of the Equalizer chip for the ATLAS ITk Pixel Detector transmission link has been realized in the form of an ASIC designed in a 65 nm CMOS process and fabricated by TSMC. This prototype transceiver design including equalization capabilities is called the Gigabit Cable Receiver (GBCR). The GBCR incorporates four upstream receiver channels operating at 5.12 Gbps, out of which three are baseline channels including a CTLE only, while the fourth is a test channel which includes an FFE+DFE module in series with the CTLE to act as a complimentary equalizer for further enhanced performance. In order to distinguish between the analog CTLE block and the tapped FFE+DFE circuit of the test channel using compact terminology, the latter will in what follows be referred to as the discrete equalizer (DEQ). An overview of the GBCR chip test channel in relation to the the full data transmission link can be seen in Figure 5.13. After passing through the cable and undergoing distortion, the transmitted signal is transferred to the test channel, where the input ports are connected to a CTLE for initial ISI removal before the subsequent discrete equalizer stage. The DEQ is delivering data to the Clock and Data Recovery (CDR) circuit, which extracts a reference clock signal to be fed to the equalizer and outputs data to the driver, which will transfer this data to the VTRx+ module on the opto-board. The CDR is a phase-locked loop circuit with a voltage controlled oscillator (VCO) designed to generate a low-jitter clock, and is migrated from a separate on-going data transmission project at CERN [52].
Chapter 5. ITk Challenge 2: Data Readout Speed

Figure 5.13: Block diagram of the GBCR chip test channel.

This section will be devoted to the design of the discrete equalizer only. The design of this stage builds on a receiver front-end circuit topology created by Berkeley Wireless Research Center (BWRC) developed for operation above 60 Gbps with minimal power consumption [53]. In order to accommodate the different design targets required by the equalizer of the ATLAS ITk data transmission system, the original circuit design has been significantly modified.

With the ability to reach multi-gigabit speeds, the original discrete equalizer design uses two interleaved branches to achieve double data rate functionality with each branch operating at half the clock frequency, but sampling data at rising and falling edges for the even and odd branch respectively. However, to obtain the desired simplicity of the ATLAS ITk equalizer, this design has been modified to consist of one branch only, operating at full data rate. In addition, the equalizer does not include an adaptation engine to update the tap weights, but instead the weights are fixed and matched to the cable and preceding CTLE circuit. Figure 5.14 shows a block diagram of the discrete equalizer.

Figure 5.14: Block diagram of discrete equalizer.

The discrete equalizer incorporates a 2-tap FFE complemented by a 3-tap DFE utilized to cancel pre-cursor and post-cursor ISI respectively. The delay elements of these tapped filters are realized through latches. The original circuit included dynamic latches to achieve sufficiently high speed; thus, operating the circuit at a lower data rate radically below the design value caused a significant voltage droop as the

---

9 A latch is an electronic logic circuit with two stable states that can be used to store state information.
latches were unable to hold a state for a longer period of time. For this reason, the dynamic latches were replaced with static, simply by extending the latch design with an additional hold branch that can store the state indefinitely instead of relying on the state being temporarily stored in parasitic capacitances (see Figure 5.15). The initial stage of the design consists of two feed-forward latches to provide UI-delayed signals to the pre-cursor-canceling FFE. The feedback filter of the DFE includes three additional latches for each of the post-cursor taps. The tap weights are set in an integrator block and a summer circuit is also included to sum up the tap values and add the correction to compute the equalized voltage level to be subtracted from the received signal prior to sampling.

![Dynamic latch consisting of sample branch only.](image1)

![Static latch with additional hold branch.](image2)

Figure 5.15: (a) Dynamic latch consisting of sample branch only. (b) Static latch with additional hold branch.

Each tap consists of a differential pair and the magnitude of the DFE tap coefficients are set by changing the tail current of each branch. In contrast, the FFE tap coefficients are instead set by varying the cascode gate bias voltage of a cascoded differential pair. This is to achieve tunable analog gain by varying load while keeping the overdrive\(^{10}\) of the input pair roughly constant for linearity across the gain range, which is a major challenge in integrating an FFE into the receiver [53]. The different tap designs are displayed in Figure 5.16.

The tap coefficients for the FFE and DFE, along with the bias voltages for the blocks included in the discrete equalizer, are part of the configuration parameters, which are set using digital-to-analog converters (DACs). The DEQ configuration includes 40 bits, which have been distributed across the parameters according to desired granularity and range determined to ensure proper circuit functionality across all front-end-of-line (FEOL) corners, as well as the high radiation corner, and variations in supply voltage and temperature. An overview of the configuration settings is shown in Table 5.3.

---

\(^{10}\)Overdrive voltage is defined as the voltage between the transistor gate and source and needs to be kept within a certain interval for the transistor to operate in a desired regime.
Figure 5.16: Schematics of an FFE tap where the weight is adjusted through the bias voltage of a differential pair (a) and a DFE tap where the weight is adjusted through the tail current (b).

<table>
<thead>
<tr>
<th>Config. parameter</th>
<th>Description</th>
<th>Default (TT)</th>
<th>Min.</th>
<th>Max.</th>
<th>Bits</th>
<th>LSB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vbias</td>
<td>Load voltage all blocks</td>
<td>350 mV</td>
<td>270 mV</td>
<td>430 mV</td>
<td>4</td>
<td>10 mV</td>
</tr>
<tr>
<td>Vsum</td>
<td>Load voltage summer block</td>
<td>500 mV</td>
<td>420 mV</td>
<td>580 mV</td>
<td>4</td>
<td>10 mV</td>
</tr>
<tr>
<td>Ibias</td>
<td>Tail current summer block</td>
<td>0.4 mA</td>
<td>0.05 mA</td>
<td>0.85 mA</td>
<td>5</td>
<td>0.025 mA</td>
</tr>
<tr>
<td>FFeTap</td>
<td>FFE tap weight</td>
<td>500 mV</td>
<td>340 mV</td>
<td>660 mV</td>
<td>5</td>
<td>10 mV</td>
</tr>
<tr>
<td>Itap0</td>
<td>DFE tap 0 weight (summer)</td>
<td>10 µA</td>
<td>0 µA</td>
<td>80 µA</td>
<td>4</td>
<td>5 µA</td>
</tr>
<tr>
<td>DFEtap1</td>
<td>DFE tap 1 weight</td>
<td>500 mV</td>
<td>380 mV</td>
<td>620 mV</td>
<td>4</td>
<td>15 mV</td>
</tr>
<tr>
<td>DFEtap2</td>
<td>DFE tap 2 weight</td>
<td>400 mV</td>
<td>280 mV</td>
<td>520 mV</td>
<td>4</td>
<td>15 mV</td>
</tr>
<tr>
<td>DFEtap3</td>
<td>DFE tap 3 weight</td>
<td>300 mV</td>
<td>180 mV</td>
<td>420 mV</td>
<td>4</td>
<td>15 mV</td>
</tr>
<tr>
<td>sgnp1</td>
<td>DFE tap 1 positive sign</td>
<td>VDD</td>
<td>VSS</td>
<td>VDD</td>
<td>1</td>
<td>VDD</td>
</tr>
<tr>
<td>sgnp2</td>
<td>DFE tap 2 positive sign</td>
<td>VDD</td>
<td>VSS</td>
<td>VDD</td>
<td>1</td>
<td>VDD</td>
</tr>
<tr>
<td>sgnp3</td>
<td>DFE tap 3 positive sign</td>
<td>VDD</td>
<td>VSS</td>
<td>VDD</td>
<td>1</td>
<td>VDD</td>
</tr>
<tr>
<td>sgnn1</td>
<td>DFE tap 1 negative sign</td>
<td>VSS</td>
<td>VDD</td>
<td>VDD</td>
<td>1</td>
<td>VDD</td>
</tr>
<tr>
<td>sgnn2</td>
<td>DFE tap 2 negative sign</td>
<td>VSS</td>
<td>VDD</td>
<td>VDD</td>
<td>1</td>
<td>VDD</td>
</tr>
<tr>
<td>sgnn3</td>
<td>DFE tap 3 negative sign</td>
<td>VSS</td>
<td>VDD</td>
<td>VDD</td>
<td>1</td>
<td>VDD</td>
</tr>
</tbody>
</table>

Table 5.3: Discrete equalizer configuration.

In order to ensure radiation hardness, all transistors were sized such that their dimensions were in compliance with the design guidelines specified in Table 4.1.

### 5.3.2 Equalizer Tuning

Physically, the equalizer taps correct for the deviation of the channel impulse response from a perfect delta pulse. Therefore, the impulse response can be utilized to optimize the tap values. Containing the same information as the impulse response in time domain, the transfer function describes how a signal is affected by a circuit in frequency domain. A way to characterize a physical channel in terms of its frequency response is to measure its S-parameters (or scattering parameters). The S-parameters of a communication network describes the electrical behavior in terms of transmission and reflection capabilities when undergoing stimuli by electrical signals, which also defines the transfer function of the system. Characterization of the customized ATLAS ITk
pixel cable prototypes listed in Table 5.2, in terms of estimating the amount of distortion they cause, was therefore conducted by extracting their respective S-parameters from measurements with an N5242A 26.5GHz network analyzer from Keysight Technologies. Figure 5.17 shows the transfer functions of the twinax cable prototypes, constructed from their measured S-parameters. As is evident from the plots, the cables exhibit low-pass filter characteristics, where the available bandwidth decreases with decreasing conductor diameter and increasing cable length. Thus, as the cables in the ITk system are required to be both low-mass (thin) and long enough to pass between the pixel FE modules and the opto-board several meters apart, they will give rise to severe signal distortion. From the transfer function each cable’s impulse response can be constructed and appropriate equalizer tap weights computed. Designing a filter specifically matched to the cable in this way will make the receiver equalizer of the data transmission link successfully eliminate ISI.

Figure 5.17: Transfer functions of ATLAS ITk twinax 34 AWG prototypes of different lengths.

Figure 5.18: eGUllizer modeling tool.
In order to facilitate characterization of the cable prototypes by using the extracted S-parameter data as well as to model the equalizer function, a graphical user interface (GUI) called "eGUIlizer" was developed, using Python and MATLAB®. The eGUIlizer framework allows for easy parameter tuning and visualization of the channel transfer function to compute available bandwidth and pulse response for determining the number of taps required depending on the cable, as well as eye diagrams to investigate eye opening achieved with equalization. It can also automatically compute optimized tap values of a 2-tap FFE combined with a 3-tap DFE (BWRC model) by solving a system of linear differential equations based on the pulse response and allow the user to complete the tuning manually as desired. Many of the core functionalities of the eGUIlizer build on a data transmission link evaluation framework developed by BWRC (see [54]). The eGUIlizer tool is pictured in Figure 5.18.

### 5.3.3 Simulations and Design Verification

The schematic and test bench for the discrete equalizer were created in the design and simulation environment Cadence Virtuoso, where performance simulations and subsequent layout, as well as parasitic extraction were also done. A summary of important circuit specifications can be found in Table 5.4. Details of the circuit verification process follow below.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature range</td>
<td>-15 °C to +65 °C</td>
</tr>
<tr>
<td>Supply voltage</td>
<td>1.2 ± 0.1 V</td>
</tr>
<tr>
<td>Radiation tolerance</td>
<td>&gt; 200 Mrad</td>
</tr>
<tr>
<td>BER</td>
<td>&lt; 10^{-12}</td>
</tr>
<tr>
<td>Power consumption</td>
<td>0.6 mW</td>
</tr>
</tbody>
</table>

Table 5.4: Verified circuit specifications for the discrete equalizer.

The circuit design was verified to function in all standard FEOL process corners as well as for the additional 200 Mrad radiation exposure corner model ("RAD200") developed by CERN (as described under Section 4.1.4). The standard two-letter naming convention for FEOL corners is used here, where the first letter refers to the NMOS corner and the second letter refers to the PMOS corner. The extremes under consideration are the typical (T) corner, as well as the fast (F) and slow (S) corners exhibiting higher and lower carrier mobilities than normal respectively. Because the corners "TT", "FF" and "SS" affect both device types evenly, they normally do not have a large impact on the logical correctness of the circuit. On the other hand, the "skew" corners "FS" and "SF", where the switching speed of the NMOS and PMOS devices is imbalanced, can cause the slews of the output edges to differ significantly. A slight indication of such behavior can be seen for the "FS" corner in Figure 5.19, where the results of the corner simulations are displayed. Meanwhile, these simulations confirm overall clean open eyes in all process corners.
5.3. Equalizer Architecture and Circuit Implementation

Figure 5.19: Eye diagrams of discrete equalizer output during process corner simulations with 5.12 Gbps PRBS input data.

Following the schematic verification, a circuit layout was made according to the CERN design rules along with the area constraints for integration into the GBCR chip. Parasitic extraction simulations were performed to create an accurate circuit model by calculating parasitic effects in both devices and interconnects. As can be seen in Figure 5.20, displaying the eye diagram at the output of the discrete equalizer, the final layout results in a circuit behavior essentially unaffected by parasitics.

Figure 5.20: Simulated eye diagram at discrete equalizer output without (blue) and with (red) parasitic effects for 5.12 Gbps PRBS input data.

Figure 5.21: Eye diagram at output of discrete equalizer at various temperatures in response to 5.12 Gbps PRBS input data.

It is important for the circuit to function properly under variations in operating conditions, such as differences in voltage supply or temperature. The operable tem-
perature range\textsuperscript{11} was confirmed to be -15–65 °C with jitter below 10 ps in the TT corner for the temperature range of interest 0–50 °C, as seen in Figure 5.21. In addition, the stability against variations in supply voltage was verified and the circuit was confirmed to function within $\pm \sim 10\%$ of the 1.2 V design value in terms of eye opening and data recovery, as can be seen in Figure 5.22.

![Figure 5.22: Bitstream (left) and eye diagram (right) at output of discrete equalizer in response to 5.12 Gbps PRBS input data for different levels of supply voltage 1.1 – 1.3 V.](image)

The design goal of a transmission system is to have zero errors, but in practice the bit error ratio (BER) should be less than the target of the particular system. The BER is defined as the number of bit errors divided by total number of transferred bits during some time interval and is a unitless performance measure, which for the GBCR is restricted to $10^{-12}$, meaning that at maximum one erroneous received bit is allowed per $10^{12}$ bits transmitted. Tolerated BERs are typically of this order of magnitude, which makes it time consuming to simulate the BER of a circuit as a very large number of bits needs to be transferred through it. However, instead of counting the number of bit errors by comparing transmitted and received bits, the BER can be estimated from the eye diagram. This is done by constructing an additional graph as displayed in Figure 5.23. For the same reason the eye diagram got its name due to the resemblance of an eye, this type of diagram is referred to as a \textit{bathtub curve} due to its shape. A bathtub curve displays the evolution of the eye diagram over time and can thus predict the amount of eye closure in the horizontal and vertical dimension after a given number of transmitted bits. The vertical opening of the eye diagram defines the noise margin, while the horizontal opening defines the jitter margin. A bathtub curve characterizes the relationship between the BER and the fraction of the UI available for sampling to provide evidence that a design functions as intended. By demonstrating the probability of bit errors over the unit interval, the bathtub curve can assist in determining where to sample to achieve a target bit rate. From Figure 5.23, it is evident that the CTLE alone does not satisfy the requirement of a BER $< 10^{-12}$, as the eye is almost completely closed (both lines nearly intersect) after $10^{12}$ bits. In

\textsuperscript{11}The circuit has the ability to operate over a further extended range of temperatures depending on the configuration settings; however, a wider range does not include temperatures of interest for the design application.
contrast, the eye of the discrete equalizer will remain sufficiently open to satisfy the BER requirement.

Figure 5.23: Bathtub curves of CTLE output (a) and discrete equalizer output (b) constructed with MATLAB’s Communications System Toolbox™. Blue color represents the horizontal eye opening and red color represents the vertical eye opening as shown in the upper eye diagram.

5.4 Equalizer Prototype Performance

5.4.1 The GBCR Prototype ASIC and Test Board Setup

The GBCR prototype ASIC has four upstream receiver channels, one downstream transmitter channel, an Automatic Frequency Calibration (AFC) module, and an I²C slave module to read and write to the internal chip registers. The AFC is able to calibrate a VCO in each receiver channel to generate a 5.12 GHz clock signal. The four receiver channels share one AFC module, with one VCO per channel as part of a CDR block, as shown in Figure 5.24.

Figure 5.24: GBCR block diagram.
The GBCR board is displayed in Figure 5.25 along with the bare GBCR ASIC, which is housed inside a 40 pin package. The input and output (I/O) interfaces of the GBCR board consist of 26 SMA connectors. Apart from the differential I/O of the receiver channels 1-4, these connectors also include differential I/O for the transmitter channel, external calibration clock input, and test output for the VCO of channel 4. In addition, there is one SMA pair in channel 1 that was intended to be used as a probe point for the differential output from the summer block of the discrete equalizer in order to verify its functionality. Due to tight time constraints before the submission of the ASIC for fabrication, a buffer was never designed to enable this and therefore this I/O pair is left unused.

The GBCR chip underwent characterization using the test setup displayed in Figure 5.26. Multi-gigabit PRBS input signals are generated with a Xilinx Kintex-7 FPGA board by configuring onboard transceivers through the IBERT IP core of the integrated Vivado serial I/O analyzer [55]. The input is connected to the GBCR board channels via 30 cm long SMA connector cables and the same type of cables are used to connect the channel outputs to a Tektronix TDS7044B Digital Phosphor Oscilloscope. Alternatively, for tests of signal recovery after distortion caused by a cable, the input signal is first passed through a 5 m 34 AWG twinax and then to the channels on the test board via SMA cables. The cable interfaces via custom SMA adapters. The GBCR board is powered from an Agilent U8001A DC power supply through clip cables hooked to the onboard power pins. The chip is configured by writing to the registers through an I²C bus from a customized LabVIEW program running on a PC, which communicates with the GBCR board through a USB-I²C interface module.

![Figure 5.25: GBCR prototype test board.](image1)

![Figure 5.26: Illustration of GBCR test setup at LBNL.](image2)
5.4.2 Summary of Test Results

The discrete equalizer block and the CDR block were designed independently and integrated at the top level. The interfaces between these blocks are differential input from the output stage of the discrete equalizer, which is the data slicer (see Figure 5.14), to the CDR and differential clock signal from the CDR supplied to the internal discrete equalizer blocks. In order to meet the submission deadline, a top level simulation without and with parasitic extraction to verify the functionality of the test channel was not done after integration of the two blocks. Even though this was understood to be a risk, there was no possibility to delay the submission, and thus this was the only chance to test the discrete equalizer circuit in silicon. The major difference between the simulations demonstrated in Section 5.3.3 and a proper simulation at the top level is that the former used an external clock source instead of the recovered clock signal provided by the CDR.

Results from measurements with the fabricated chip show that the baseline channels 2-4, including the CTLE only, performed according to expectations\(^{12}\) and the distortion caused by a 5 m 34 AWG twinax cable connected between the signal generator FPGA board and the channel inputs is reduced, as seen through the re-opening of the eye diagrams in Figure 5.27. Meanwhile, the test channel (channel 1), including the discrete equalizer, did not achieve expected performance as predicted by simulations. In fact, injecting a 5.12 Gbps PRBS input signal through channel 1 did not produce any measurable output, as shown in Figure 5.28. This remains true even when tuning the equalizer by changing the bias and tap settings of the 40 bit configuration (see Table 5.3).

![Eye diagrams](figure.png)

Figure 5.27: Eye diagrams of 5.12 Gbps PRBS input to 5 m 34 AWG twinax cable, distorted output from passing through the same cable, and corresponding equalized output from channel 1 (test channel with discrete equalizer) and channels 2-4 (with CTLE only) with oscilloscope settings 150 mV per division in the vertical dimension and 125 ps per division in the horizontal.

\(^{12}\)The interfacing SMA cables used during testing introduce some distortion which is why the eye diagram appears more closed than expected, even on the transmitting end before passing through the twinax cable.
Two problems causing the issues with channel 1 were identified:

i) **Overloading of data slicer input.** The signal path to the slicer input inside the discrete equalizer block was loaded with excessive capacitance in the top level due to the connection of this node to outputs with no buffer, which was the result of an integration error of the discrete equalizer into the full GBCR circuit. This node was not intended to be brought out as an output, and the mistake could have been discovered by a top level simulation with layout parasitics.

ii) **Low CDR clock output voltage swing.** The VCO output from the CDR block has a lower voltage swing than required for the DFE slicer to function properly. This mistake could have been avoided and the slicer design corrected if time would have permitted a proper schematic top level simulation to be performed prior to device fabrication.

Once these problems were understood, a more realistic simulation was done and this simulation is close to reproducing the measured behavior. Therefore, in the end the GBCR prototype test channel is fully understood, but cannot be used to validate the intended performance of the discrete equalizer in silicon. Changes to the operating conditions to compensate for the low CDR clock swing were able to partly recover performance, but not enough for a full speed demonstration. The following subsection presents further details on the analysis of the hardware tests and post-fabrication simulations to better understand the measured behavior of the discrete equalizer. Unless otherwise stated, the measurements and simulations have been performed by injecting a PRBS bitstream directly into the input of channel 1 without first letting it pass through a cable, in order to gain initial understanding of the circuit behavior without excessive distortion.
5.4.3 Test Result Analysis

A first attempt to combat problem (i) above, the issue with capacitive overloading of the DFE slicer input due to unintentionally connecting an unbuffered output during design integration, was to cut the circuit traces between the DFE summer and the I/O pads using a Focused Ion Beam (FIB) technique, as illustrated in Figure 5.29. However, this alone was not enough to recover the intended performance.

![Figure 5.29: (a) Oscilloscope image of eye diagrams of PRBS input to channel 1 and corresponding output at the DFE summer test pads before FIB modification. (b) GBCR chip layout with discrete equalizer channel marked in blue. The mistakenly connected differential output from the DFE summer is marked in red. (c) Illustration of FIB modification (overlaid on closer view of GBCR chip layout) to cut the traces (yellow) between the DFE summer output and the I/O pads in order to reduce the capacitive loading on this node.]

Since there are no probe points for the different blocks of the discrete equalizer, understanding problem (ii) above through hardware tests, which are limited to studying the channel input and output only, remains a challenging task. However, the origin of the issue could be found through simulation studies, where it was finally traced to the design error of the data slicer block inside discrete equalizer circuit. Even when the input to the slicer is a clean data signal coming from the DFE summer, the output is flat, as seen in Figure 5.30.

![Figure 5.30: Simulation results of 5.12 Gbps PRBS input signal to GBCR channel 1 and corresponding output data (left). Simulation results of slicer input data from DFE summer block and corresponding output from the slicer (right).]

In order to troubleshoot this issue, square pulses of decreasing frequencies were injected through channel 1 with a HP 8133A pulse generator to investigate where there was a bandwidth cutoff. As demonstrated in Figure 5.31, 160 MHz is the highest
frequency signal that can be transmitted through channel 1 and properly recovered, albeit attenuated. The corresponding data rate of merely 320 Mbps is significantly lower than the performance requirements for the GBCR chip.

![160 MHz input](image1)

![320 MHz input](image2)

**Figure 5.31:** Square pulse inputs of 160 MHz and 320 MHz (pink) and corresponding outputs of channel 1 (cyan).

The interdependency of the discrete equalizer and CDR makes a top level simulation of the full GBCR especially important. More specifically, the discrete equalizer requires a proper clock signal to function, but this signal is supplied by the CDR, whose output in turn depends on a proper input signal from the discrete equalizer. Failure to verify that this interdependence was functioning caused the slicer design error to pass unnoticed. Instead, the slicer was simulated with an externally supplied clock source to emulate the VCO output of the CDR block. As seen in Figure 5.32, this clock has a significantly higher amplitude than that produced by the VCO with the discrete equalizer connected in a loop with the CDR. While the common mode voltage of the differential clock signals externally supplied ($V_{cm} = 800$ mV) and generated by the VCO ($V_{cm} \approx 860$ mV) are not too far apart, the single-ended amplitude swings between $\pm 600$ mV for the former and $< \pm 200$ mV for the latter. Figure 5.33 shows results from simulations of the discrete equalizer output when the circuit is supplied with external clocks of different amplitudes, demonstrating how the data recovery ability worsens with decreasing clock swing and why the clock cannot simply be treated as a digital signal without a level requirement.

![External clock](image3)

![Recovered clock](image4)

**Figure 5.32:** Simulation of externally supplied clock source used during design of slicer (purple) and recovered clock generated from VCO when discrete equalizer connected (green).
5.4. Equalizer Prototype Performance

The clock amplitude is important since the discrete equalizer uses dynamic logic to a great extent in order to reach high speeds, with clocked transistors on both pull-up and pull-down paths. Figure 5.34 (a) shows a dynamic gate topology similar to the one used inside the data slicer where a clocked PMOS transistor serves as resistive pull-up. This type of logic operates in two phases: *pre-charge* when the clock is low, turning on the PMOS device and pulling the output high, and *evaluation* when the clock is high, turning off the PMOS device and evaluating the output through the NMOS transistor stack. As can be seen in Figure 5.34 (b), the PMOS device will operate mostly in the cut-off region, \( V_{\text{SG}} < |V_{\text{th}}| \approx 400 \text{ mV} \) when supplied with a clock generated internally from the VCO. This explains why no output is visible in Figure 5.28. In contrast, 5.34 (c) shows that a higher amplitude clock will periodically turn on the PMOS \( V_{\text{SG}} > |V_{\text{th}}| \approx 400 \text{ mV} \) and allow the dynamic operation to function as expected, which is why simulations with the externally supplied clock signal did not show any issues with the slicer.

Figure 5.34: Dynamic gate topology with clocked pull-up PMOS transistor (a). Simulation results of operating conditions for pull-up PMOS inside data slicer in response to a 5.12 Gbps PRBS input when supplied with internally recovered clock (b) and external clock source (c).
5.4.4 Discrete Equalizer Prototype Improvements

Solutions exist to combat the issues with the data slicer without redesigning it. The loss in output voltage that occurs in the slicer can be mitigated by increasing the over-drive voltage between the source and gate $V_{SG} = V_S - V_G$ of the PMOS pull-up to force it out of the cut-off region. This can be done by increasing the voltage swing of the clock signal to impact $V_T$, which is what the original slicer design was intended for. Instead of supplying an external clock, an amplifier can be inserted in the loop between the discrete equalizer output and the CDR input to make sure the interdependency between these blocks remains. Alternatively, $V_{SG}$ can be increased by increasing the supply voltage to the slicer $V_{DD}$, since $V_S = V_{DD}$ for the PMOS. Figures 5.35 and 5.36 show how successful data recovery is achieved through these solutions and confirm the cause of the problem with the slicer.

![DEQ input](image)

**Figure 5.35:** Simulated input (red) and output (blue) to discrete equalizer when slicer supply voltage $V_{DD_{slicer}}$ increased to 1.5 V, while the supply voltage to all other blocks is kept at the design value of 1.2 V.

![DEQ output](image)

![PRBS](image)

![CTLE](image)

![DEQ](image)

![CDR](image)

![recovered clock](image)

![CLK](image)

![PRBS input](image)

![DEQ output](image)

![amplified clock](image)

![recovered clock](image)

**Figure 5.36:** Comparison of clock signal (blue), input data (red) and output from DEQ (green) for original circuit design (upper) and with an amplifier inserted in the clock path between the CDR and the DEQ (lower).
Nevertheless, while these modifications are possible in simulation, they are not possible to implement in hardware without re-fabricating the ASIC. Solutions for the existing prototype chip are limited to modifications of environmental and operating conditions. Such solutions include increasing the supply voltage for the full circuit, as opposed to the slicer alone, or increasing the temperature at which the circuit operates in order to decrease the threshold voltage $|V_{th}|$ to allow operation above cut-off with a low $V_{SG}$. The transistor threshold voltage decreases with temperature due to the decrease in Fermi potential, as can be seen in Figure 5.37 (a). Figure 5.37 (b) confirms that running the chip at higher operating temperatures improves performance, but not enough for proper functionality within a reasonable temperature range, with the maximum operational temperature of the ITk set to 60 °C.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{pmos_threshold_volt.png}
\caption{(a) Simulations of threshold voltage of PMOS pull-up transistor in data slicer vs. operating temperature. (b) Simulations of slicer input (red) and output (blue) at various temperatures.}
\end{figure}

Similarly, although increased supply voltage up to the chip limit of 1.5 V mitigates the detrimental effects caused by the data slicer to some extent, in terms of increased ability to detect shorter pulse widths, the improvements are not sufficient to achieve adequate data recovery. This has been demonstrated in simulation, as shown in Figure 5.38, and confirmed by hardware tests as seen in Figure 5.39.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{vdd_sim.png}
\caption{Simulations of channel 1 response to 5.12 Gbps PRBS input at different supply voltage levels.}
\end{figure}
Figure 5.39: Oscilloscope images of channel 1 output (cyan) in response to 5.12 Gbps PRBS input (yellow) at different supply voltage levels.

Figure 5.40 shows simulation results demonstrating that, for a square pulse up to 640 MHz corresponding to a data rate of 1.28 Gbps, full data recovery is possible if VDD is increased to 1.45 V. However, these simulations have been performed without added parasitic effects and the results confirming this data rate limit have not been reproducible in hardware measurements, in which the upper data rate that can be recovered remains 320 Mbps (cf. Figure 5.31) even during operation with a higher supply voltage.

Figure 5.40: Simulations of square pulse at different frequencies and PRBS at different data rates as inputs to channel 1 (red) and corresponding outputs (blue) at various levels of supply voltage VDD from 1.2 V (design value) to 1.5 V (chip limit).
Figure 5.41: Channel 1 output (cyan) in response to 5.12 Gbps PRBS input (yellow) at VDD = 1.5 V in simulation (a) and oscilloscope image from hardware tests of the GBCR prototype (b).

Figure 5.41 shows a comparison between a 50 ns simulation of the channel 1 output in response to a 5.12 Gbps PRBS input sequence and the corresponding signals during hardware test of the GBCR prototype displayed on the same time scale, both with the supply voltage raised to 1.5 V. The simulations give better results, in terms of greater bandwidth with higher frequency components being recovered, but this is likely due to parasitic effects not being accounted for in the simulations, in addition to the SMA cables used during the hardware measurements reducing the overall system bandwidth.

Figure 5.42: Block diagram of modified design of the discrete equalizer with data slicer removed.

For a possible future iteration of the discrete equalizer, a better solution to eliminate the slicer issue is to completely remove the slicer block and connect the summer output directly to the CDR input, as shown in Figure 5.42. This is possible as the
slicer is not part of the DFE feedback path, and its main function is to operate as a sense amplifier to amplify small amplitude signals to recognizable logic levels. In fact, the slicer block is essentially just a remain from the original BWRC circuit design, where it was used to produce rail-to-rail outputs for the output stage. However, the slicer does not fill an instrumental function in the GBCR circuit as the discrete equalizer output is connected to the CDR and not in need of further amplification, and it can therefore be bypassed without issues, yielding satisfactory data recovery and a clean eye diagram as demonstrated in Figure 5.43. The jitter is reduced compared to using a CTLE only; however, not to the extent shown in simulations with an external clock. This is most likely partly due to further tuning of the configuration parameters being needed to operate with the recovered clock and after the circuit modification of removing the slicer block.

![Eye diagrams](image)

Figure 5.43: Eye diagrams from 50 ns simulations of 5.12 Gbps PRBS input (red), 5 m 34 AWG twinax output (green), CTLE output (orange), and discrete equalizer output for original design with VDD increased to 1.5 V and modified design with slicer bypassed (blue).

### 5.4.5 Future GBCR Prototype

Existing problems with the test channel of the GBCR, pertaining to the data slicer design, have been identified and removing the slicer block is suggested as the preferred solution for a possible future prototype of the discrete equalizer. However, the next iteration of the GBCR chip will not contain a discrete equalizer. Underlying that decision is the fact that, due to scheduling reason, the active cable development will continue as an alternative approach rather than a primary solution. The active cable was meant to be deployed to make the challenges of the electrical transmission system decoupled from the front-end chip development, on a technical level as well as with regards to scheduling. Meanwhile, the integration has been shown to be more complex than expected, given the timescale of the project. With the Pixel Detector production scheduled to start in 2021, the time limit did not permit development of a
fully-functioning aggregator chip. In addition, the main motivation behind the aggregator chip was to reduce cable mass through multiplexed data lines, but in the primary plan auxiliary mass required for the active cable, such as for powering, was not accounted for. Ultimately, the active cable solution was therefore shown to not result in a significant mass reduction as compared to a readout scheme without aggregation. The baseline architecture currently being implemented is instead readout of individual pixel front-end modules at 1.28 Gbps without the need for data aggregation, so-called "direct transmission" as illustrated in Figure 5.44.

The reduction of the required cable bandwidth arising from the lower data rate compensates for the increased distortion caused by a longer cable as the distance between the driver and the location of the receiver will increase by around 1 m. Figure 5.45 shows the transfer function of a 6 m twinax, which is the longest length considered in the new baseline ITk transmission system (with margins added), demonstrating an 11 dB loss at the 1.28 Gbps upper data rate limit, corresponding to a minimum required bandwidth of 640 MHz. This is compared to a 5 m twinax demonstrating an 18.5 dB loss at the 5.12 Gbps upper data rate limit of the original ITk transmission system including the active cable solution, corresponding to a minimum required bandwidth of 2.56 MHz. As equalization is increasingly efficient the higher the bandwidth of the system, it remains crucial for high-speed applications while being superfluous for low-bandwidth systems, where it has little or no effect. Further, if the data transmission rate is low enough, pre- and post-cursors may not be present in the impulse response and a discrete equalizer would be completely redundant. Consequently, with the data rate reduced to 1.28 Gbps, the CTLE technology alone will be able to eliminate enough ISI to provide a clean enough signal for clock and data recovery, without need for further equalization through a more advanced filter such as an FFE and/or DFE. Thus, a discrete equalizer would simply add unnecessary costs and complexity to the system. While not utilizing multiplexing of data lines increases the number of cables in the ITk readout system (the uplinks are estimated to consist of a total of 34 772 twinax cables), it will have a small effect on the total cable mass as the cable diameter can be reduced (higher AWG) due to the lower data rate which requires less bandwidth per cable. A receiver chip for this new baseline readout system is under
development, for which CTLE is the only equalizer type being implemented and is currently being re-optimized for 1.28 Gbps equalization to compensate for loss up to 20 dB at 640 MHz.

![Transfer functions of 34 AWG twinax cable prototypes of 5 m and 6 m.](image)

Figure 5.45: Transfer functions of 34 AWG twinax cable prototypes of 5 m and 6 m.

### 5.5 Pulse Amplitude Modulation Studies

A possibility to reduce the cable mass inside the tracker volume of the ATLAS ITk further could be to use a pulse amplitude encoding scheme. Standard signal modulation schemes for serial communication are commonly based on binary so-called Non-Return-to-Zero (NRZ) encoding, in which ones and zeros are presented by one specific condition each, usually a positive voltage and a negative voltage respectively, with no other neutral condition. In such schemes, one baud (Bd) carries one bit. The baud is a measure of the number of times a signal changes state per unit time, equivalent to the number of bits per symbol. Thus, if more than one bit of data is transferred per symbol, the bit rate at a given baud rate can be increased.

![Eye Diagrams](image)

Figure 5.46: Illustration of traditional NRZ modulation scheme, PAM2, (left) and PAM4 using Gray encoding (right).

Traditional NRZ can be seen as a two-level pulse amplitude modulation (PAM) system, where data is encoded in the amplitude of a series of signal pulses shifting between two distinct amplitude levels. Lately, in order to reach speeds in the higher multi-gigabit regime, PAM techniques of higher degree are instead becoming extensively popular, with more data encoded into the same timeframe using multiple signaling levels. Four-level PAM, or PAM4, is currently the most popular multilevel signaling alternative to NRZ. PAM4 uses four levels to encode two bits of data, i.e.
5.5. Pulse Amplitude Modulation Studies

one baud carries two bits, and this is commonly achieved by using so-called Gray encoding, wherein each amplitude level 1-4 corresponds to the logical bit mapping ('00', '01', '11', '10'). Such a coding pattern facilitates error correction, since successive symbols differ by one binary bit only. In contrast to NRZ/PAM2 which produces only one "eye", PAM4 instead produced three "eyes", as seen in Figure 5.46, along with an illustration of the bit encoding of the amplitude levels.

![Cable transfer functions](image)

Figure 5.47: Results from simulations of PAM4 signaling using the S-parameter data extracted from the ATLAS ITk pixel twinax cable prototypes.

By using a PAM4 transmission scheme, each data line transferring data at a baud rate of 1.28 GBd over a twinax cable in the ATLAS ITk pixel system would operate at a bit rate of 2.56 Gbps instead of 1.28 Gbps as for the current NRZ scheme. This would be beneficial, since the number of twinax cables required to achieve the same bit rate could be cut in half. In order to evaluate whether PAM4 is a viable option for the ATLAS ITk transmission system, S-parameters from the twinax prototypes presented in Table 5.2 were used to investigate the distortion caused by the cables with this scheme. An emulator was created in MATLAB, where a pseudo-random bit sequence was encoded according to Gray mapping to generate PAM4 symbols, and this sequence was then fed through 30 AWG cables of length 1 m, 3 m, 5 m and 9 m. The results of this study for a data rate of 1.28 Gbps, corresponding to a baud rate of
640 MBd per second, are displayed in Figure 5.47. The bandwidth will be reduced to 320 MHz, as compared to 640 MHz for transmission at the same data rate with a NRZ scheme. It is evident that severe distortion is caused for cable lengths above 5 m, but this could be battled by equalization.

This study aimed to investigate if PAM4 would be a feasible alternative to reduce the number of cables for the ATLAS ITk pixel data transmission link, given the noise present in the system. The PAM4 technique trades off bandwidth for SNR as it reduces the voltage difference between the signal levels representing symbols, making noise a significant factor as the eyes become more sensitive to noise than in a traditional NRZ system. In general, any signal degradation has greater impact on PAM4 than it would have on NRZ. Further, decoding more than two voltage levels is typically more difficult, and adds cost and complexity to the hardware. For instance, PAM4 transmission includes 16 symbol transitions, 6 different rise/fall times, the need for three slicers at the receiver to detect all three eyes, and also makes it difficult to predict bit-error-ratios from symbol-error-ratios.

PAM4 begins to pay off when there is escalation in cabling cost and material and could therefore be considered for the ATLAS ITk pixel data transmission link. However, determining whether PAM4 could be a viable option for this link requires a more thorough investigation of all system components and of higher AWG cables as well as a receiver design utilizing CTLE and adapted for PAM4 decoding.
Chapter 6

Summary and Conclusions

The ATLAS inner tracker (ITk) will be commissioned during the Phase-II Upgrade and has been designed to cope with the challenging demands following the High Luminosity Large Hadron Collider (HL-LHC) installation, where the integrated luminosity will be increased by a factor of 10. The ATLAS ITk will cope with around 5 times higher pile-up than the present inner detector system, but should achieve at minimum identical performance with higher η coverage. Two of the most challenging demands on this system is to ensure proper tracking functionality under extreme radiation exposure, as well as to handle the increased data readout speed. For this to be possible, both the detector hardware as well as the software used to reconstruct events from data retrieved from the detector modules must be carefully designed. This thesis has presented solutions pertaining to each of the main challenges facing the ITk in terms of improvements of detector hardware and simulation software for particle track reconstruction, in order to enhance the overall detector performance adequately for the HL-LHC operation starting in 2026.

6.1 ITk Challenge 1: Radiation Damage

A simulation model has been developed to account for radiation induced defects in 3D silicon pixel sensors. An algorithm for charge carrier transport within the silicon was created to emulate the charge collection process and how it is impacted by radiation induced trapping centers. From simulations with this model, the amount of signal lost as a consequence of radiation exposure of the sensors at various fluence levels was computed. These studies show that the expected fluence received by the innermost pixel layer at the end of the lifetime of the ATLAS ITk will result in a signal loss around 40%. Consequently, taking these effects into account is crucial for accurate reconstruction of particle trajectories through the detector as new physics discoveries depend on precise tracking data. The simulation model developed and integrated into the main ATLAS simulation framework will thus carry great importance for the tracking detector performance, as well as for making important decisions on how to operate the detector in extreme radiation environments.
6.2 ITk Challenge 2: Data Readout Speed

A discrete equalizer circuit based on an existing design from Berkeley Wireless Research Center, containing a feed-forward equalizer (FFE) and a decision feedback equalizer (DFE), was designed and fabricated as part of the gigabit cable receiver (GBCR) chip for the planned active cable solution of the ATLAS ITk pixel data transmission readout link operating at 5.12 Gbps. Simulation studies showed how this discrete equalizer was able to efficiently remove inter-symbol interference and improve the performance of the receiver in terms of eye diagram opening and significant jitter reduction, and displayed stable operation during variations in temperature and supply voltage, as well as extreme process corners and after radiation exposure up to 200 Mrad. Due to time constraints of the integrated circuit submission, a top level simulation of the complete data chain with the discrete equalizer integrated into the full GBCR circuit was not performed prior to chip fabrication, resulting in a design error of the data slicer contained in the discrete equalizer, which caused a malfunction of the circuit at higher data rates. Subsequent simulation studies identified the preferred solution of this issue for a possible future prototype of the discrete equalizer as removal of the slicer block and showed how satisfactory performance could thus be restored. Meanwhile, the transmission hardware that motivated investigating the discrete equalizer as a complement to the analog continuous-time linear equalizer (CTLE) is no longer being developed as a baseline solution for the ITk Pixel Detector readout system; thus, the next iteration of the GBCR chip will not contain such a circuit. However, while the discrete equalizer is not being implemented at this point, the simulation studies and design process provided valuable insights of how a current industry-standard solution for high-speed transmission, previously unexplored within high energy physics applications, can be deployed in a particle detector readout system. By customization and adaptation of the discrete equalizer to the unique demands of particle tracking detector readout, improved performance of a future system operating at higher data rates has been proven possible by including such an equalizer.
Appendix A

Readout and Data Acquisition Systems for PET Detectors

The fact that high energy particles are able to penetrate various solid objects is something which is of essential use within the field of medical imaging, where internal body structures and physiological processes are revealed non-invasively, i.e. without introducing an instrument into a patient, in order to diagnose and treat disease. Positron Emission Tomography (PET) is one such imaging modality for which the radiation detection and readout scheme share many similarities with a tracking detector for HEP applications, as illustrated in Figure A.1.

Figure A.1: High level illustration of detection principle, data readout and reconstruction for a HEP tracking detector versus a PET detector.

In addition to the work presented in the main part of this thesis, the author has carried out two smaller projects within the field of PET, which will be compactly summarized in this appendix. Starting with a short review covering the basics of PET, each project will then be described by first introducing the background to put the work carried out into context, followed by a summary of the results. The first project on the development of a DOI-TOF PET module was carried out with the Imaging Physics Team at the National Institute of Radiological Sciences (NIRS) and the second project on a readout system setup for characterization of PET scintillation crystals was done at Universidad Carlos III de Madrid (UC3M).
A.1 PET Basics

Molecular imaging is a branch of nuclear medicine and an imaging modality which encompasses both diagnostic imaging and treatment of disease. It is a type of functional imaging that visualize biological processes and physiological activities within a certain tissue or organ to allow detection of abnormal cell activity, such as cancerous tumors. This stands in contrast to techniques such as magnetic resonance imaging (MRI) or computerized tomography (CT) using X-rays, which only reveal the anatomic structure of the patient or subject. Multi-modality imaging, i.e. performing simultaneous image acquisition using functional and anatomic modalities combined, gives a more complete insight into the workings of the human body at a cellular and molecular level in addition to the anatomic structure.

In conventional diagnostic imaging, an external energy source, such as magnetic fields, X-rays or ultrasound waves, is used to produce images of bone and soft tissue. Molecular medicine techniques instead utilize radiation coming from interaction with biological tissue and certain properties of short-lived radioactive isotopes injected into a patient causing energetic particles to be emitted as a consequence of their decay. Positron Emission Tomography (PET) is one such technique which can detect these particles and correlate the emissions to regions of biologic activity within the body. Since the likelihood of absorption of the isotopes is related to the metabolic rate of the cells, estimates of the concentration of a radiotracer within the body can be used to form a functional image in order to identify tumors, fracture points in bone or other signatures associated with disease.

![Figure A.2: Illustration of Positron Emission Tomography (PET).](image-url)

Figure A.2 illustrates a PET scan procedure, during which the patient is injected with a positron emitting isotope incorporated with an organic substance such as glucose, typically fluorodeoxyglucose (F-18), that is used as a marker of metabolic utilization. The tracer substance will accumulate in certain target body regions and the
radionuclide will undergo positive beta decay, i.e. a proton in the parent nucleus decays into a neutron that will remain in the daughter nucleus and a neutrino and a positron will be emitted in the process. The positrons will annihilate with electrons inside the body, thereby producing two gamma rays back-to-back, each with energy equivalent to the electron rest mass of 511 keV. Such signature photons will be registered by detectors arranged around the patient, consisting of finely segmented scintillation crystals which absorb the energy of incoming 511 keV gamma rays and re-emit it in the form of light. The scintillators are coupled to photodetectors, which count light photons by converting them into an amplified current signal as electron-hole pairs are released in the depletion region of a p-n junction in the detector. Image data is acquired by detecting two annihilation photons within a certain coincidence window (~ns), and each such coincidence event is assigned to a line of response (LOR), joining two detectors where a signal was registered. By superimposing several LORs (~10⁶) from the detector count data, positional information of the recombination is gathered, from which the radiotracer activity distribution can be reconstructed. The PET images are formed through tomography techniques, by combining 2D images of the structures in a thin section of the body to create a detailed image of the object. Such images of the activity distribution throughout the body indicate the location of abnormally functioning cells, with higher metabolic activity than healthy cells and can reveal rapidly growing tissue, like tumors, metastasis, or infection.

![PET detector module flood histogram created from centroid computations of counts registered in photodetectors X₁, X₂, Y₁, Y₂. The photodetectors are displayed as light gray squares with solid lines in the flood histogram, and the crystals as smaller squares with dashed lines. The orange circles correspond to the peak of the count distribution within each crystal.](image)

The readout system of a PET scanner can be optimized by using different schemes to extract the signals from the sensors and translate it to position information. One-to-one coupling between the crystal element and the photodetector can be used, but this requires a large number of readout channels which increase the system costs and adds complexity. Alternatively, several crystals can be coupled to the same photodetector in order to minimize the number of electronic channels. This is beneficial, because a photon absorbed by the scintillator will cause isotropic emission of light that propagates down a particular crystal of interaction but also spreads across multiple photodetector elements, and this light sharing can be used to for optically multiplexed readout. The light sharing property can be enhanced by inserting a light guide between the crystal and the photodetector to diffuse the scintillation light. The distribution of
current across the photodetectors will depend on the location of the scintillating light flash, and given the outputs from the photodetectors a weighted centroid positioning algorithm can be used to estimate the photon interaction position. By uniformly irradiating the photodetectors with a gamma source and using their outputs to create a two-dimensional distribution of horizontal ($X$) and vertical ($Y$) positions of detected events, a probabilistic map, or flood histogram, where each peak corresponds to a single crystal in the detector array, can be constructed as illustrated in Figure A.3.

Accurate image reconstruction from PET detector data requires calibration of the light spread pattern in the scintillator block to physical crystal locations. Each detector module is therefore calibrated with a crystal look-up table that delineates events in one crystal from those in adjacent crystals by segmenting the flood histogram into crystal regions containing one peak each so that each combination of photodetector outputs generated from a unique light pattern response is mapped to a corresponding crystal of interaction in the array. During image acquisition the gamma ray ($X$, $Y$) position of each event can directly be mapped with the crystal look-up table and assigned to a crystal and to a correct LOR. Light-sharing performance can be assessed with the flood histogram to evaluate the crystal identification ability of a module. Figure A.4(a) illustrates a blurry flood histogram, which is difficult to separate into crystal regions, while Figure A.4(b) shows a clearer flood histogram where crystal regions can easily be identified and sectioned to create a crystal look-up table.

![Figure A.4: Illustration of blurry flood histogram with wide count distributions (a) and high resolution flood histogram with narrow count distributions divided into crystal regions (b).](image)

## A.2 Development of PET-DOI detector with TOF capability

### A.2.1 Background

As one of the most prominent methods to ensure effective treatment, PET imaging can also be used for range verification of the dose delivered to the patient during particle therapy, in which beams of energetic ionizing particles (for instance protons) are used to supply a radiation dose to target tumors in order to damage the DNA of the cancerous tissue cells and thereby eliminate them. This monitoring can be done either by transferring the patient to a separate PET scanner, which can result in errors and loss in accuracy, or by using an in-beam system during therapy, where the angular
A.2. Development of PET-DOI detector with TOF capability

coverage is limited by the volume occupied by the treatment beam equipment. To overcome the issues associated with the in-beam setup, the Imaging Physics Team at the NIRS has developed a novel solution in the form of the dedicated OpenPET system [58], with a geometry that creates a physically opened space where the treatment beam can pass, while maintaining high detector sensitivity. Different versions of the OpenPET system are pictured in Figure A.5.

Figure A.5: Different geometries of the OpenPET system. [58]

The technology that enabled the flexible arrangement of detector modules required for the OpenPET system is the successful concept of so-called depth-of-interaction (DOI) detectors, which were developed to increase sensitivity and spatial resolution. A PET scanner should be designed to utilize a maximum fraction of photons for imaging with respect to the total number of available photons created as a result of the positron emission. This is especially important in order to reduce the amount of radioactive substance injected into the patient and also to lower statistical fluctuations, resulting in a more narrow crystal response in the flood histogram. Each detected event in a PET scanner is associated with a particular crystal of interaction where a photon was absorbed, but does not give the position within the crystal and thus, PET LORs are probability densities of some width. In order to achieve high sensitivity, a crystal large enough for sufficient light yield, i.e. where enough scintillation photons are created, is required while keeping the surface small enough for high spatial resolution. This means that longer crystals are advantageous; however, the reconstruction algorithm assigns the interaction positions over all depths within a crystal to the center front and this causes mis-positioning of the LOR, a so-called parallax error, as illustrated in Figure A.6 (a). Such errors degrade the radial resolution in the peripheral field of view where incoming gammas are not perpendicular to the crystal surface.

A way to reduce the parallax error and improve the spatial resolution of the detector is to supply some depth information to the reconstruction system so that an absorbed gamma ray can be localized to a position along the length of the scintillation crystal element instead of defaulting the location to the center of the crystal. In this way, uniform spatial resolution can be obtained in all directions of the detector. One approach to construct such a DOI detector is to design the crystal as a multi-layered structure, segmented also in the vertical dimension, where each layers has a relative offset of one-half a crystal pitch in X and/or Y-directions as shown in Figure A.6 (b). The Imaging Physics Team at NIRS has created such a structure by inserting reflectors that ensure optical isolation to limit crosstalk between adjacent crystals between every two lines of crystal segments and shifted differently depending on crystal layer [61]. With this detector design, a 3D interaction position \((X, Y, \text{interaction layer})\) of a
All crystals are expressed on one 2D position histogram without overlapping.

Gamma ray within each crystal can be projected onto a 2D position map to generate a flood histogram, as demonstrated in Figure A.6 (c). Areas between reflectors interface by optical coupling material, such as glue, silicon rubber or air. The general crystal arrangement is such that four crystals are covered by a reflector and light sharing between them due to gaps causes a centroid shift of the light distribution so that the flood positions corresponding to crystals in each layer are properly separated.

An additional improvement to reach better image resolution in a PET detector system is to also include time-of-flight (TOF) capability, i.e. the ability to measure the arrival time difference between two photons in order to constrain the emission point to a particular segment along the LOR. Early PET detectors were only able to determine to which LOR an event belongs, but faster photomultipliers and high-density scintillators enabled TOF, which can now be used to improve the SNR by using only the most probable portion of the LOR instead of its entire length, as demonstrated in Figure A.7. Thus, adding TOF capabilities to the detector modules of the OpenPET system could lead to a significant noise reduction. By combining the DOI and TOF technologies, the detector performance can be greatly enhanced in terms of improved accuracy of the estimated annihilation position as the LOR is both narrowed and shortened, making the position uncertainty dramatically decreased.

Figure A.6: (a) Illustration of parallax error for incident photon orthogonal to detector surface (I), with high degree of non-orthogonality with respect to the detector surface for a 1-layer detector (II) and for a 2-layer DOI detector (III) [59]. (b) 4-layer DOI detector module [60]. (c) Principle of 4-layer DOI detector structure showing how the 2D position maps in each layer are superimposed to create the complete flood histogram for the crystal.
A.2. Development of PET-DOI detector with TOF capability

A.2.2 Result Summary

The Imaging Physics Team at NIRS is developing a DOI-TOF detector expected to enhance the performance of the OpenPET system. However, the flood histogram of a 4-layer DOI crystal block suffered from incomplete crystal identification when coupled to a light-sharing TOF-PET detector module produced by Hamamatsu Photonics [62] due to low resolution of the flood histogram created by exposing the module to a gamma source. A correction scheme to improve the resolution of the flood histogram was developed for a 1-layer crystal block coupled to the TOF module for similar application extended to a 4-layer DOI crystal.

The TOF-PET module consists of a Multi-Pixel Photon Counter (MPPC) array with 64 8-bit ADC channels distributed over 4 readout ASICs. The module is coupled to a 3.035 mm}$^2 \times 10$ mm Lutetium Fine Silicate (LFS) scintillator with 256 crystal segments (64 per ASIC, 4 per channel) and a silicon rubber (RTV) light guide is placed in between this crystal array and the photodetector. The setup is displayed in Figure A.8 and was used to identified peaks in the energy spectra for each channel when the module was exposed to gamma radiation from various sources by placing them at a distance of 10 cm from the crystal surface while collecting data. Apart from the background peaks from the lutetium crystal itself of 88 keV, 202 keV and 307 keV, a $^{22}$Na source with expected peaks at 511 keV and 1274.5 keV, as well as a $^{137}$Cs source with a peak at 662 keV, were used to calibrate the module by identifying these peaks.

![Figure A.8: TOF-PET test setup.](image-url)
The difference in response between the ADC channels of the MPPC module due to slight variations in gain was adjusted by performing a \textit{gain correction} in which the energy spectra were aligned to achieve a uniform response. The reference to correct against was determined by computing the mean ADC count value for all channels at each identified energy peak and performing an exponential fit to these mean value data points. In addition to the gain variations, the MPPC response is also non-linear due to a limited number of microcells, which causes saturation at higher energy levels. In order to improve the response linearity and compensate for this effect by artificially extending the dynamic range, a \textit{saturation correction} of the acquired data was performed. This was done by applying a correction to all channels against a linear fit over the lower energy ranges in the gain corrected mean value curve. The correction results are illustrated in Figure A.9 (a). The energy spectrum for one selected MPPC channel before and after corrections is shown in Figure A.9 (b), where the 511 keV photon peak can be distinguished after the corrections are applied. Additional corrections were also made, where events with less than four hits were discarded as the position cannot be accurately reconstructed with fewer number of output signals, and where a more narrow energy window was used in order to suppress noise. A reduction of the effect from background radiation was also done by subtracting background measurement data from the source energy spectra.

![Estimated MPPC Response](image)

Figure A.9: (a) Response curves for the 64 MPPC channels of the TOF-PET module, normalized to the photon energy of 511 keV, and saturation (blue) and gain (red) correction curves. (b) Measured energy spectrum in selected MPPC channel before (upper) and after (lower) corrections applied.

Figure A.10 shows the flood histogram before and after corrections using data from a $^{22}\text{Na}$ source with a 2 mm thick light guide. The crystal identification capability has been improved after corrections were applied, in terms of significantly better separation of the individual crystals peaks, as well as lower noise level and improved response uniformity. An attempt was made to extend the correction scheme to the flood histogram of the 4-layer DOI detector, but this requires a different calibration to be applied to each MPPC array. The energy resolution of the light-sharing modules may not be sufficient for accurate gain correction, or at least makes it complex, but
future improvements are expected by individual calibration of each MPPC channel during separate source scans. The analysis performed during this project, in terms of characterization of the response of the MPPC channels and implemented corrections to achieve uniform gain and saturation compensation to preserve linearity across the entire energy range, is to be used for future development of the DOI-TOF module in order to ensure accurate crystal identification so that detected events can be assigned to a correct crystal of interaction during image acquisition.

Figure A.10: Flood histogram constructed from raw data samples (a) and after corrections applied (b). Magnified areas included for clarity.

A.3 FPGA-based Readout and Data Acquisition System for Characterization of Scintillation Crystals for PET

A.3.1 Background

Pixelation of scintillation crystals is the most common topology for gamma radiation detectors used in PET scanners, as it limits the lateral spread of photons and thereby improves spatial accuracy of the detection in the transverse plane. However, the assembly of crystals into smaller rectangular pieces makes the fabrication process complex and expensive. A promising alternative to reduce costs and complexity of PET detectors is to use sub-surface laser engraving (SSLE) to segment monolithic scintillation crystals using a laser beam. In such techniques, the laser is focused below the surface of the material to create strong local heat accumulation, which causes controlled formation of small localized fractures from the thermal stress. A grid of these kinds of fractures provides the reflective structure required to produce pixels, without mechanically separating the segmented crystals. Figure A.11 shows a pixelated crystal and a crystal fabricated using SSLE. [63]
A.3.2 Result Summary

In order to characterize scintillators segmented using the SSLE technique, a readout system was set up at UC3M to acquire and analyze data from in-house engraved LYSO (Lutetium Yttrium Orthosilicate) crystals. The core of the system is a commercial FPGA-based data acquisition unit from Nutaq called a PicoDigitizer [64], containing 32 125 MHz 14-bit ADC channels on a Virtex-6 FPGA that offers the fast readout speed required for an optimal PET geometry covering maximal surface area, which leads to an increased number of events and high pileup. Xilinx Design Suite was used to program the FPGA using provided IP cores for recording data to memory and retrieving it from memory to a host PC. The readout scheme was optimized to filter out data below a certain threshold, in order to only transfer interesting event data that can be used for image reconstruction to the PC. During acquisition, a gamma source was placed in front of the crystal, which was coupled to an array of 4 silicon photomultipliers (SiPMs). An illustration of the test setup and PicoDigitizer system is shown in Figure A.12.

A GUI was developed in MATLAB® for data readout control and subsequent analysis. The GUI, called the PicoAnalyzer, is displayed in Figure A.13 and allows the user to specify how many samples to acquire, and to construct and save a flood histogram and energy spectrum from the acquired data, as well as one-dimensional X-
and $Y$-profiles and three-dimensional $X$-$Y$ profile in order to study the peak separation. The complete readout and data acquisition system, including the PicoDigitizer unit and PicoAnalyzer GUI, can be used for future characterization and performance evaluation of LYSO crystals segmented with SSLE techniques.

Figure A.13: Matlab-based PicoAnalyzer GUI.
References


References


References


