A University of Sussex PhD thesis

Available online via Sussex Research Online:

http://sro.sussex.ac.uk/

This thesis is protected by copyright which belongs to the author.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Please visit Sussex Research Online for more information and further details
First ATLAS measurement of the effective $B_s^0 \rightarrow \mu^+ \mu^-$ lifetime, and firmware development for Phase-II monitoring and control

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in the Department of Physics, School of Mathematical and Physical Sciences

Author: Ioannis Xiotidis

Supervisor: Prof. Alessandro Cerri

30th March 2022
To my best friend
and eternal support
Nikoleta
ACKNOWLEDGEMENTS

This thesis is summarizing a long effort and plenty of working hours committed to my PhD for the University of Sussex. None of the studies and results would have come into flourish without the continuous support from all the people around me, both academically and personally. For this reason I would like to devote this section to express my gratitude to them.

The first person I would like to express my unlimited and deep gratitude is my supervisor Alex. I know during those four years we had great times but also tough times, times were I felt that we could really achieve everything and times that I felt it's not going anywhere. However, in all those years I always felt your continuous support and push towards me becoming a better person and physicist, hence thank you very much. In addition I would like to thank all the academics at Sussex, Fabrizio, Iacopo, Antonella, Lily and Kate, who embraced me in the group and provided me with any help I needed throughout those years. Furthermore I would like to thank Mark and Kerim who both supported my research even though I never had to work with them directly.

I would like to write a few lines as special thanks to Tom, Benedict and Batool for different reasons. The first one to thank is Tom, for all the endless discussions over beer, the amazing D&D sessions and the infinite hours spent in the lab trying to debug any weird behavior of our boards. The next one is Benedict, whom I am thanking very much for teaching me all the interesting beer flavours, pubs and the interesting discussions about politics. Finally, from the post-docs at Sussex I would like to thank Batool, for the amazing dinners and walks in the city and most importantly for convincing me to stay in academia with my next position at Imperial. Furthermore I would like to express my thankfulness to Umberto and Fabiola with whom I had the pleasure to work on the physics analysis of my PhD.

All the time I spent in Brighton wouldn’t be the same without the PhD students in the office. The first one I’d like to thank is Mario G., who honoured me with the proposal to become his groomsman at his wedding with Olivia, and constantly complained during our Geneva time that my lasagne were not freshly made. Next one to deeply thank is Mario S., who taught me all those nice Sicilian words that for sure will bring me in troubles if I use them in Italy and his wife Lucia. In addition I would like to thank Dani for the company in the office and the random pub excursions for gin and tonics. A special thanks goes also to all the other PhD students, with whom I spend tons of time either in the office or at pubs around Brighton: Fabio, Ondra, EW-Fab, Iker, Marco, Meirin, Giuseppe, min-Fab, Nicola, Emma, Olly.

During my PhD time I had the opportunity to spend plenty of time at CERN. There I had the chance to meet people from all over the world which allowed me to have some very interesting stories to tell. The first and most important person to thank is Sebi-one, for all the amazing time we spent chatting in our balcony in Meyrin, for the endless fifa games and the fantastic dinners at home. In addition I would like to say thank you to Alex and Brianna, which although are from Brighton our friendship started on a Swiss night when trains decided to become a bit more Swiss that normal. Also a special thanks to Alessandra for all the funny Skype chats and discussions during collaboration meetings. A special thanks for various reasons has to go to all
the CERN people I worked with which is impossible to mention without forgetting someone: Naoki (super thanks!), Calliope, Christos, Melissa, Matt, Nikos, Alberto, Lucas, Despina.

At this point, I would like to express my deep gratitude to a certain set of people which I consider to be family. Thanks to my mentor and supervisor in Greece Kostas for the titanic help that made me a member of the CERN family. Thanks to Petros for pushing me and guiding me towards the right decisions. Thanks to Vakis for making me attempt to walk the path I am walking. Thanks to all my friends from high school in Alexandroupolis for the interesting discussions in our hometown and long skype chats: Christos, Alex, Vasilis G., Vasilis T., Kostas, Nikitas, Evgenios, Giotis and Ntinos.

An infinite big thank you to my parents Maria and Vasilis for allowing me to leave my dreams and teaching me how to hold my destiny in my own hands. Thank you to my brother Stefanos for making fun of me, so that I can always keep my feet on the ground, thanks to his wife Vasia for the tsipoura during Christmas celebrations and thanks to mini-Bill for becoming part of our small but interesting family. I would like also to give my special thanks to my grandparents, Stefanos, Foteini, Ioannis and Zoi, which are responsible for the person I am and I miss them a lot.

The last person which I would like to thank, but honestly no word can express it accurately, is Nikoleta. I really feel like I am the luckiest person in the world to be able to have you with me in all this journey. I know that during this time I was not always a pleasant company but having your support gave me the courage to continue and reach the end of the path. Σε ευχαριστώ που φωτίζεις την ζωή μου.

I would like to close this section with the following lines from Elitis describing the best the work of this thesis: Ο κόσμος ο μικρός, ο μεγάς!
STATEMENT

I, Ioannis Xiotidis, hereby declare that this thesis has not been and will not be, submitted in whole or in part to another university for the award of any other degree.

Brighton,
30th March 2022

Ioannis Xiotidis
First ATLAS measurement of the effective $B_s^0 \rightarrow \mu^+\mu^-$ lifetime, and firmware development for Phase-II monitoring and control

by Ioannis Xiotidis

ABSTRACT

This thesis documents a study on the first ATLAS measurement of the $B_s$ meson effective lifetime in the $\mu^+\mu^-$ final state. The result is based on 26.3 fb$^{-1}$ of $\sqrt{s} = 13$ TeV LHC proton-proton collision data collected in 2015 and 2016. The $B_s \rightarrow \mu^+\mu^-$ decay provides a fundamental New Physics phenomena probe in the flavour sector thanks to a very accurate Standard Model prediction for its branching fraction and decay time. The thesis work develops also firmware and hardware tests for the Hardware Tracking for the Trigger (HTT) ATLAS Phase II upgrade.

The physics analysis sections discuss the methodology to measure the $B_s \rightarrow \mu^+\mu^-$ effective lifetime on the same data set used by ATLAS to search for $B_s \rightarrow \mu^+\mu^-$ decays. The effective $B_s \rightarrow \mu^+\mu^-$ lifetime is measured with a binned maximum likelihood fit and applying the “sPlot” technique for background subtraction. The main lifetime systematic uncertainties are estimated, with international collaborators contributing part of the study through data-driven techniques (exploiting $B^+ \rightarrow J/\psi K^+$ decays).

On the firmware development front the work revolves around preparations for the test facility of the UK produced HTT boards. FPGA firmware has been developed to operate legacy ASICs as a training ground in view of the final HTT devices. The same legacy hardware is used to establish IPbus as the monitoring protocol of the HTT system. To this effect, the protocol’s latency and throughput were characterised. Finally, the characterization of the HTT ATCA back-plane has been performed at a line rate of 10Gbps.

The ATLAS $B_s \rightarrow \mu^+\mu^-$ effective lifetime measurement documented in this thesis sets the roadmap for higher statistics ATLAS analyses. The work on the upgrade project contributed to the establishment of the main UK testing facility for HTT hardware.
# Contents

**Introduction** 1

1 Theoretical background 3
   1.1 The Standard Model 3
      1.1.1 Fermions 4
      1.1.2 Bosons 6
      1.1.3 Particle interactions 7
         Quantum Electrodynamics 9
         Weak interactions 10
         Electroweak unification 11
         Higgs mechanism 13
         Quantum Chromodynamics 15
   1.2 Hadrons in the SM 16
      1.2.1 Baryons 17
      1.2.2 Mesons 18
      1.2.3 Quark mixing 19
   1.3 Di-muon $B_s$ decays 20
      1.3.1 Branching fraction and $B$-mixing 22
      1.3.2 Effective lifetime 24
      1.3.3 SM predictions on branching fraction and lifetime 25
   1.4 SM limitations 25
      1.4.1 The hierarchy problem 25
      1.4.2 Neutrino masses 26
      1.4.3 CP violation 26
   1.5 New physics prediction with $B_s \to \mu\mu$ 27

2 Experimental apparatus 29
   2.1 The Large Hadron Collider 29
      2.1.1 Acceleration complex at CERN 30
      2.1.2 Facts and figures about the LHC 32
## 2.2 The ATLAS detector

- **2.2.1 ATLAS coordinates** .......................................................... 34
- **2.2.2 Magnet system** ............................................................... 35
- **2.2.3 Inner detector** ............................................................... 36
  - Insertable B-Layer ................................................................. 37
  - Pixel detector ................................................................. 38
  - SemiConductor tracker ....................................................... 38
  - Transition radiation tracker ................................................. 38
- **2.2.4 Calorimeters** ................................................................. 39
  - Electromagnetic calorimeter ................................................. 40
  - Hadronic calorimeter ....................................................... 40
- **2.2.5 Muon spectrometer** ......................................................... 41
- **2.2.6 Trigger and Data acquisition system** .............................. 42
  - Level-1 trigger ................................................................. 44
  - High Level Trigger ............................................................ 45
  - Detector Control System ................................................... 45

## 2.3 ATLAS Upgrade ................................................................. 45

## 3 The ATLAS Upgrade Project .................................................. 46

- **3.1 Run-2 performance** .......................................................... 46
- **3.2 Phase-I upgrade overview** ................................................ 47
  - **3.2.1 New Small Wheel** ...................................................... 48
  - **3.2.2 Liquid Argon** .......................................................... 48
  - **3.2.3 Muon detectors** ....................................................... 48
  - **3.2.4 TDAQ system** .......................................................... 48
- **3.3 Phase-II motivation** ........................................................ 50
- **3.4 Upgraded sub-systems** ..................................................... 51
  - **3.4.1 Inner Tracker** .......................................................... 51
  - **3.4.2 Calorimeters and Muon system** .................................... 52
  - **3.4.3 The TDAQ system upgrade** ......................................... 52
    - Level-0 ................................................................. 53
    - DAQ system .............................................................. 54
    - Event Filter ............................................................. 55

## 4 The Hardware Tracking for the Trigger System ....................... 57

- **4.1 System motivation and functionality** .................................. 57
- **4.2 Hardware tracking theory** ................................................ 58
  - **4.2.1 Track finding** ........................................................ 58
  - **4.2.2 Track fitting** ........................................................ 59
  - **4.2.3 Common issues in hardware tracking** ............................ 60
- **4.3 HTT Technical description** ............................................... 61
  - **4.3.1 HTT Interface** ....................................................... 61
CONTENTS

4.3.2 Track Processor ................................................. 62
4.3.3 Pattern Recognition Mezzanine ............................... 63
4.3.4 Associative Memory chip ..................................... 64
4.3.5 Track Fitting Mezzanine ..................................... 64
4.4 Performance using simulation .................................. 65
  4.4.1 HTTSim overall use ........................................... 65
  4.4.2 HTT estimated performance ................................. 66
     Pattern-matching ................................................ 66
     Track parameter resolution post first-stage tracking ...... 66
4.5 HTT operation services ......................................... 67

5 IPbus interface ....................................................... 69
  5.1 Protocol overview ............................................... 69
  5.2 Firmware and software components ......................... 70
    5.2.1 IPbus firmware block ................................... 70
    5.2.2 IPbus software component ............................... 71
       ControlHub ..................................................... 71
       $\mu$HAL ....................................................... 71
  5.3 IPbus packet structure ...................................... 72
    5.3.1 Packet header ............................................ 72
    5.3.2 Control packets ......................................... 74
       Transaction Header ......................................... 74
       Read transaction ........................................... 75
       Write transaction .......................................... 76
       Read/Modify/Write transaction ............................ 76
  5.4 IPbus topologies .............................................. 77
  5.5 Performance studies ......................................... 78
  5.6 Application in the HTT ...................................... 80

6 Testing legacy hardware .......................................... 81
  6.1 Sussex HTT contribution ..................................... 81
  6.2 Legacy hardware ............................................... 82
    6.2.1 ATCA shelf .............................................. 82
    6.2.2 Pulsar-IIb ............................................... 84
    6.2.3 Xilinx evaluation cards ................................. 85
    6.2.4 Pattern Recognition Mezzanine 06 ..................... 86
       Associative Memory chips 06 ................................ 87
  6.3 Hardware setup ................................................ 87
    6.3.1 Workbench ............................................... 88
    6.3.2 Crate ..................................................... 88
  6.4 Firmware development for PRM06 .......................... 90
    6.4.1 Carrier card block ..................................... 91
## CONTENTS

6.4.2 Mezzanine card block ........................................... 92  
MAC layer ......................................................... 93  
IPbus core ......................................................... 94  
Inter-Integrated Circuit (I2C) master ............................ 94  
JTAG master ....................................................... 96  
Data preparation logic ............................................ 97  
High speed serial I/F and In-System IBERT ....................... 100  
Simulation ......................................................... 101  

6.5 Lab tests .......................................................... 101  
6.5.1 IPbus latency and throughput measurement ............ 103  
6.5.2 AM thermal tests ............................................. 106  
6.5.3 ATCA back-plane evaluation ......................... 107  
Bit Error Rate theory ........................................... 108  
Back-plane comparisons ........................................ 109  
6.5.4 Integration into HTT online software ...................... 112  
Sussex software .................................................. 113  
Example firmware ............................................... 114  

6.6 Contributions ..................................................... 114  

7 $B_s \to \mu^+ \mu^-$ experimental overview ......................... 116  
7.1 CMS measurements ............................................. 116  
7.1.1 Branching fraction ....................................... 116  
7.1.2 Effective lifetime ......................................... 119  
Two-dimensional unbinned maximum likelihood method .......... 119  
The one-dimensional binned fit on sPlot signal projection ...... 120  
7.2 LHCb result ...................................................... 122  
7.2.1 Branching fraction ....................................... 122  
7.2.2 Effective lifetime ......................................... 124  
7.2.3 Full Run-2 improvements ............................ 125  
7.3 ATLAS result .................................................... 128  
Samples and triggers ............................................ 129  
Analysis strategy ................................................ 131  
Invariant mass fit ............................................. 133  
Result extraction .............................................. 141  
7.3.1 Preparation for full Run-2 ............................ 143  
Trigger studies .................................................. 144  
Current status .................................................. 147  
7.4 Conclusions on $B_{(s)} \to \mu^+ \mu^-$ analyses ................ 150  

8 The $B^0_s$ effective lifetime measurement ....................... 151  
8.1 Analysis strategy ............................................ 151  
8.2 Samples and triggers ....................................... 153
8.2.1 Signal estimation .................................................. 154
8.2.2 BDT cut optimization ............................................. 155
8.3 Mass model .......................................................... 157
8.3.1 Toy sample generation .......................................... 159
8.4 Lifetime fitting metric .............................................. 164
8.4.1 MC template generation ....................................... 166
8.4.2 Choice of the lifetime goodness of fit estimator .......... 167
    Toy-based approach ............................................ 170
    Poisson error .................................................. 172
    Mixed definitions of goodness of fit estimator ............. 172
8.5 Fitter selection and validation with toy MC .................. 173
8.5.1 First $\chi^2$ on the high statistics bins ................. 174
8.5.2 Toy-based $\chi^2$ performance ............................ 175
8.5.3 Poisson error in $\chi^2$ performance ...................... 176
8.5.4 Mixed approaches performance ........................... 177
8.6 Fitter application on control channel ....................... 179
8.7 Fitter expected performance .................................. 181
8.8 Summary for fitter configuration ............................. 183

9 Systematic uncertainties ........................................... 184
9.1 Fit accuracy ....................................................... 184
9.2 Invariant mass fit and decay time distribution assumptions 185
9.2.1 Signal mass model ........................................... 186
9.2.2 Background models ......................................... 187
    Assumptions on background inv. mass models ............ 187
    Decay time distribution and background contamination 188
    Neglected background sources ............................. 190
9.3 MC modeling ..................................................... 191
9.3.1 Control channel discrepancies ........................... 191
    $\eta$ dependence ............................................ 192
9.3.2 Weight effects ............................................... 193
9.4 Heavy-Light state asymmetry ................................. 193
9.5 Combined uncertainty .......................................... 194
9.6 Final result ....................................................... 195
9.6.1 Comparison with other experiments ..................... 197

Conclusions ........................................................... 199

A Analysis samples and selections ................................ 201
A.1 MC samples ....................................................... 201
A.2 Event preselection .............................................. 202
A.3 Simulation samples re-weighting ............................. 204
In particle physics the most accurately tested and successful theory of nature is called the Standard Model (Standard Model (SM)). With the SM physicists are able to describe all structures that define our universe along with the interactions between its fundamental building blocks. In 2012, the final missing piece of this remarkable theory was discovered at CERN. Although the Higgs boson resolved many questions about the particle masses in the universe, there are still many parts missing from the ultimate goal of describing all phenomena in nature. To explain effects that cannot be described with the SM, new theories need to emerge. Any theory/model attempting to explain SM shortcomings is referred to as Beyond the Standard Model (BSM). Examples of BSM theories are supersymmetry (SUSY), dark matter theories, etc. Currently, none of the BSM theories were proofed, allowing searches to become more exciting. Searching for BSM phenomena can be performed with two main approaches. The first approach, called direct searches, looks for excesses from SM predictions in various physics phenomena. The second approach focuses more on probing the effects of new phenomena at lower energies in existing physics processes.

An important branch of particle physics contributing mostly in indirect searches of new physics (NP) lies at the energies of the so called B-sector. More specifically the decays of the $B_s(\bar{B}s)$ and $B_d(\bar{B}d)$ mesons into a muon pair ($\mu^+\mu^-$) are important candidates in the detection of NP. The calculation of the decay probability and the corresponding lifetime of the two mesons shows a great interest from a theoretical and experimental point of view. Both decays are heavily suppressed and well explained in the SM framework proving any observed deviation in their branching fractions an effect of NP. Additionally, the $B_s - \bar{B}_s$ system is characterized by a sizeable difference in the decay widths between the heavy and light mass eigenstates; in the SM only heavy states are allowed to decay, yielding to a very precisely calculated lifetime for the $B_s$ meson. Any experimental observation of the $B_s$ lifetime to have a significant discrepancy with the theoretically calculated heavy state lifetime would be as well an effect of NP phenomena.

The state of the art experimental device searching for NP in the particle physics field is the Large Hadron Collider (LHC) with its four large experiments (ATLAS, CMS, LHCb, ALICE) at CERN. As part of the ATLAS collaboration in this thesis the work described concerns mainly the experimental observation of the effective lifetime in the $B_s \rightarrow \mu^+\mu^-$ decay. CMS and LHCb, have already published their effective lifetime and $Br$ results with the data sets collected in the 2015
and 2016 data taking. In ATLAS the main focus was given to release initially the \( Br \) analysis results on the 2015/2016 data set and subsequently follow with an independent analysis on the effective lifetime on the same data set. In parallel to the lifetime analysis the preparation for the full Run 2 analysis requires to determine the set of triggers to be used.

In 2027 the ATLAS detector is planned to undergo an extensive upgrade mainly due to the increase on the amount of potentially colliding protons (luminosity) per interaction. At this phase the LHC accelerator will move to its ultimate operational capabilities resulting data taking to become a daunting task. The strategy in ATLAS to maintain the data taking rate under control with the best possible quality of data, is to upgrade the whole Trigger and Data Acquisition (TDAQ) system. Part of this upgrade (phase-2) is migrating tracking from a software only task into a hybrid software plus hardware system, called Hardware Tracking for the Trigger (HTT).

The second main focus of this thesis is to describe the characterization of the firmware based monitoring protocol (IPbus) planned to be used in HTT. Furthermore, the preparation of the test facility in Sussex for quality assurance tests on the custom hardware boards assembled in the UK is described.

The outline of the thesis is separated to the following chapters. In chapter 1 a more detailed description of the theoretical background on the SM and the \( Br \) decays is presented. The experimental tools used for the collection of data, mainly the LHC acceleration and a detailed explanation of the current state of the ATLAS detector, is shown in chapter 2. A more detailed description of the phase-2 upgrade planned to be held in 2027 can be found in the chapter 3.

In chapter 4 and 6 a detailed overview of the HTT system along with the facility under development in Sussex is given. The description of the IPbus protocol characterization which concludes the technical part of the thesis can be found in chapter 5.

The main focus of the thesis on the physics front can be found in chapter 7 and 8. Initially a short overview is given for the \( Br \) analysis in ATLAS followed by a more detailed description of the effective lifetime analysis for the 2015/2016 data set. The last chapter shows the extensive studies on the systematic uncertainties along with the final result of the effective lifetime analysis.

To conclude, all the work shown in this thesis was developed within the heavy flavour group (effective lifetime analysis) and the Hardware Tracking for the Trigger group (hardware/lab activities) of the ATLAS collaboration, in which the author contributing and collaborating with physicists and engineers from various countries. The author’s individual work has therefore been often part of a bigger collective effort. A detailed summary of the author’s contributions can be found in Appendix <C>.
THEORETICAL BACKGROUND

This chapter summarizes the theoretical background relevant to this thesis. In more detail an overview of the SM of particle physics will be provided, since it’s the most complete theory of particle physics to date. The next sections will focus mostly on a particular type of particles called hadrons and more precisely B mesons decaying to muon pairs. Finally, the limitation of the SM will be explained and connected with predictions that can be made with the use of di-muon B decays, in the so called Beyond the Standard Model (BSM) theories.

1.1 The Standard Model

The main goal for particle physics is to describe nature with a complete theory that can mathematically formulate and predict events occurring in the universe. The most up to date set of theories providing outstanding accurate predictions are so called renormalisable Quantum Field Theory (QFT) [1] on which the SM is based on. The observed universe is filled with tiny components (elementary particles), called fermions [2]. Fermions interact with each other via the four known fundamental forces which are mediated via a different genre of particles, called bosons [3]. Three out of the four forces can be described by the SM (electromagnetic, weak interaction, strong interaction), and therefore are mediated via bosons. The remaining force (gravity) due to its weak strength at the elementary particle scale is still missing a complete theory under the SM framework. However, gravity is important in the universe since it’s attractive behavior is responsible for the large structures we observe and can be described with general relativity [4].

In the SM all elementary particles are excitations of quantum fields existing in the four-dimensional Minkowski [5] space-time continuum. A categorization of the elementary particles can be achieved based on their properties which get affected by the so called quantum numbers. Quantum numbers are mathematically notated with Poincaré transformations [1]. Separating fermions and bosons originates from the different values of the spin quantum number. Fermions are all particles following the Fermi-Dirac [6] statistics and have a half-integer spin, whereas bosons are following the Bose-Einstein [7] statistics and carry integer spins.
1.1 The Standard Model

1.1.1 Fermions

The first category of particles discussed in this section are the fermions which have half-integer spins. There are currently twelve different fermions discovered to date with a corresponding anti-fermion for each. Fermions interact with each other by mediating bosons. Depending on which interaction they partake fermions can be further classified in sub-categories.

The first category are the so called leptons. Leptons are a set of six particles. Three of them (electrons, muons and tau) interact through both the electromagnetic and weak interaction. The remaining three leptons are called neutrinos and interact only via the weak force. All the leptons interacting with the electromagnetic force carry an electric charge of ±1 [3].

The second category of fermions are called quarks. As some of the leptons, quarks are also interacting through the electromagnetic force, however they contain fractions of an integer charge. All quarks are interacting via the weak interaction and in addition to the leptons they can also interact through the strong force. Being sensitive to the strong force translates to a different type of charge called colour, with the corresponding theory describing those interactions Quantum Chromodynamics (QCD) discussed below.

A summary of all the SM predicted and discovered fermions with their observed properties can be seen in table 1.1 and 1.2. An additional categorization of fermions is the grouping per generation. Generations are couples of fermions, starting from the least heavy to the most heavy ones, which have different flavour numbers. For neutrinos in each category the mass hierarchy does not necessarily assume the same order as for the remaining particles [8]. Particles from the first generation are the particles composing most of the matter seen in the universe [3].

The first categorization for quarks can be based on their electric charge. Quarks with positive

<table>
<thead>
<tr>
<th>Generation</th>
<th>Name</th>
<th>Electric charge</th>
<th>mass/GeV</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td>electron (e⁻)</td>
<td>-1</td>
<td>0.0005</td>
</tr>
<tr>
<td></td>
<td>electron neutrino (νₑ)</td>
<td>0</td>
<td>&lt;10⁻⁹</td>
</tr>
<tr>
<td></td>
<td>muon (μ⁻)</td>
<td>-1</td>
<td>0.106</td>
</tr>
<tr>
<td></td>
<td>muon neutrino (νμ)</td>
<td>0</td>
<td>&lt;10⁻⁹</td>
</tr>
<tr>
<td></td>
<td>tau (τ⁻)</td>
<td>-1</td>
<td>1.78</td>
</tr>
<tr>
<td></td>
<td>tau neutrino (ντ)</td>
<td>0</td>
<td>&lt;10⁻⁹</td>
</tr>
</tbody>
</table>

Table 1.1: All SM predicted leptons categorized in generation based on their mass values [3].

<table>
<thead>
<tr>
<th>Generation</th>
<th>Name</th>
<th>Electric charge</th>
<th>mass/GeV</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td>down (d)</td>
<td>-1/3</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>up (u)</td>
<td>2/3</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>strange (s)</td>
<td>-1/3</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>charm (c)</td>
<td>2/3</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>bottom (b)</td>
<td>-1/3</td>
<td>4.5</td>
</tr>
<tr>
<td></td>
<td>top (t)</td>
<td>2/3</td>
<td>174</td>
</tr>
</tbody>
</table>

Table 1.2: All SM predicted quarks categorized in generations based on their mass values [3]. Listed quark masses are obtained from the re-normalisation scheme [9]. Only the top quark mass quoted here is being obtained from experimental observations [10].
charge are referred as up-type quarks and respectively quarks with negative charge as down-type quarks. The anti-quarks although having flipped electric charge fall in the same category as their mass counterpart.

Fermions in the SM are relativistic and therefore cannot be described with the standard quantum mechanical Schrödinger equation (eq. 1.1) [3].

\[
i \frac{\partial \psi(x, t)}{\partial t} = -\frac{1}{2m} \frac{\partial^2 \psi(x, t)}{\partial x^2} + \hat{V} \psi(x, t), \tag{1.1}\]

where \(\psi(x, t)\) is the time dependent particle wavefunction, \(m\) the non-relativistic particle mass and \(\hat{V}\) the potential energy operator. For this reason the first attempt to describe fermions mathematically was performed with the use of the Klein-Gordon (KG) equation [11]. However, the KG is a second order equation, yielding negative solutions with non-physical negative probability densities [3].

Solution to the problems of the KG equation was given by Dirac with eq. 1.2 which provides a first order solution in both space and time derivatives [3]

\[
\hat{E} \psi = (\alpha \cdot \hat{p} + \beta m) \psi, \tag{1.2}\]

where \(\hat{E}\) is the particle energy, \(\psi\) the relativistic wavefunction, \(\hat{p}\) and \(m\) are the particle four-momenta and mass, and finally \(\alpha, \beta\) indicate matrix constants that are required to full-fill certain criteria discussed below. Since, eq. 1.2 describes a relativistic particle it must also fulfill the Einstein energy-momentum equation which effectively is described by the KG. This requirement translates to strong constraints on the nature of the \(\alpha\) and \(\beta\) constants. The mathematical formulation of the constraint on the constants can be seen in eq. 1.3-1.5

\[
\alpha_x^2 = \alpha_y^2 = \alpha_z^2 = \beta^2 = I \tag{1.3}\]

\[
\alpha_j \beta + \beta \alpha_j = 0 \tag{1.4}\]

\[
\alpha_j \alpha_k + \alpha_k \alpha_j = 0 (j \neq k) \tag{1.5}\]

where \(I\) is the unity matrix. Introducing the constraints to the constants, forces them to be matrices of trace equal to zero, which subsequently forces those matrices to be of even dimension. Adding on the top that the Hamiltonian \((\hat{H}_D = \alpha \cdot \hat{p} + \beta m)\) for any quantum system has to be Hermitian [11], the \(\alpha\) and \(\beta\) matrices need to be Hermitian as well [3]. Expanding eq. 1.2 to the energy-momentum form, generates four mutually anticommuting matrices. If the dimension of \(\alpha_x, \alpha_y, \alpha_z\) and \(\beta\) was \(2 \times 2\) that would allow only three traceless matrices, therefore the dimension must be of \(4 \times 4\). The necessity of four dimensions for the constants in the Dirac Hamiltonian makes the wavefunction \((\psi)\) being a four-component vector called the Dirac spinor [3].

\[
\psi = \begin{pmatrix}
\psi_1 \\
\psi_2 \\
\psi_3 \\
\psi_4
\end{pmatrix} \tag{1.6}
\]
1.1 The Standard Model

The most common form of the Dirac equation however is not the one using the matrices $\alpha$ and $\beta$ but rather its covariant form, seen in eq. 1.7.

\[(i\gamma^\mu \partial_\mu - m)\psi = 0\]  

(1.7)

where $\gamma$ are the Dirac matrices, $\mu$ the Lorentz index showing that the element has four-dimensions and finally as usual $\psi$ the wavefunction and $m$ the mass [12].

The Dirac equation can be easily solved for any half-integer spin free fermion providing elegant solutions containing naturally the spin and magnetic moments [3]. Like the KG, the Dirac equation has negative energy solutions; but unlike KG the probability densities expected are physical and positive. The explanation of the negative energy Dirac spinors is fully compatible with the picture of anti-matter. Therefore, as it was also proved experimentally each fermion seen in table 1.1 and 1.2 has a corresponding anti partner which has the same properties as matter particles but with flipped electric charge.

1.1.2 Bosons

In classical physics forces are transmitted with the use of scalar potentials [1]. However, this picture cannot be applied directly to particle physics due to anomalies that can be created arising from the close distance of interactions [3]. For this reason the use of mediators carrying energy and momentum is required. The mediators of the four fundamental forces are the so called gauge bosons. The SM application of QFT provides an accurate description of each of the forces in the universe except gravity, due to its huge difference in strength at the elementary particle scale. A summary of all the bosons, carrying the fundamental forces, discovered to date can be seen in table 1.3, along with their properties in terms of electric charge, spin and mass [3].

The strength value seen in the table is relative to the strong interaction. For each of those forces a fermion carrying the appropriate charge interacts with the exchange of one of the spin-1 bosons seen above. Attempts to include gravity in the picture are made with the hypothetical graviton being a spin two massless particles since it has to couple with the energy-momentum tensor introduced in general relativity.

In 1964 an additional boson was proposed by P. Higgs [13], which in contrast to the existing carriers of the forces had to be scalar (spin=0) and with no charge. Interaction of all the known particles, except the neutrinos (their mass can be obtained either by including right handed neutrinos in the SM or assuming that neutrinos are Majorana particles [3]), with the Higgs boson generates the discovered particle masses. In addition, also the self-interaction of the Higgs

<table>
<thead>
<tr>
<th>Force</th>
<th>Boson name</th>
<th>Strength</th>
<th>Spin</th>
<th>Mass/GeV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong</td>
<td>Gluon (g)</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Electromagnetism</td>
<td>Photon ($\gamma$)</td>
<td>$10^{-3}$</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Weak</td>
<td>$W^\pm$ and $Z^0$</td>
<td>$10^{-8}$</td>
<td>1</td>
<td>80.4 and 91.2</td>
</tr>
</tbody>
</table>

Table 1.3: List of fundamental forces with their bosons and their spin and mass properties [3].
boson with itself yields a mass for it as well. In 2012 the A Toroidal LHC ApparatuS (ATLAS) and Compact Muon Solenoid (CMS) in a joint announcement, announced the discovery of a Higgs-like particle decaying to four muons with the properties summarized in Table 1.4. Since

<table>
<thead>
<tr>
<th>Boson name</th>
<th>Spin</th>
<th>Mass/GeV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higgs ($H^0$)</td>
<td>0</td>
<td>125</td>
</tr>
</tbody>
</table>

Table 1.4: Higgs boson properties after the discovery at CERN [14].

then more decay channels of the Higgs boson were discovered at European Organization for Nuclear Research (CERN), yielding a better understanding of its properties. The Higgs boson discovery is important for the SM, thus a more detailed discussion on its theoretical properties will be provided in the sections below.

### 1.1.3 Particle interactions

Following the introduction of the free particles in the SM, is the description of the interaction between these particles. Interactions between free particles in the SM happen with the exchange of bosons, only if the appropriate charge is being carried. Studies on the different type of charges which allows fermions to partake in different interactions and therefore be categorized as leptons and quarks yields to three different QFT theories describing each interaction.

If a fermion carries an electric charge then the theory called Quantum Electrodynamics (QED) is used which allows charged particles to exchange virtual photons in any interaction; the word virtual is used to show that the created boson exists only within the principle of uncertainty and thereafter it's energy and momentum are calculated based on the incoming and outgoing particles.

For weak interactions the charge is called isospin and carried by all fermions. The QFT describing the weak interaction uses as mediators of the force the $W^\pm$ for electrically charged interactions and the $Z^0$ boson for neutral. Involving the electric charge into weak interactions yields to the so called Electroweak unification which will be discussed later in the chapter.

A universal property of all QFT’s describing interactions in the SM is the invariance under the gauge group [12], hence its interaction can be described with the symmetry group seen in eq. 1.8.

\[ SU(3)_C \otimes SU(2)_L \otimes U(1)_Y \] (1.8)

Where, $SU(3)_C$ is dedicated to the strong interaction and $SU(2)_L \otimes U(1)_Y$ is referred to the combined electroweak interaction.

Finally, when fermions carry the colour charge they can partake also in the strong interaction. The mediators for strong interactions, named gluons, are described in QCD, are spin one particles with zero mass. There are three different type of colours (red, green, blue) in QCD. Quarks carry a single colour whereas gluons carry two colours simultaneously, to allow interactions between coloured quarks. With this configuration the existence of eight gluon carriers of the force is needed to allow all the colour combinations in interactions.

In general quantum mechanical transitions from one particle state to another are described via
1.1 The Standard Model

a quantity called transition rate $\Gamma_{fi}$, where i is the initial particle state and f the resulting state. The mathematical formulation for the transition rate is based on Fermi’s golden rule [3], seen in eq. 1.9

$$\Gamma_{fi} = 2\pi |T_{fi}|^2 \rho(E_f)$$ (1.9)

where $T_{fi}$ is the transition matrix element and $\rho(E_f)$ the density of allowed states, given a certain final state energy $E_f$. The matrix element contains all the physics properties that are taking part in the interaction whereas the density of states indicates the allowed final states the system can transition to. Expanding the matrix element with the use of Perturbation Theory [15] can be performed showing all possible transitions between the two states.

$$T_{fi} = \langle f | V | i \rangle + \sum_{i \neq j} \frac{\langle j | V | j \rangle \langle j | V | i \rangle}{E_i - E_j} + \ldots$$ (1.10)

In eq. 1.10 the first term shows the interaction of particle with a potential (V) whereas the second term describes the scattering with an intermediate state (j) [3]. The potential for the scattering process is produced by some other particle containing many parameters like instant changes of the field and force applications when the source undergoes a sudden move, which cannot be directly described with the first term. In QFT interactions happen with the exchange of gauge bosons which are responsible for carrying momentum between the two quantum states.

The QFT picture can be understood if time-ordered perturbations are applied [12], which in a two-to-two process $\alpha + \beta \rightarrow c + d$ can be expressed with the following matrix element,

$$T_{fi} = \frac{\langle i | V | j \rangle \langle j | V | i \rangle}{E_i - E_j} = \frac{\langle d | V | X + b \rangle \langle c + X | V | a \rangle}{(E_a + E_b) - (E_c + E_X + E_b)}$$ (1.11)

where $E_{iijjX}$ notes the energy of the incoming, outgoing and mediator particles (ab, cd, X) and the bracket notation shows the transitions from the different quantum mechanical state, which subsequently can be visualized in fig. 1.1.

If Lorentz Invariance is imposed on the matrix element ($M_{fi}$) and all time-ordered diagrams are summed together, the following expression can be produced which shows the interaction matrix element in a two-by-two process.

$$M_{fi} = \frac{g_ag_b}{q^2 - m_X^2}$$ (1.12)

Eq. 1.12 essentially indicates that any interaction depends on the physics applied in each vertex ($g_a, g_b$: interaction strength) and a quantity called the propagator ($1/(q^2 - m_X^2)$) [3].
The only component that makes the calculation of the interaction rates quite challenging is the part of summing over all the potential time-ordered perturbation diagrams. A systematization that helps the calculations was proposed by Feynman with the so called Feynman rules [12]. With the use of those rules schematic representations describe the incoming and outgoing particles as currents carrying momentum and the gauge bosons mediating the interaction, represent the sum of all time-ordered perturbation diagrams.

**Quantum Electrodynamics**

All the Feynman rules to explain interactions with the use of gauge bosons were developed as part of QED formulated by Feynman [16], Schwinger [17] and Tomonaga [18] with the primary goal of describing electromagnetism at the scale of elementary particles.

A free charged particle that fulfills the Dirac eq. 1.7 does conserve U(1) global symmetry however the symmetry is broken when the gauge transformation is local. For this reason to restore gauge invariance an additional term is introduced in the Dirac Hamiltonian to include interactions between any fermion and an electromagnetic field.

\[
\hat{H} = (m \gamma^0 - i \gamma^0 \gamma \cdot \nabla) + q \gamma^0 \gamma^\mu A_\mu \tag{1.13}
\]

The right hand side term is the free Dirac Hamiltonian whereas the additional part describes interactions of free particles with the photon field [3]. Using the QFT formulation eq. 1.13 can be expressed in the form of a Langrangian density where the kinematic properties of the photon are included in the electromagnetic tensor \( (F_{\mu\nu}) \)

\[
\mathcal{L} = \bar{\psi}(i \gamma^\mu D_\mu - m)\psi - \frac{1}{4} F_{\mu\nu}F^{\mu\nu}, \tag{1.14}
\]

where \( \psi \) notes the fermion interacting with the photon field \( (F_{\mu\nu}) \), \( m \) the fermion mass and \( D_\mu = \partial_\mu - i e A_\mu \) is called the covariant derivative. The inclusion of an additional term in both the Hamiltonian and the Landrangian allows the calculations of the QED matrix element in a scatter process \( (1 + 3 \rightarrow 2 + 4) \) where two particle interact and exchange a photon

\[
M = -[q \bar{u}_3(p_3)\gamma^\mu u_1(p_1)] \frac{g_{\mu\nu}}{q^2} [q \bar{u}_4(p_4)\gamma^\nu u_2(p_2)] = -Q_{12} Q_{34} e^2 \frac{j_{12} j_{34}}{q^2}, \tag{1.15}
\]

where \( \bar{u}/u(n) \) indicate the Dirac spinors partaking (n) in the interaction with the photon field, \( Q_n \) shows the charge of the incoming and outgoing fermions and finally \( q^2 \) the Mandelstam variable for the propagator energy. In addition \( j_n \) summarizes the incoming and outgoing particles in the form of currents to obtain a better visual understanding of the Feynman rules for interactions. At the level of the propagator no mass term is present since that would make the Langrangian not gauge invariant anymore and thereafter the photon is a massless particle. If the current formulation from Feynman’s rules is being used then eq. 1.15 gets simplified and yields the Lorentz Invariant description of the electromagnetic interaction [3]. Transforming the incoming and outgoing particles into currents essentially implies that interactions in QED are mathematically equivalent to vector-vector terms, since both the current (spinors) and the propagator are vector objects. Additionally the strength of the interaction can be extracted...
1.1 The Standard Model

which depends on the momentum exchange $q^2$. At low values of $q^2$ the parameter is equal to
the fine structure constant $a = 1/137$.

Weak interactions

QFT theories conserve various symmetries present in the universe. One of those symmetries
concerns the uniformity of space and is called parity. The application of parity in quantum
mechanics appears as an operator which effectively converts any spatial co-ordinate to its op-
posite, as in eq. 1.16.

$$
\psi(x, t) \rightarrow \psi'(x, t) = \hat{P}\psi(x, t) = \psi(-x, t)
$$

When dealing with fermions the mathematical description of parity is expressed with the Dirac
$\gamma^0$ matrix acting on the fermion spinors. Effectively, this yields into particles having a positive
parity eigenvalue and anti-particles a negative [3].

Applying the parity operation in the framework of QED make the fermion currents have an
opposite sign $(j_1 j_2 \hat{P} \rightarrow j_1 j_2) = (j_1 j_2)$ for any interaction. Therefore it can be easily
proven from eq. 1.15 that parity in any QED process is conserved.

$$
M = -Q_{12} \frac{j_{12} j_{34}}{q^2} \frac{\hat{P}}{q^2} M = -Q_{12} \frac{Q_{34}}{q^2} \frac{j_{12} j_{34}}{q^2} = M
$$

However, in 1957 an experiment on measuring the products of a $\beta$-decay, conducted by Wu
[19] proved that parity was not always conserved in weak interaction processes. This discovery
meant that the fermionic currents interacting through the weak force are different than the
fermionic currents interacting through QED.

Introducing parity violation in weak interactions means that the vector representation for a cur-
rent from QED is not sufficient. In general the interaction matrix element is a scalar product
described by a $4 \times 4$ matrix, which is required to be Lorentz Invariant. Those constraints restrict
the number of potential matrices to a number of 5 possible combinations, called bilinear cov-
ariants and give rise to the possible type of currents listed in table 1.5 [3]. Since QED is a vector-

<table>
<thead>
<tr>
<th>Type</th>
<th>Form</th>
<th>Components</th>
<th>Boson spin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scalar</td>
<td>$\bar{\psi}\phi$</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Pseudoscalar</td>
<td>$\bar{\psi}\gamma^5\phi$</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Vector</td>
<td>$\bar{\psi}\gamma^\mu\phi$</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Axial vector</td>
<td>$\bar{\psi}(\gamma^\mu\gamma^5 - \gamma^5\gamma^\mu)\phi$</td>
<td>6</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 1.5: Lorentz invariant bilinear covariants with the corresponding allowed degrees of free-
dom and spin of the carrier boson [3].

vector interaction and conserves parity the next type of current that can be used for the descrip-
tion of the weak fermionic currents is the axial vector-vector interaction. Therefore the weak
charged current can be broken into a vector and axial vector component $j^\mu = g_V j^\mu_V + g_A j^\mu_A$.
With the linear combination of the vector and the axial vector it can be seen that parity is con-
served in the vector-vector and the axial vector-axial vector products. Allowing only the vector-
axial vector product in the matrix element to give raise to parity violations. From experiments
1.1 The Standard Model

It’s known that the weak charged current is a vector-axial vector interaction where the current takes the form showed in eq. 1.18

\[ j^\mu = \frac{g_W}{\sqrt{2}} \bar{u}(p') \frac{1}{2} \gamma^\mu (1 - \gamma^5) u(p) \]  

(1.18)

where \( g_W \) is the weak coupling constant. Following the definition of the weak charged current and according to Feynman’s rules, a propagator term is required to fully describe weak interactions. In contrast to the definition of the spin-1 photon for QED the weak charged current propagator contains an extra degree of freedom due to its mass. Therefore, an extra term in eq. 1.12 needs to be introduced converting the propagator term to eq. 1.19 [3].

\[ -i \frac{q^2 - m_W^2}{q^2 - m_W^2} \left( g_{\mu\nu} - q \gamma^\mu q^\nu \right) \]  

(1.19)

Electroweak unification

As stated also above both QED and weak interaction are QFT theories which are invariant under certain symmetries present in the universe. QED is invariant under the U(1) symmetry with the conserved quantity to be the electric charge. For weak interactions the equivalent quantity that is conserved is the weak isospin \( (I_W) \) associated with the local SU(2) phase symmetry. The SU(2) gauge transformations in QFT are given via the Pauli matrices which are of dimension two. The wave-functions (e.g. particles) need to be coupled in doublets, due to the Pauli matrices dimensions, which have to differ by one unit of charge [12]. For this reason in both leptons and quarks it can be seen that in each generation a doublet of particles exists, like the electron-electron neutrino \((e - \nu_e)\) and up-down quarks. In addition, the SU(2) gauge invariance introduces three possible vector bosons for weak interactions called \( W^{(1)}, W^{(2)} \) and \( W^{(3)} \) [3].

Another important quantity for elementary particles is the so called chirality. Chirality in QFT is introduced by the Dirac \( \gamma^5 = i \gamma^0 \gamma^1 \gamma^2 \gamma^3 \) matrix which depending on how it acts on particles and anti-particles categorizes them as Right-Handed (RH) or Left-Handed (LH) states defined by eq. 1.20.

\[ \psi_{L/R} = \frac{1 \mp \gamma^5}{2} \psi \]  

(1.20)

The weak interaction depends heavily on this separation between the fermions, since it can occur only on LH particles and RH anti-particles, renaming effectively the SU(2) symmetry into \( SU(2)_L \). Since LH fermions are grouped in doublets the theory imposes RH fermions to be organized in singlets in order to be unaffected from the weak interaction. Thereafter, the new interaction term in the Dirac equation takes the following form seen in eq. 1.21

\[ ig_W T_k \gamma^\mu W^k_\mu \phi_L = ig_W \frac{1}{2} \sigma_k \gamma^\mu W^k_\mu \phi_L \]  

(1.21)

where \( \phi_L \) shows the LH chiral particles, \( T_k = \frac{1}{2} \sigma_k \) the local gauge transformation and its formulation with the Pauli matrices. Including in eq. 1.21 the existence of three gauge bosons due to the SU(2) symmetry, three possible currents can arise, for each Pauli matrix [3]. Since
the weak-charged currents impose a change of one electrical unit within the fermion doublets taking part in the interaction, the physical W-bosons can be defined as in eq. 1.22 [1].

\[ W_\mu^\pm = \frac{1}{\sqrt{2}} (W_\mu^{(1)} \mp i W_\mu^{(2)}) \]  

(1.22)

Until this point only the weak charged currents were discussed, which propagate through the \( W^\pm \) spin-1 bosons. The remaining \( W^{(3)} \) field can be subsequently associated with a weak neutral current. However, that would imply that only LH fermion couple with the neutral current propagator, which is directly in contrast with the discoveries from experiments. For this reason the \( W^{(3)} \) generator cannot be associated directly with the real Z-boson discovered in 1973 at CERN [20], which can couple to both LH and RH fermions. A solution to this issue was given by Glashow, Salam and Weinberg [21] who proposed electroweak unification. Combining the \( A_\mu \) photon field with the \( W^{(3)} \) bosonic field, a unified framework can arise which allows the coupling of the physical Z-boson to both LH and RH fermions. For this, a transformation to the U(1) symmetry is required. The new symmetry that arises is called \( U(1)_Y \), where \( Y \) is the hypercharge which is a combination of the electric charge and the third isospin component, seen in eq. 1.23 [12].

\[ Y = 2(Q - T_3) \]  

(1.23)

Introducing the \( U(1)_Y \) symmetry leads to the transformation seen in eq. 1.24 for a given fermion with a wavefunction \( \psi(x) \)

\[ \psi(x) \rightarrow \psi'(x) = \hat{U}(x)\psi(x) = e^{ig'Y\frac{1}{2}\zeta(x)}\psi(x), \]  

(1.24)

where \( Y \) indicates the hypercharge defined in eq. 1.23, \( g' \) the electroweak interaction strength and \( \zeta(x) \) are the generators of the group symmetry (e.g. the gauge bosons). Eq. 1.24 yields an interaction term for the Hamiltonian seen in eq. 1.25

\[ g'\frac{Y}{2}\gamma^\mu B_\mu \psi \]  

(1.25)

which is equivalently formulated as for the QED interaction term, but with the use of a new gauge field \( B_\mu \) that couples to the new hypercharge [3]. In the unified model of QED and weak interactions the physical Z-boson and the photon will be linear combinations of the \( B_\mu \) and \( W^{(3)} \) noted in eq. 1.26 and eq. 1.27

\[ A_\mu = B_\mu \cos\theta_W + W^{(3)}_\mu \sin\theta_W \]  

(1.26)

\[ Z_\mu = -B_\mu \sin\theta_W + W^{(3)}_\mu \cos\theta_W \]  

(1.27)

where \( \theta_W \) is the so called Weinberg angle, with a value precisely measured by experiments and found to be \( \sin^2\theta_W = 0.23129 \pm 5 \cdot 10^{-5} \) [12].

In the current formulation of QED and weak interactions the Lagrangian density describing them has the form seen in eq. 1.28

\[ \mathcal{L} = \sum_{f=l,q} \bar{f} i\gamma^\mu D_\mu f - \frac{1}{4} B_{\mu\nu}B^{\mu\nu} - \frac{1}{4} W^{i}_{\mu\nu} W^{i\mu\nu} \]  

(1.28)
where the sum is running over the three generations of leptons and quarks. No fermions and bosons following the Lagrangian density of the unified electroweak model have any mass. However, this is in clear contradiction with experimental observations. A solution to this problem is discussed in the next sub-section below.

**Higgs mechanism**

A solution to the discrepancy of the theoretical massless particles and the observed massive ones in the electroweak model was given in 1964 by Brout, Englert [22] and Higgs [13]. According to the authors, a spontaneous symmetry breaking mechanism was introduced with the use of complex scalar fields allowing both fermions and bosons to gain mass by interacting with it. All mass terms in the Lagrangian remain invariant under the SU(2)\textsubscript{L} symmetry group. In addition, the mechanism allowed the self interaction of the scalar field, giving raise to a new spin-0 gauge boson called the Higgs. Generally, in QFT Lagrangian powers of higher order on the fields indicate interaction vertices of particles.

The new scalar field needs to be invariant under the SU(2)\textsubscript{L} group, so must be a doublet following eq. 1.29.

\[
\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \phi_1 + i \phi_2 \\ \phi_3 + i \phi_4 \end{pmatrix}
\]  

(1.29)

According to the rules from SU(2)\textsubscript{L}, the upper and lower part of the doublet have to differ by one unit of electric charge. The interaction term in the Lagrangian then can be written as eq. 1.30,

\[
\mathcal{L} = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi)
\]  

(1.30)

where \( D_\mu \) is the covariant derivative defined similarly to the QED case and \( V(\phi) \) the Higgs potential defined in eq. 1.31.

\[
V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2
\]  

(1.31)

The \( \lambda \) parameter in the potential field is bound to be positive if a finite minimum for the field is required. However, for \( \mu^2 \) there is not such a constraint, therefore \( \mu^2 \) can be either positive or negative with different implications on the resulting potential. If \( \mu^2 \) is positive the field has a cone shape with the minimum value at zero. Effectively this would translate into a massive particle with only self-interaction allowed, meaning that all the particles participating in the electroweak model would remain massless. In contrast, if \( \mu^2 \) becomes negative then the shape of the potential assumes a sombrero-like form with a non-negative Vacuum Expectation Value (VEV) [3]:

\[
\phi^\dagger \phi = \frac{1}{2} (\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2) = \frac{\nu^2}{2} = \frac{-\mu^2}{2\lambda}
\]  

(1.32)

As illustrated in fig. 1.2 the solutions on the VEV are degenerate allowing an infinite amount of points. Choosing any of those points would lead to the spontaneous breaking of the symmetry effectively allowing interaction terms between the complex field and the electroweak model’s fermionic and bosonic fields, as well as self-interacting terms for the complex field itself.

The spontaneous symmetry breaking mechanism is an elegant property for the SM, because
1.1 The Standard Model

Figure 1.2: Complex field shape depending on value of $\mu^2$, (a) when the value of $\mu^2$ is positive and (b) when the value of $\mu^2$ is negative. Any of the points at the VEV breaks spontaneously the U(1) symmetry [3].

by applying just perturbations on any of the VEV points an additional gauge boson can be introduced which will be responsible for the particle masses. The photon field is bound to be massless and hence any interaction with the complex field $\phi$ doublet needs to be 0. For this reason the complex field with its perturbations takes the form seen in eq. 1.33.

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}$$

With the current formulation the boson masses in the electroweak model arise naturally by replacing the derivatives with the appropriate covariant derivatives [12]:

$$\partial_\mu \rightarrow D_\mu = \partial_\mu + ig_W T \cdot W_\mu + ig'_Y B_\mu$$

Performing the calculations on the Lagrangian leads to eq. 1.35 for the W, photon and Z masses. As expected the photon mass is 0 and the masses for both W and Z bosons depend on the VEV value.

$$m_W = \frac{1}{2} g_W v \quad \text{and} \quad m_A = 0 \quad \text{and} \quad m_Z = \frac{1}{2} v \sqrt{g_W^2 + g'^2}$$

For the fermionic masses the situation is a bit more complicated in comparison to bosons. However, the Higgs mechanism is able to predict their masses, through the SM lagrangian term:

$$\mathcal{L} = \sum_{f=q,l} \lambda_f (\bar{f}_L \phi f_R + \bar{f}_R \phi f_L)$$

Where $\lambda_f$ is the so called Yukawa coupling which shows the strength of interaction between the fermions and the Higgs field. Substituting eq. 1.36 into eq. 1.29 gives rise to the masses of the fermions:

$$m_f = \lambda_f \frac{v}{2}$$

There were many arguments favoring the Higgs mechanism as the method to generate particle masses in the SM, however without the discovery of the Higgs boson no certain proof was present. All the discussions in the physics community concluded when in 2012 both general
1.1 The Standard Model

purpose experiments at CERN, ATLAS and CMS, announced jointly the discovery of a Higgs like particle with the mass of 125 GeV decaying into a full muon final state. Effectively, the discovery meant that mass production in the SM could be achieved via the elegant way of the spontaneous symmetry breaking mechanism.

Quantum Chromodynamics

The final missing part in the description of the SM is a theory describing the strong interactions. The best theory currently in place is called QCD and as the other parts of the SM, conserves a symmetry of the universe called SU(3):

$$\psi(x) \rightarrow \psi'(x) = e^{ig_s a(x) \hat{T}} \psi(x)$$

(1.38)

where $\hat{T}$ are the 3x3 Gell-Mann matrices [23] used as generators of the symmetry group, and $a(x)$ are eight functions of space-time. Since, the generators are 3x3 dimension matrices an additional degree of freedom arises for the fermions sensitive to the strong interaction which yield to the conserved QCD quantity, called colour, with three possible values, red, green and blue [3].

Applying an SU(3) phase transform in the Dirac eq. 1.39, will maintain invariance if the transformation seen in 1.40 is applied.

$$i\gamma^\mu \left[ \partial_\mu + ig_s G^a_\mu T^a \right] \psi - m \psi = 0$$

(1.39)

$$G^k_\mu \rightarrow G^{k'}_\mu = G^k_\mu - \partial_\mu a_k - g_S f_{ijk} a_i G^j_\mu$$

(1.40)

where $G^k_\mu$ are new massless bosonic fields representing the mediators of the strong interaction and $g_S$ is the interaction strength of the strong force. The final term in eq. 1.40 is needed due to the generators of SU(3) not commuting. This property of SU(3), categorizes effectively QCD as a non-Abelian [24] gauge group, allowing self interactions for the propagators. The terms noted as $f_{ijk}$ are called structure constants of the SU(3) group.

The boson mediating the strong interaction is called the gluon and is in practice a massless spin-1 gauge boson with no electric charge but with two color charges (color-anticolor). To encompass the existence of gluons in the Lagrangian a new term has to be introduced following the same logic as for the photon in QED. The new terms seen in eq. 1.41 need to foresee the self-interacting nature of gluons as per the Dirac equation.

$$\mathcal{L} = \bar{q}_i (i D^\mu \gamma^\mu - m) q_j - \frac{1}{4} F^{a}_{\mu\nu} F^{a}_{\mu\nu}$$

(1.41)

As in QED the strong interaction has a coupling value indicating the interaction strength. The main difference for QCD is that the coupling $\alpha_S$ is very large in comparison to the QED value. The main consequence of this large value is that in QCD perturbation theory cannot be applied and therefore calculations are not straightforward. However, the coupling constant in QCD proves to be energy dependent (eq. 1.42) breaking effectively the theory into two components. The high-energy component, where the value of $\alpha_S$ decreases and where perturbations can be
used as in all other SM theories, and the low-energy component which is handled by lattice QCD.

\[
\alpha_S(q^2) = \frac{\alpha_S(\mu^2)}{1 + \left( \frac{11N_c - 2N_f}{12\pi} \right) \alpha_S(\mu^2) \ln \frac{q^2}{\mu^2}}
\]

The \( N_c = 3 \) parameter corresponds to the number of colours and \( N_f \) are the number of loops present in the interaction. An experimental measurement and trend of \( \alpha_S \) can be seen in fig. 1.3 which shows how the coupling depends on the value of the Mandelstam \([3]\) \( q \). Having the coupling constant being dependent on the energy introduces a property in QCD called asymptotic freedom; this means that quarks and gluons behave like free particles at small distances (high energies) allowing the use of perturbation theory when their properties are calculated. Additionally, another interesting property arising from the large values of the coupling constant at low energies is the so called colour confinement; coloured-charged particles cannot be observed as isolated states.

### 1.2 Hadrons in the SM

Due to colour confinement in QCD free particles that carry colour charge cannot be observed. Therefore, in nature only colourless bound states can be found. Mathematically a colourless bound state behaves like a singlet state although it contains multiple quarks and gluons and thus it can be detected as a free particle in contrast to its components. Those states are called hadrons and contain at least two quarks and gluons. States with two quarks can be constructed only with the inclusion of a quark and an anti-quark and are normally referred as mesons. States with three quark components are called baryons and are allowed to have only either three quarks or three anti-quarks, since any intermediate composition does not yield to a colourless singlet.

In the Large Hadron Collider (LHC) a vast majority of quark - anti-quark pairs are created in
1.2 Hadrons in the SM

Figure 1.4: Combinations of up, down, strange quarks used to form baryons with a 3/2 spin values and various charges. The appropriate terminology in isospin mathematics is called uds baryon decuplet [3].

Each collision. Initially, those pairs will propagate with high velocities in an attempt to separate each other. However, with the constraints from QCD the outgoing pairs cannot break and hence at a certain point it’s preferable to break into two pairs. This process will happen to all the created pairs until compound hadronic states (e.g. mesons/baryons) are formed. The resulting hadrons will point all to a certain direction creating effectively what is called a jet. The direction of the created jet will have the initial direction of the quark starting the process; this process is called hadronization. Normally, in proton-proton collisions a large number of jets are created making jet reconstruction one of the most daunting tasks [25].

1.2.1 Baryons

The most abundant baryons in the universe are the proton and the neutron, that form the nuclei of all observed atoms. Having those two baryonic states being almost equal in mass prompted Heisenberg in 1932 [26] to introduce the concept of isospin, describing the neutron and the proton as the same particle without different excitations of the isospin quantum number. Later on, with the introduction of QCD, isospin symmetries were found to be a subset of the quark colour symmetries [23].

The use of isospin led, for many years, to the organization of the baryons into various structures as part of the allowed isospin values. The isospin is defined with the use of the Pauli matrices seen in eq. 1.43 [3]. The isospin mathematics, directly taken from the spin calculations, allow only $±1$ transition in its values.

$$\hat{T} = \frac{1}{2} \sigma$$  \hspace{1cm} (1.43)

Introducing the spin and charge quantum numbers in a given baryonic state with certain values of the isospin yields the organizations of decuplets and octects seen in fig. 1.4. As mentioned above baryons are the components of all the matter observed in the universe, therefore matter is sometimes referred to as baryonic matter. However, the number of baryons in the universe...
1.2 Hadrons in the SM

Figure 1.5: Combinations of up, down and strange quarks and one of their anti-quark as a function of the various quantum numbers [3]. The quantum numbers used are the sping (S) and charge (Q), creating the different lines.

is conserved. Thereafter any interaction in the SM has to conserve the number of baryons. The bigger problem arising from the conservation of the baryonic number in any process is called baryogenesis; the number of baryons in the universe is larger than the number of anti-baryons. Breaking the symmetry between matter and anti-matter is one of the biggest challenges in modern particle physics.

The three quark baryon states are the dominant states observed, in addition though several discoveries were made of more complex bound states called pentaquarks and tetraquarks [27]. Multi-quark states do not occur naturally in the universe, but their creation in the labs allows for more detailed studies of the strong force.

1.2.2 Mesons

Mesons are quark states with a quark and an anti-quark. Initially a meson, proposed by Yukawa in 1934 [25], was falsely believed to be the carrier of the strong force holding the atomic nuclei in place and not allowing electromagnetism to destroy it. Some of the most known mesons are the pion (π), the kaon (K) and the B-meson which the physics work of this thesis concerns. Mesons are force carriers with spin, in contrast to the baryonic spin, with an integer number following the Bose-Einstein statistics. Initially their discovery was falsely announced (it was the muon) in 1936 by Anderson [12], followed by the real observation of the π meson in 1947 at the University of Bristol [3].

As the baryons, mesons are bound states of quarks and hence the first categorization, before the quark model was based on the isospin algebra. Similarly to baryons, the masses are extremely close to each other, leading to the belief that different mesonic states were excitations of isospin states. Following the same model, fig. 1.5 can arise for mesons as per fig. 1.4 for the baryons.
1.2.3 Quark mixing

Both hadron types (1.2.1, 1.2.2) have as basic components quarks and gluons. Quarks, as the leptons, are fermions with the same rules being expected to govern both charged currents, in terms of the interactions under the weak force. However, from many experiments discrepancies between the two were observed. A key difference for instance is the coupling strength of the weak interaction in leptonic currents. The lepton coupling strength was found to be the same between all the generations of leptons ($G_e = G_\mu = G_\tau$). The same effect is not present in currents representing quarks, since a 5% difference was observed in a up-down current with the corresponding $\mu\nu_\mu$ current [3].

Those discrepancies were initially described by Cabibbo [28] with his hypothesis on the different fermion eigenstates. According to the Cabibbo theorem quarks in weak interactions do have the same coupling strength however no assumption can be made that the mass eigenstates of the system are equivalent to the weak interaction eigenstates. Mathematically, this hypothesis is formulated by allowing the two different eigenstates to mix with a certain mixing matrix seen in eq. 1.44 [3].

$$
\begin{pmatrix}
  d' \\
  s' \\
  b'
\end{pmatrix} =
\begin{pmatrix}
  \cos \theta_C & \sin \theta_C \\
  -\sin \theta_C & \cos \theta_C
\end{pmatrix}
\begin{pmatrix}
  d \\
  s \\
  b
\end{pmatrix} \tag{1.44}
$$

with the primed states being the weak eigenstates on the left and the non-primed the mass eigenstates on the right. Additionally, $\theta_C$ is the Cabibbo angle measured to be $\sin \theta_C \approx 0.225$ [28]. Correcting the quark couplings with the Cabibbo angle produces coupling strengths that are equal, as observed in the leptons. However, accepting the Cabibbo model requires the existence of the charm quark, which at the time was not yet discovered, in order to allow the foreseen strangeness changing neutral currents. Without the additional quark, decays such as $K \to \mu\nu_\mu$ and $K \to \mu^+\mu^-$ would were equally probable. Solution to the disagreement between theory and observation was given by Glashow, Iliopoulos and Maiani in 1970 [29] with the introduction of the Glashow-Iliopoulos-Maiani (GIM) mechanism. According to this mechanism an additional quark exists which belongs to the up-type category of the second generation. The missing charm quark observed simultaneously by two different experiment in its $f/\psi(c\bar{c})$ bound state [30] [31].

Although the Cabibbo matrix (eq. 1.44) described many of the experimental results involving weak interactions with quarks, not all phenomena could be explained. Thus in 1964 Kobayashi and Maskawa proposed an extra quark generation, so to describe the CP violation observed in Kaon decays [32]. With the addition of the extra generation of quarks an extension of the Cabibbo matrix defined in eq. 1.44 was introduced seen in eq. 1.45.

$$
\begin{pmatrix}
  d' \\
  s' \\
  b'
\end{pmatrix} =
\begin{pmatrix}
  V_{ud} & V_{us} & V_{ub} \\
  V_{cd} & V_{cs} & V_{cb} \\
  V_{td} & V_{ts} & V_{tb}
\end{pmatrix}
\begin{pmatrix}
  d \\
  s \\
  b
\end{pmatrix} \tag{1.45}
$$

The resulting mixing matrix is known as Cabibbo-Kobayashi-Maskawa (CKM) matrix. The form seen in eq. 1.45 applies on the down type quark interaction, however an equivalent matrix ex-
1.3 Di-muon $B_s$ decays

exists from up-type quarks. CP violation is being explained by introducing a phase in the transition matrix with a probability analogous to $|V_{ij}|^2$, meaning that $V$ in the matrix are the transition amplitudes between mass and weak eigenstates. As expected diagonal elements concern the same generation of quarks and therefore, aligned with experimental evidence, have higher probability values than the off diagonal terms [3].

To simplify the expected values in the matrix, various methods were developed with the most convenient being the Wolfenstein parameterisation [33], seen in eq. 1.46 which by definition indicates the hierarchy of the amplitude values between generations

$$V_{CKM} = \begin{pmatrix}
1 - \frac{\lambda^2}{2} & \lambda & A\lambda^3(\rho - i\eta) \\
-\lambda & 1 - \frac{\lambda^2}{2} & A\lambda^2 \\
A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1
\end{pmatrix} + \mathcal{O}(\lambda^4), \quad (1.46)
$$

the Wolfenstein parameters are all real numbers with the following definitions: $\lambda = \sin \theta_C$, $\rho = \text{Re}\left\{\frac{s_{13}e^{i\delta}}{\sqrt{s_{12}s_{23}}}\right\}$, $\eta = -\text{Im}\left\{\frac{s_{13}e^{i\delta}}{\sqrt{s_{12}s_{23}}}\right\}$ and $A = \frac{s_{23}}{s_{13}}$, where $s_{nm}$ are the real parts of the CKM elements.

Measuring the values of the CKM can happen through a large number of decays, since its effect is present in all weak interactions containing quarks. The currently calculated values from the various channels can be seen in eq. 1.47 [34].

$$V_{CKM} = \begin{pmatrix}
0.97446 \pm 0.00010 & 0.22452 \pm 0.00044 & 0.00365 \pm 0.00012 \\
0.22438 \pm 0.00044 & 0.97359 \pm 0.00010 & 0.04214 \pm 0.00076 \\
0.00896 \pm 0.00024 & 0.04133 \pm 0.00074 & 0.999105 \pm 0.000032
\end{pmatrix} \quad (1.47)
$$

With the CKM matrix, it was found that in the SM only weak charged currents are allowed to change the flavour at the lowest Feynman diagram order (tree-level). To achieve a change in the flavour for neutral currents higher order diagrams have to be used and thereafter are naturally suppressed.

The CKM provides a mechanism for CP-violation within the SM (with the $\delta$ complex term in its elements), however this is not sufficient to described the overall CP-violation observed in the universe. For this reason further theories have to be developed, some of those are discussed in the sections below.

1.3 Di-muon $B_s$ decays

A set of particular mesons and their decays to muon pairs, which this thesis physics contribution concerns the most, are the $B^0_s \rightarrow \mu^+\mu^-$ and $B^0_d \rightarrow \mu^+\mu^-$. Both mesons are composed of an anti-b quark and either a strange or down quark effectively giving those two mesons similar masses, as seen in table 1.6. In the SM these particles are foreseen to have a finite decay length of the $\mathcal{O}(10^{-12} \text{ ps})$, which allows them to travel short distances after their production vertex and before their decay into the muon pair. The decay of both mesons in a muon pair is a heavily suppressed process in the SM, sensitive though to BSM physics. The suppression of these decays is driven by three reasons yielding the theoretical predicted values to be $\mathcal{O}(10^{-9})$ and $\mathcal{O}(10^{-10})$ respectively for $B_s$ and $B_d$ [3].
### 1.3 Di-muon $B_s$ decays

<table>
<thead>
<tr>
<th>Name</th>
<th>Mass (MeV)</th>
<th>Electric charge</th>
<th>Valence quark composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B^0_s$</td>
<td>$5366.88 \pm 0.14$</td>
<td>0</td>
<td>$b\bar{s}$</td>
</tr>
<tr>
<td>$B^0_d$</td>
<td>$5279.58 \pm 0.17$</td>
<td>0</td>
<td>$\bar{b}d$</td>
</tr>
</tbody>
</table>

Table 1.6: Summary table of $B_s$ and $B_d$ properties [10].

The first main reason for the suppression of those decays is arising from the quark flavours composing those states. The lowest and direct way to achieve a transition from a quark neutral current to a lepton neutral current is via the exchange of a $Z$-boson. However in SM flavour changing neutral currents are not allowed, as discussed above, and thereafter higher order diagrams need to be employed. Those higher order diagrams seen in fig. 1.6 can happen only with the exchange of $W$, $Z$ and $H$ (small contribution in comparison to the other two) bosons in box form or penguin style [35]. The lack of tree level direct transitions is the main culprit for the suppression of the $B_{s/d}$ decays over other more preferable decay channels.

The second main reason for the high suppression factor of the $B_{s/d}$ di-muon decays originates from the CKM matrix seen in eq. 1.45. As it can be seen in fig. 1.6 the quarks appearing within the loops foresee transitions of the strange quark into either an up or a charm or a top quark. Given the CKM the $s \rightarrow t$ transitions are much more favorable than the others. For the $B_d$ the transitions that need to happen are by definition between the first and third quark generation thereafter leading to a much more suppressed Branching Fraction (BR) than the $B_s$ decay, although the kinematics are essentially the same.

The last remaining source of suppression arises from the different helicity states [3] present in those decays. The two mesons according to table 1.5 are categorized as pseudo-scalar states meaning that the helicity quantum number is forced to be the same in the final state. Since the di-muon decays happen via the weak currents the weak constraint for left-handed chirality final states appears in the decay. In the limit of massless particles, negative helicity states corre-
1.3 Di-muon $B_s$ decays

pond to left-handed chirality states and vice-versa. However, the muons are massive particles with a mass much smaller than the $B_s$ mass, making the decay possible under the weak interaction chirality constraint only with a very large suppression factor, and with a preference for certain muon helicity states.

As mentioned both decays can happen in the SM only via higher order Feynman diagrams, making both sensitive to New Physics (NP) phenomena that can take place within the loops. Those effects can appear in two forms for the di-muon B decays, the first one is by modifying the value of the branching fraction foreseen in the SM and the second one is by changing the effective lifetime of the decay. For those reasons in the following two sections the theoretical calculation of the branching fraction and the lifetime of the $B_s$ is provided with as much as possible model independence.

1.3.1 Branching fraction and B-mixing

A very useful observable that is sensitive to NP in the $B_s^0 \rightarrow \mu^+ \mu^-$ system is the branching fraction. The definition of a particle branching fraction is the number of decays in a specific decay channel over the total number of $B_s$ decaying to all possible channels. The various $B_s$ decay branching fraction definitions tend to be challenging due to the B-meson properties.

One of those properties concerns the quark mixing introduced by the CKM matrix. The $B_s$ are always produced in the $B_s^0 \rightarrow \bar{B}_s^0$, with $\bar{B}_s$ the anti-B meson, which at $t = 0$ are described with the use of the $|B_s(0)\rangle$ and $|\bar{B}_s(0)\rangle$. As time passes those states can oscillate between their mass and weak eigenstates, changing from one to the other, as it can be seen in fig. 1.7.

The description of this system in the frame of quantum mechanics can be achieved with the use of the time-dependent Schrödinger equation \[\text{[25]}\], in eq. 1.48. From now on in order to simplify the notation of this document the $B_s$ notation will be used, but the same calculations

\[\text{Figure 1.7: Oscillations of the matter ($B_s^0$) state to the anti-matter ($\bar{B}_s^0(b\bar{s})$) state with the exchange of W bosons. Where both states indicated are showing the flavour states of the meson - anti-meson mixing.}\]
A convenient way to define the decay rates, in order to simplify the calculations, is with the light mass eigenstates. The $R_i$ seen in eq. 1.53 [38], expressed as a function of the mass states, which have very well defined lifetimes, and can be using the definition of the heavy and light eigenstates seen in eq. 1.50 the decay rate can be
\[
\tau = (M - i \frac{T}{2}) \begin{pmatrix} B_0^0 \\ B_0^+ \\ B_0^- \end{pmatrix}
\]

(1.48)
The M and T appearing in the eq. 1.48 are hermitian matrices describing the mass and lifetime of the system. Both matrices have non-zero off-diagonal elements allowing for positive times to have both particle and anti-particle states superimposed. The superposition of the two particles is described with the heavy and light mass eigenstates which are at $t = 0$:
\[
|B_H\rangle = p|B^0_s\rangle - q|\bar{B}^0_s\rangle \quad \text{and} \quad |B_L\rangle = p|B^0_s\rangle + q|\bar{B}^0_s\rangle
\]

(1.49)
with the eigenvalues $(m_{H,L} - i\Gamma_{H,L}/2)$ and the coefficients p and q being constrained by $|p|^2 + |q|^2 = 1$. The time evolution in terms of the heavy and light states is then:
\[
|B_H(t)\rangle = |B_H\rangle \exp\left\{-i(m_H - i \frac{\Gamma_H}{2}) t\right\} \quad \text{and} \quad |B_L(t)\rangle = |B_L\rangle \exp\left\{-i(m_L - i \frac{\Gamma_L}{2}) t\right\}
\]

(1.50)
In pp collisions at the LHC only the time-integrated $B_s \rightarrow \mu^+\mu^-$ branching fractions can be measured without considering the flavour states of the B-mesons [36]. The eq. 1.51 shows the experimentally measured $\mathcal{B}(B_s \rightarrow \mu^+\mu^-)$ [37], where with $\mathcal{B}$ the BR is noted.
\[
\mathcal{B}(B^0_q \rightarrow \mu^+\mu^-)_{\text{exp}} = \frac{1}{2} \int_0^\infty \left| \langle \Gamma(B^0_q \rightarrow \mu^+\mu^-) + \Gamma(\bar{B}^0_q \rightarrow \mu^+\mu^-) \rangle dt \right| \frac{1}{2} \int_0^\infty \langle \Gamma(B^0_q(t) \rightarrow \mu^+\mu^-) \rangle dt
\]

(1.51)
The $\langle \Gamma(B^0_q(t) \rightarrow \mu^+\mu^-) \rangle$ is called the time-dependent decay rate, without any assumption on the flavour. In the SM the theoretical calculations on the branching fraction are performed on the time-independent equation, yielding a discrepancy between the experimentally measured quantity and the theoretically calculated prediction [36]. The theoretical prediction is considering the CP averaged (independent of the evaluation of the system over time) decay rate which can be seen in eq. 1.52 [37],
\[
\mathcal{B}(B^0_q \rightarrow \mu^+\mu^-)_{\text{theo}} = \frac{\tau_{B^0_q}}{2} \langle \Gamma(B^0_q(t) \rightarrow \mu^+\mu^-) \rangle_{t=0}
\]

(1.52)
with $\tau_{B^0_q}$ being the average lifetime of a flavour independent B-meson decaying to a muon pair.
Using the definition of the heavy and light eigenstates seen in eq. 1.50 the decay rate can be expressed as a function of the mass states, which have very well defined lifetimes, and can be seen in eq. 1.53 [38].
\[
\langle \Gamma(B^0_q(t) \rightarrow \mu^+\mu^-) \rangle = R_H^{\mu^+\mu^-} \exp\left\{-\Gamma_H^q t\right\} + R_L^{\mu^+\mu^-} \exp\left\{-\Gamma_L^q t\right\}
\]

(1.53)
The $R_{H/L}^{\mu^+\mu^-}$ parameters appearing in eq. 1.53 are the corresponding decay rates of the heavy and light mass eigenstates.
A convenient way to define the decay rates, in order to simplify the calculations, is with the definition of the $y_q$ and $\Gamma_q$ variables seen in eq. 1.54 [37].
\[
y_q = \frac{\Gamma_q^q - \Gamma_H^q}{2\Gamma_q} = \frac{\Delta\Gamma_q}{2\Gamma_q} \quad \text{and} \quad \Gamma_q = \tau_B^{-1} = \frac{\Gamma_L^q + \Gamma_H^q}{2}
\]

(1.54)
Another important reason for introducing the variables \((y_q, \Gamma_q)\) in eq. 1.54 is because they are experimentally measured quantities [36]. Substituting eq. 1.54 into eq. 1.53 yields to eq. 1.55 [37]

\[
\langle \Gamma(B_q^0 \rightarrow \mu^+\mu^-) \rangle = (R_H^{\mu^+\mu^-} + R_L^{\mu^+\mu^-}) \exp \left\{-y_q t \frac{\cosh \Gamma_q t}{\tau_{B_q^0}} + \Delta y^{\mu^+\mu^-} \sinh y_q t \frac{\Gamma_q t}{\tau_{B_q^0}} \right\}
\]

(1.55)

with \(\Delta y^{\mu^+\mu^-}\) being defined as

\[
\Delta y^{\mu^+\mu^-} = \frac{R_H^{\mu^+\mu^-} - R_L^{\mu^+\mu^-}}{R_H^{\mu^+\mu^-} + R_L^{\mu^+\mu^-}}
\]

(1.56)

The result from eq. 1.55 can be used in order to derive a correction factor on the experimentally measured quantity seen in eq. 1.51 to allow comparisons with the theoretically predicted value. The tedious calculations are beyond the scope of this thesis and therefore are not presented, however the result can be seen in eq. 1.57 [36].

\[
\mathcal{R}(B_q^0 \rightarrow \mu^+\mu^-)_{\text{theo}} = \left[ \frac{1 - y_q^2}{1 + y_q \Delta y^{\mu^+\mu^-}} \right] \mathcal{R}(B_q^0 \rightarrow \mu^+\mu^-)_{\text{exp}}
\]

(1.57)

In the case where the value of \(y_q\) is equal to zero the two quantities are effectively the same. This was observed in the case of the \(B_d\) meson, however it was proven experimentally that for the \(B_s\) there is a sizeable difference. In the SM the value of \(\Delta y^{\mu^+\mu^-}\) is foreseen to be equal to +1 which means that only the \(B_H\) states are allowed to decay since the \(\mu^+\mu^-\) final state is CP odd. However, measuring a different value than the prediction can lead to significant discrepancies between the value predicted by the SM for \(\mathcal{R}(B_s^0 \rightarrow \mu^+\mu^-)\) and the decay time, since both values depend directly on the value of the \(\Delta y^{\mu^+\mu^-}\) parameter [36]. Physically that would be translated that heavier particles (not foreseen by the SM) appear within the \(B_0^s\) loop alternating the observable properties of the decay, in the same fashion as the output electron angles in an electron-electron scattering without the existence of the Z-boson [3].

### 1.3.2 Effective lifetime

As mentioned above measuring the \(\mathcal{R}(B_s)\) can provide access to the value of the \(\Delta y^{\mu^+\mu^-}\) parameter. An alternative method probing NP concerns the measurement of the effective lifetime of the \(B_s\) meson to a muon pair. For \(B_s \rightarrow \mu^+\mu^-\) the effective lifetime can be defined as [39]:

\[
\tau_{\mu^+\mu^-} = \frac{\int_0^\infty t \langle \Gamma(B_q^0 \rightarrow \mu^+\mu^-) \rangle \, dt}{\int_0^\infty \langle \Gamma(B_q^0 \rightarrow \mu^+\mu^-) \rangle \, dt}
\]

(1.58)

and by substituting eq. 1.55.

\[
\tau_{\mu^+\mu^-} = \tau_{B_s} \left( \frac{1 + 2 \Delta y^{\mu^+\mu^-} y_s + y_s^2}{1 + \Delta y^{\mu^+\mu^-} y_s} \right)
\]

(1.59)

As mentioned above if the SM prediction is correct only the heavy states are allowed to decay therefore \(\tau_{\mu^+\mu^-}\) will be equal to \(\tau_H\). Any discrepancy measured in the effective lifetime will
be due to NP phenomena. The main physics focus of this thesis concerns the measurement of the $B_s \rightarrow \mu^+\mu^-$ effective lifetime with the ATLAS detector in the data set collected during 2015/2016. Due to the importance of the effective lifetime measurement the ATLAS collaboration aims for a publication. Since many of the tools used in the lifetime analysis are shared with the already published branching fraction analysis [40], the publication is considered a follow up.

### 1.3.3 SM predictions on branching fraction and lifetime

In the framework of the SM precise values of both the $\mathcal{B}(B^0_s \rightarrow \mu^+\mu^-)$ and the effective lifetime $\tau_{\mu^+\mu^-}$ are provided. For the branching fraction the calculation includes all the recent progress performed in the lattice QCD field yielding to the following values [41].

\[
\mathcal{B}(B^0_s \rightarrow \mu^+\mu^-) = (3.10 \pm 0.17) \times 10^{-9} \quad (1.60)
\]

\[
\mathcal{B}(B^0_d \rightarrow \mu^+\mu^-) = (1.6^{+1.6}_{-1.4}) \times 10^{-10} \quad (1.61)
\]

The largest contributions in the uncertainties seen in eq. 1.60 and eq. 1.61 arise from the CKM contributions and decay constants [34].

For the $B_s \rightarrow \mu^+\mu^-$ effective lifetime the predicted value from the SM assumes that $\mathcal{A}_{\mu^+\mu^-}$ is equal to +1 and in addition that the value of the mean $\tau_{B_s}$ the expected value seen in eq. 1.62 can be obtained.

\[
\tau_{\mu^+\mu^-} = 1.661 \pm 0.032 \quad ps \quad (1.62)
\]

The spectrum of allowed values is determined by the value of $\mathcal{A}_{\mu^+\mu^-}$ which is allowed to vary between -1 and +1, with any deviation from +1 hinting at NP. A $5\sigma$ measurement for the effective lifetime is still to be performed making the analysis an interesting task.

### 1.4 SM limitations

The SM was proven to be a remarkably accurate theory being able to provide accurate predictions to many of the experimentally observed quantities in particle physics. However, as a theory the SM is far from being perfect and explaining all the phenomena observed. There are a set of limitations where in some the effective lifetime measurement of the $B^0_s \rightarrow \mu^+\mu^-$ can provide useful information. A short summary of such limitations is provided in this chapter, where some of them are more correlated with the effective lifetime than others.

#### 1.4.1 The hierarchy problem

The first major limitation of the SM arises from the Higgs boson itself [14]. According to the theory the Higgs mass is defined as $m_H^2 = m_{H_{0}}^2 + \delta m_H^2$ where the first term is the so-called bare mass of the Higgs boson arising naturally from the mass term in the Lagrangian. The second
part concerns the radiative correction which are produced from the interaction of the Higgs boson with all the other particles. An example of such interactions can be seen in fig. 1.8 which shows the fermion loops that can appear in the Higgs propagator. Incorporating such corrections with fermions yields a value of the radiative corrections on the Higgs mass which is way larger than the value observed experimentally at CERN. The calculation of the radiative correction can be seen in eq. 1.63

$$\delta m^2_{H|f} = -\frac{|\lambda_f|}{8\pi} \Lambda^2_{UV} + \cdots$$

(1.63)

where the value of $\Lambda^2_{UV}$ was set to the Planck scale ($\approx 10^{18}$ GeV) and is called the ultraviolet momentum cut-off.

The large discrepancy between the experimental observation and the radiative correction term is what is called the Higgs hierarchy problem. In the SM solving the issue is achieved with the so called fine-tuning process of the bare mass in order to cancel the large value arising from eq. 1.63. Fine-tuning the bare mass is not forbidden however it comes in open contrast with the naturaleness argument of QFT which allows such approaches only if the theories have very specific features.

### 1.4.2 Neutrino masses

In the discussion of the Higgs mechanism being the only process to provide masses to particles foreseen in the SM the neutrino seemed to be a massless particle, in the lepton doublets. However, recently the Super-Kamiokande [42] and the SNO [43] collaborations have made a discovery that neutrinos could oscillate their flavour states. The only process allowing this behavior foresees that neutrinos have masses and must oscillate between mass and flavour states as the quarks. Allowing, the neutrinos to be massive is directly in contrast with the SM prediction of massless neutrinos and lepton flavour universality.

### 1.4.3 CP violation

Another big mystery of the universe that cannot be explained by the SM are the sources of CP violation. Within the SM the only place that CP violation is allowed concerns the phase shift appearing in the CKM matrix (shown in sec. 1.2.3). Having this only source is not sufficient to describe the matter anti-matter imbalance observed and therefore more sources are required. A potential solution could arise from the masses observed in the neutrinos or via QCD. As
previously discussed neutrinos obtain mass either with the known Higgs mechanism (requires RH neutrinos) or via different type of mechanisms which imply that neutrino and anti-neutrino are the same particles (Majorana). In QCD CP violating phenomena are not forbidden, however none were observed.

1.5 New physics prediction with $B_s \rightarrow \mu\mu$

A large number of theories exists which use the $B_s \rightarrow \mu^+\mu^-$ process to describe BSM scenarios solving the open issues of the SM that were discussed above. A small subset of those theories will be discussed briefly in this section to indicate the importance of precision measurements in the current experiments.

An overview of such theories concerning the branching fraction measurement can be seen in fig. 1.9 with the most known discussed in more detail below, indicating that, depending on the used theory model, different measurements of $B_{(s)}$ branching fractions are expected. A more detailed discussion about the different ways that both the effective lifetime and branching fraction of $B_{(s)} \rightarrow \mu^+\mu^-$ contributes to NP theories can be found in the following ref. [49],[50],[51],[52]. However, below some of the most common contributions are discussed, here.

The most popular BSM theory is concerning both the branching fraction and the lifetime are the so called Minimal Flavour Violation (MFV) [44] theories. According to the MFV theories all the operators contributing to the $B_s \rightarrow \mu^+\mu^-$ decays are SM-like and NP phenomena can arise from the introduction of extra "Wilson" coefficients [38]. Additionally, many of the coefficients predict variations in the $\Delta \Gamma_{\mu^+\mu^-}$ parameter, which directly affect both the lifetime and the branching fraction, predicted by the SM.
Other commonly known theories for BSM predict the existence of new particles that can appear in the $B_{(s)}$ to di-muons Feynman diagrams, hence affecting the $\Delta \Gamma_{\mu\mu}$ parameter. Those particles can appear in the decay channel either via closed loops like 1.6 or by allowing flavour changing neutral currents at tree level. Examples of such theories are the Higgs doublet models [53], SUSY models [45], and the leptoquark models [54].

The first of the three examples modifies the SM Higgs boson field, by introducing two complex fields where both have a non-zero VEV. The immediate consequence of this modification is that during the spontaneous symmetry breaking mechanism five new particles arise (2 charged scalars, 2 neutral scalars and 1 pseudoscalar). All of those new particles are allowed to enter the $B_{(s)}$ loops or undergo tree level flavour changing neutral currents, affecting both the lifetime and branching fractions.

In SUSY models the philosophy remains the same, where the existence of the SUSY particles affects the $B_{(s)}$ decays. In many SUSY models the mass of the particles are beyond the currently achievable energy of the accelerators, meaning that precision measurements might give a hint at the energy scale of those particles. Minimal SUSY models foresee modifications on the Higgs boson equivalent to those of the Higgs doublet theories.

Finally, the leptoquark models are currently highly popular and interesting. Those models predict extra particles which carry both leptonic and baryonic quantum numbers, allowing flavour changing neutral currents at tree level. The exact description of the properties of leptoquark particles depends on the couplings they form with SM particles.

The current state of the art experiments have not hinted any NP phenomena in the di-muon decays of the $B_{(s)}$ mesons. However, from the current experimental results on both the branching fraction and effective lifetime, plenty of exploration room is left. Therefore, experiments like ATLAS at CERN are needed to explore all the aspects of the SM, and shed light on the remaining unanswered questions. The details of the ATLAS detector will be discussed in the following chapter, along with its scheduled upgrade programs.
EXPERIMENTAL APPARATUS

All the studies that are presented in this thesis are in the context of CERN and more precisely the ATLAS experiment. In this following chapter I describe the detector used in order to perform all the studies in this thesis. The CERN facilities are described, followed by a detailed explanation of all the sub-systems of the detector. Finally, the data acquisition system is described.

2.1 The Large Hadron Collider

Some of the most powerful machines used in particle physics to explore NP phenomena are particle accelerators. Currently, the largest machine is the LHC [55] that was built at the CERN between 1998 and 2008. The location of the LHC is a tunnel that crosses the Swiss-French border adjacent to Geneva. The 50 to 175 meters subsurface tunnel with a circumference of 27km was used to host the previous largest accelerator in the world, the Large Electron Positron (LEP) collider. The LHC is accelerating two opposite direction hadron (proton) beams currently up to 13 TeV in energy. The main reason why proton-proton interactions were chosen was due to the higher beam concentration (luminosity) that it can be achieved rather than with proton and anti-protons. To achieve this high energy, 1624 super-conducting magnets are placed around the beam pipes. The main purpose of the LHC magnet system is to bend the beams with the help of dipole magnets and to focus with the use of quadrupole and other type of magnets. One of the main challenges for the magnet operation is to keep the temperature below 1.7 Kelvin in the super-conductivity regime of the used super-fluid helium to withstand the intense magnet currents necessary to produce an 8.3 Tesla magnetic field.

The LHC started its first operation in 2008 when the first record of many was broken, for the highest energy ever achieved (1.17 TeV for 1h). Since the LHC is a complex machine with many factors affecting its performance a staged operation system was set in place. The operation was split into so called runs. The main purpose of the runs is to allow maintenance of the machine along with foreseeable upgrades that will push the LHC to its best operational limits and give access to energy ranges never probed before. The vast majority of upgrades and maintenance happen in the periods between the runs which are called Long Shutdown (LS).

In 2013, the LHC concluded successfully its first run (Run-1) and moved to the first major up-
2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is the main discovery machine at CERN, its operation is dependent on a chain of smaller accelerators that are present at CERN. Protons extracted from hydrogen atoms are accelerated through this chain in order to reach the LHC ring and be finally brought to 13-14 TeV. Fig. 2.2 illustrates the layout of the CERN accelerator complex described below.

Although the LHC is the main discovery machine at CERN, its operation is dependent on a chain of smaller accelerators that are present at CERN. Protons extracted from hydrogen atoms are accelerated through this chain in order to reach the LHC ring and be finally brought to 13-14 TeV. Fig. 2.2 illustrates the layout of the CERN accelerator complex described below.

The beams at CERN originate from a bottle of hydrogen ($H_2$) atoms which are then stripped from their electrons and fed to the first acceleration step, called the Linear Accelerator (Linac) 2. Linac2 is one of the oldest accelerator machines at CERN dating its construction back in 1978. Protons entering Linac2 reach 50 MeV energy and gain about 5% in mass due to special relativity (when particles reach energies close to the speed of light then the $E = mc^2$ formula applies). Following Linac2 the protons enter into the Proton Synchrotron (PS) booster reaching an energy of 1.4 GeV in their first circular accelerator machine. The next step into the journey towards the LHC is the Super Proton Synchrotron (SPS), which is the second-largest accelerator currently operating at CERN. At the SPS the protons reach an energy of 450 GeV and are separated into "bunches": particle packets along the beam direction, spaced by 25 ns. The final step of acceleration is the LHC where the protons from SPS get split into two beams circulating the LHC in opposite directions. Once the beams reach their maximum energy collisions are...
2.1 The Large Hadron Collider

Figure 2.2: CERN Accelerator complex and LHC experiments. The LHC is the largest circle (dark grey) with the four major experiments noted with yellow dots. The smaller circles are the accelerators used to inject beams at the LHC input energy [57].

allowed in the four interaction points, where the LHC experiments lie.

Another, operation mode of the LHC is the acceleration of heavy ions instead of protons, mainly fully-stripped lead ions $^{208}_{82}$Pb$^{82+}$. The acceleration procedure for heavy ions is very similar to the protons, with some small variations. Instead of sending the heavy ions through the Linac2 path they are instead injected into Linac3, where the first acceleration occurs along with stripping the electrons. The resulting nuclei are then injected as long pulses into the Low Energy Ion Ring (LEIR), where the shortening into bunches is happening as well as the acceleration from 4.2 MeV to 72 MeV. From this point onward the path to the LHC coincides with the protons. The heavy ion bunches are sent to PS and SPS in order to reach the LHC.

Mainly the use of the LHC is explored at the four interaction points where the four big detectors are located. In each of those points on the LHC ring large detectors were built to study the collisions. The two largest general-purpose detectors are ATLAS [58] and CMS [59]. The two other major experiments are the Large Hadron Collider beauty (LHCb) [60], specializing on probing flavour physics produced at the LHC, and A Large Ion Collider Experiment (ALICE) [61], focusing on heavy ion physics. Some smaller experiments are located near the four main interaction regions, outside the main experiments. For instance the Large Hadron Collider forward (LHCf) [62] experiment is located on both sides of the ATLAS detector, with the purpose of measuring forward particles produced in the collisions and mimicking cosmic rays in laboratory conditions. Another example is the TOTal cross-section, Elastic scattering and diffractive
dissociation Measurement (TOTEM) [63], by the CMS cavern. The main purpose of TOTEM is to measure the total elastic and single or double diffractive cross-section of proton-proton collisions. The reason why TOTEM is needed is due to the limitation of the CMS detector in a very forward reason because of the front magnets, hence low angle particles escape the detector acceptance and cannot be detected.

2.1.2 Facts and figures about the LHC

Accelerator performance is mainly a function of two physics related parameters. The first, that was also mentioned previously, is the Centre-of-Mass (CoM) energy that the accelerator can reach. The CoM is measured in electron volt (eV) and noted usually with the use of the Mandelstam variable $\sqrt{s}$ [3]. The second important parameter is the so-called instantaneous luminosity, noted with $L_{\text{inst}}$, and describes the number of physics processes a collider is able to provide per unit time. The instantaneous luminosity can be obtained by measuring a given physics process with the following formula [64]:

$$L_{\text{inst}} = \frac{1}{\sigma_{\text{process}}} \cdot \frac{dN_{\text{collision}}}{dt} \quad (2.1)$$

where the Standard Units (SI) for it are $cm^{-2} \cdot s^{-1}$. In eq. 2.1 the other two parameters indicate the event frequency of a given physics process ($dN/dt$), measured in $s^{-1}$ and the probability of the process to occur, called cross-section ($\sigma$) and measured in $cm^{-2}$ or inverse barn $b^{-1} = 10^{24} cm^{-2}$. Hence, the most frequently used unit for luminosity in HEP is $b^{-1} s^{-1}$. Fig. 2.3 illustrates the dominance of the LHC in terms of energy.

The quantity, that better expresses the discovery potential of a given accelerator set-up is the integrated luminosity given by the following formula:

$$L_{\text{int}} = \int_{0}^{T} L_{\text{inst}}(t) \, dt \quad (2.2)$$

By combining eq. 2.1 and eq. 2.2 the estimated number of events of a given process can be calculated on a given operation time:

$$N_{\text{process}} = L_{\text{int}} \cdot \sigma_{\text{process}} \quad (2.3)$$

From eq. 2.3 it can be seen that the important parameter to maximize with an accelerator is the integrated luminosity because this will yield more data for a given process. Since experiments at colliders are not 100% efficient in collecting data, the expression above should be considered a theoretical maximum rather than the actual exact data collected by the experiment for a given process. Therefore, what the physics experiments are mostly interested at is the so called recorded luminosity, which includes all the detector related effects in data taking, like acceptance, efficiency and dead-time.

The LHC machine was designed to operate at a CoM of $\sqrt{s} = 14$ TeV and instantaneous luminosity of $L_{\text{inst}} = 1 \ast 10^{10} b^{-1} s^{-1}$ [56] for proton-proton collisions. For the lead-lead collision the respecting values are $\sqrt{s} = 5.02$ TeV and $L_{\text{inst}} = 1 \ast 10^{3} b^{-1} s^{-1}$ [65].
2.2 The ATLAS detector

In 1992 two initially proposed collaborations, Experiment for Accurate Gamma Lepton and Energy measurements (EAGLE) and Apparatus with Super Conductor Toroids (ASCOT) joined effort to build a general-purpose detector for the LHC. The resulting detector, named **ATLAS** [58], started to be constructed in 2003 and completed in 2008.

The detector location is in a 100m deep pit, called Point-1, near the Swiss site of **CERN** next to Meyrin. **ATLAS** is 46m long and 25m radius, weighting 7000 tonnes with more than 3000 km of cabling.

The design of the detector follows the standard forward-backward symmetry of general purpose detectors. It covers a $2\pi$ angle around the interaction points of the two circulating beams from the LHC. Since in every interaction different type of particles can be detected, **ATLAS** consists of multiple sub-detector layers, each one dedicated to a different purpose.

The detector consists of two main regions. The central region, which is called the barrel, and the two "end-caps" covering the front and the back of the detector. The different sub-detector layers as encountered by an out-going particle are the following; the first and closest system to the LHC beam is called the Inner Detector (ID). The main purpose of the **ID** is to perform high

---

1 The **LHC** during the first run of operation (Run-1) managed to reach the energy of $\sqrt{s} = 8\text{TeV}$ and for the following Run-2 the energy of $\sqrt{s} = 13\text{TeV}$. 

---

Figure 2.3: The comparison between Tevatron and the LHC. In the x-axis of the figure the logarithm of CoM energy is given for Tevatron and the LHC. On the y-axis the total cross-section of various different physics processes is quoted.

Reaching the LHC design CoM energy remains a technical challenge and therefore has not yet been achieved. For the luminosity the situation is radically different; since 2017 the LHC managed to double the design value ($2.06 \times 10^{14} \text{b}^{-1}\text{s}^{-1}$) [66].

---
2.2 The ATLAS detector

Figure 2.4: Layout image of the ATLAS detector. The arrows point to the different detector layers indicating the various sub-systems. On the picture the detector dimensions are also quoted [58].

resolution particle trajectory identification (tracking) and interaction point (vertex) finding of the outgoing charged particles. A 2 T magnetic field permeates the detector, in order to bend the charged particle trajectories, thus allowing the measurement of their momentum as part of the trajectory. The next two sub-systems following the ID are called the Electromagnetic Calorimeter (ECAL) and Hadronic Calorimeter (HCAL). The ECAL is responsible for measuring the energy of electrons and photons, whilst the HCAL is responsible for the hadronic jets consisting mostly of pions and kaons. The next to last detector subsystem placed on the perimeter of the detector is dedicated to the detection of muons and called the Muon Spectrometer (MS). The MS along with information retrieved from the ID is able to measure the position and the momentum of the most penetrating particles, the muons. The last system needed in the ATLAS detector is the magnet system allowing the bending of the charged particles. The overall ATLAS layout with all the subsystems can be seen in fig. 2.4.

2.2.1 ATLAS coordinates

An important step in understanding and using the ATLAS detector is to define the coordinate system. The most common system used to define coordinates in mathematics is the widely known Cartesian system. Applying the Cartesian coordinates in ATLAS the z-axis will point along the beam pipe, the x-axis is pointing towards the centre of the LHC ring and subsequently the y-axis points upwards.

However, given the symmetry and shape of the ATLAS detector the most convenient coordinate system to be used is a modified cylindrical coordinate system. The usage of the word modified is made due to the fact that instead of using the standard \((r, \theta, \phi)\) coordinates, in ATLAS we
replace the $\theta$ angle with a quantity called pseudorapidity ($\eta$).

\[
\eta = -\ln\left(\tan\frac{\theta}{2}\right)
\]  

(2.4)

The azimuthial angle $\phi$ is measured around the beam spot with a range of $[0-2\pi]$. With this particular parameterization the $\eta$ allowed range is $[0, \pm\infty]$, where 0 is perpendicular to the beam axis and $\pm\infty$ to the forward detector regions.

The main reason for making the conversion of standard cylindrical coordinates into the modified version is due to the fact that differences in pseudorapidity are Lorentz invariant on boosts along the longitudinal axis, hence more convenient for physics analysis. In general Lorentz transformation provide a set of modification that are applied to quantities when the frame of reference is changing. For this reason it’s highly desired in particle physics to identify quantities that are Lorentz invariant so that general statements and observation can be obtained [67].

### 2.2.2 Magnet system

The first component described is the ATLAS magnet system. The design of the system is one of the biggest differences with the other main general-purpose detector installed at the LHC, CMS. The main key difference is that CMS deploys a single solenoid magnet. In contrast the ATLAS magnetic system is composed of 4 super conducting magnets each with a 22m diameter and 26m in length. The four different magnets are installed in different detector regions. The first magnet is in the central region of the detector close to the ID detector, called the central solenoid. The second one, covering the outer detector layers in the barrel region, is called the barrel toroid. Finally, the two end-cap toroids cover the two forward regions of the detector.

The magnet layout is presented in fig. 2.5a and 2.5b

The central solenoid is 5.3 m long, 2.4 m in diameter with a 4.5 cm thick coil, weighing 5 tonnes.

![Geometry of magnets surrounding the calorimeter. The 8x barrel coils along with the end-cap parts can be seen [58].](image)

![Schematic view of the super conducting spectrometer magnets [68].](image)

It generates a 2 T field, parallel to the beam axis, leading to the bending of outgoing charged particles, in the xy plane, within the ID detector. The central magnet system assists the high
2.2 The ATLAS detector

precision measurement of charged particle momentum up to 100 GeV [68]. The toroid system contains two different components. The first is the barrel toroid which is 25.3 m long and 20.1 m in diameter of its outer layer. It consists of 8 separate coils which weigh 830 tonnes in total generating a toroidal magnetic field of 0.5 T within the muon detector tracking volume. The main purpose of the barrel toroid is to bend the muons and assist with the measurement of their momenta. Lastly, the end-cap components of the toroid system are made of two sets of 8 coils, one on each side, with their cryostats. They have a 5 m axial length, a 1.65 m inner diameter and a 10.7 m outer diameter. The end-cap toroids generate a 1 T toroidal magnetic field for the forward regions of the muon system. The main goal, as for all magnet systems, is to bend the trajectory of particles that cross the MS in the $\eta$ direction.

The are several reasons why ATLAS decided to deploy a different magnetic system from CMS [69]. The most important argument is that with the use of the toroid based system a larger integral along the particle trajectory can be achieved, with a limited material budget. Keeping the material inside the detector as low as possible is important to avoid multiple scattering of particles, which would complicate the detection of certain signatures.

2.2.3 Inner detector

The closest detector system to the collision point in ATLAS is the ID system [70]. It extends from 33.25 mm above the beam axis up to 1.082 m height. Its length is 6.2 m, with a diameter of 2.1 m and is submerged into a 2 T field. The main purpose of the ID is to measure the momentum, charge and direction of flight of a high flux of outgoing particles in each collision.

The ID is separated into three components, each serving a different purpose and using a different detection technology. The first sub-detector is the Silicon Pixel Tracker (Pixel detector), followed by the SemiConductor Tracker (SCT) and finally the Transition Radiation Tracker (TRT). During LS1 the collaboration decided to insert another pixel detector layer closer to the beam pipe, called the Insertable B-layer (IBL), mainly due to degradation observed to the original innermost layer due to radiation and to assist the other pixel layers, due to the high occupancy observed during Run-1 (Fig. 2.6).

A pixel-detector-only ID would were the best possible option for high resolution tracking in a particle physics detector. However, cost is a limiting factor in detector construction leading to necessary compromises in detection methods. For this reason ATLAS decided to change its ID technologies with 2 sub-detectors being made from silicon and the final layer being based on gas detectors. The first 8 layers, which are subject to high particle occupancy, are made of pixels to maximize the resolution. Subsequently, when the density of particles reduces the strip layers will be used since they provide an acceptable, cost effective, resolution. The final layers which produce the biggest cost reduction is the gas based TRT.

In the following sub-sections the three different sub-detectors are discussed in more detail, starting from innermost to the outermost. Finally, it’s worth mentioning that the usage of the ID tracker up to now in ATLAS yielded a robust track reconstruction, with an excellent perform-
2.2 The ATLAS detector

Figure 2.6: Layout of the ATLAS Inner Detector sub-systems, dimensions of all the different layers are quoted [70].

The resolution on certain track parameters reached up to 20µm and the position of origin for all particles above 500MeV in a region of |\( \eta \) < 2.5 was possible. The performance of the ID was a key factor for the ATLAS detector operation to the discover the Higgs boson in 2012.

Insertable B-Layer

The closest detector system to the beam pipe in ATLAS is a layer of pixel detector at 3.3cm from the beam pipe axis [71]. It was inserted after the LS1 due to the high degradation of the original innermost pixel layer that was observed during Run-1. With the inclusion of the IBL ATLAS managed to avoid replacing the original innermost layer, despite to the radiation damage and improve the tracking at the ID level. Without inserting the IBL (nor replacing the original innermost layer), there would were inefficiencies and resolution reduction on parameters that the data recording system depends on.

On the technical side the IBL is a single pixel detector layer covering the whole azimuthial range around the beam pipe and an \( \eta \) range of [−3, 3]. The pixel sensor on the IBL have a size of 50x250 \( \mu m^2 \), with a resolution of 8x40 \( \mu m^2 \). The pixel module technology is Complementary Metal-Oxide-Semiconductor (CMOS).
2.2 The ATLAS detector

**Pixel detector**

The next detector layers after the IBL are the three Pixel detector layers [72]. The three layers are placed concentrically around the beam pipe at radii of 4cm, 11cm and 14cm respectively. Their installation happened during the construction of the ATLAS detector. The main task for the Pixel detector is to perform precision measurement on particle momentum, and charge, and identifying the interaction point (vertex).

The Pixel detector has the highest granularity out of all other detectors in ATLAS. The coverage is $|\eta| < 2.5$, where the range $|\eta| < 1.7$ is covered by three layers in the barrel and the regions $1.7 < |\eta| < 2.5$ by three pixel disks per side for the end-caps. With this placing each particle produced will cross at least three layers. When a particle crosses one of the layers, it deposits its energy in multiple pixel modules, making particle localization a daunting task. Technically the pixel detector consists of 80 million CMOS pixel sensors with dimensions of 50x400 $\mu m^2$ and a resolution of 14x115 $\mu m^2$.

**SemiConductor tracker**

Following the three pixel layers is the SemiConductor (SCT) [73]. The SCT is made of four concentric layers covering the barrel region $|\eta| < 1.4$ [74] and nine disks per end-cap region to cover the $1.4 < |\eta| < 2.5$ range [75]. The detector consists of 4088 modules of silicon-strip detectors. The main reason for switching from pixel modules to silicon strips is that the flux of particles at the distances of the SCT is significantly smaller. This means that the granularity needed at this stage is much lower, whereas the precision of the measurement remains comparable. Deciding to change from pixel sensors to strips constitutes a compromise on the granularity for practical issues like cost and construction.

Technically the SCT is based on the same detection concept and material as the pixel modules. Its size is 6.36x6.40 $cm^2$ with 768 strips each with a 80 $\mu m$ pitch in each module. Each layer consists of double strip modules glued back-to-back. With this placing, a particle passing through the central region is measured through as many as 8 hits, rather than 4, improving the measurement on the position. The SCT reaches an intrinsic resolution of 17$\mu m$ in the $r-\phi$ direction and 580$\mu m$ in the $z$ direction, where $r$ is the radius pointing from the centre of collision upwards. The SCT layers are placed at a radii of 300, 373, 447 and 520 mm.

**Transition radiation tracker**

The last ID sub system is the transition radiation tracker (TRT). It extends from a radial distance of 554 mm to 1082 mm from the centre of the beam pipe. The working principle of the TRT is based on the concept of gas ionization inside the straws when a charged particle crosses. The free electrons produced from the ionization process drift via the wire towards the anode while being multiplied with consequent avalanches. Having a knowledge of the electric field inside the tube allows calculation of the position of the trajectory by measuring the time needed for the electrons to drift to the anode.
The TRT is based on the use of 4mm thick straw detectors at a maximum length of 150cm. The straw tubes are made of multi-layer film reinforced with carbon fibers and contain a 30\(\mu m\) gold plated tungsten wire in the centre. The gas used in the straws is an admixture of carbon dioxide, oxygen and xenon (27\%, 3\%, 70\%) [76], allowing differentiation between electrons and photons radiated from the material between the straws.

As for most sub-detectors in ATLAS, the TRT is partitioned in a barrel region [77] (|\(\eta\)| < 1) and an end-cap region [78] (1 < |\(\eta\)| < 2). The TRT barrel is made of 50,000 individual straws, divided in half to provide a reduced occupancy and possibility of read-out on each end. For the end-cap region 320,000 straws are used. Each read-out chain provides a 170 \(\mu m\) per straw spatial resolution, with two independent thresholds available for detecting an incoming particle.

The resolution of the TRT is significantly lower than the silicon based detectors; however, the resolution loss can be recovered from the large amount of hits provided for each track. Within the TRT, charged particles provide at least 30 hits and along with the long lever arm, the momentum resolution improves significantly.

The presence of Xenon in the gas admixture allows the TRT detector to perform particle identification. Charged particles in ATLAS reach the energy required to produce soft X-ray photons when crossing through different materials. The emitted Transition-Radiation (TR) photons, are absorbed by the Xenon atoms, which subsequently get ionized. Therefore any detected photon in the TRT indicates the presence of a charged particle.

### 2.2.4 Calorimeters

In the following section the description of the two calorimeter sub-systems is provided. The main purpose of calorimeters is to stop the corresponding particles of each category and in the process, measure their energy. As in most general-purpose detectors, ATLAS has decided to employ a two calorimeter detection system with the innermost of the two for photons and electrons (electromagnetic) [79],[80] and the other calorimeter for detecting hadrons [81],[82] (hadronic) (Fig. 2.7). The ATLAS calorimeter coverage is 2\(\pi\) in \(\phi\) and |\(\eta\)| < 4.9, shared between a barrel (|\(\eta\)| < 1.475) and two symmetric end-caps (1.5 < |\(\eta\)| < 3.2).

![Figure 2.7: The layout of the ECAL and the HCAL [79].](image)
Electromagnetic calorimeter

The technology used for the ECAL is based on lead-Liquid-Argon (LAr) [80], with accordion-shaped electrodes and lead absorber plates over the full coverage. The advantage of using LAr is that it’s a highly linear material in terms of response to particles passing through, with large signal yields and resistance to radiation. Additionally, deploying the electrodes and the absorber in an accordion-shape, oriented in the radial direction, allows a complete, crack free coverage in the azimuthal direction.

The barrel region of the ECAL is divided into three longitudinal parts. Each layer has a different depth in the $\eta - \phi$ direction serving a different purpose. The first layer (with thickness of 4.3 radiation lengths ($X_0$)), provides an accurate measurement of the coordinates on the incoming particle. The second layer, much thicker than the first (16 $X_0$), serves as particle absorber. As particles are stopped, the position and magnitude of the deposited energy are measured. The final layer, which is very thin (2 $X_0$), detects showers not fully absorbed by the calorimeter (punch-through). The granularity of the last layer is much coarser than that of the other two layers.

The performance of the ECAL was measured during the development phase with electron beams. The modules used to estimate the performance were equivalent to the modules installed in ATLAS. The energy resolution is given by the following eq. 2.5 [80].

$$\frac{\sigma(E)}{E} = \frac{10\%}{\sqrt{E}} \oplus \frac{0.17 \, GeV}{E} \oplus 0.7\% \quad (2.5)$$

where $E$ is measured in GeV. The first term in eq. 2.5 originates from the statistical behavior of the showers, and is fixed for a set of energy values. The second term accounts for the electronics noise, again fixed for the energy scale of the calculation, and finally the last term accounts for the non-uniformity of the calorimeter response. The final term is generally a property of the detector (e.g. layout, construction) and is not dependent on the beam energy.

Hadronic calorimeter

The main goal of the HCAL [82] is to stop hadronic compound states, like pions. As for the ECAL, the HCAL is separated into barrel and end-cap regions, deploying different technologies.

For the barrel region a tile based approach was used where iron plates act as absorbers and plastic scintillating tiles as active material. Wavelength-shifting fibers are read out by photomultipliers coupled to plastic scintillators. The tile calorimeter is 7.4 interaction lengths ($7.4\lambda_I$) deep.

For the end-cap region a similar technology as for the ECAL was used. The main motivation for this choice is the increase of the particle flux of hadrons with pseudorapidity. The HCAL end-cap differs from the ECAL technology mostly for the absorber material (copper) and the geometry. The most forward region of the end-cap HCAL is covered by a LAr calorimeter, called the FCAL. The thickness is exactly the same as for the barrel region and equal to $7.4\lambda_I$.

As for the determination of the HCAL a very similar approach was employed as for the ECAL,
with a pion beam used instead of electrons. The resulting resolution for the Tile calorimeter is [81]:

$$\frac{\sigma(E)}{E} = \frac{52\%}{\sqrt{E}} \oplus \frac{160\text{GeV}}{E} \oplus 3.0$$ (2.6)

with $E$ expressed in GeV. With the breakdown of the three terms being analogous to what discussed for the ECAL. A similar resolution was obtained for the forwards LAr.

The main reason why the resolution of the hadronic calorimeter is worse than the resolution of the electromagnetic is due to the particle nature used for measuring it. For the hadronic calorimeter a pion beam is being used whereas in the electromagnetic an electron beam. The difference of the two particles is mainly on the way they are interacting inside matter. Pions in the calorimeter lose energy mainly due to gluon emissions and subsequently from ionisation. For this reason a hadronic shower is much more spread than the electromagnetic shower (which is mostly developing due to ionisation) leading to a worse resolution.

### 2.2.5 Muon spectrometer

The last ATLAS sub-detector system discussed here is the MS (Fig. 2.8) [83]. It consists of gas chambers (placed all around the barrel region) and two end-cap wheels (in the forward regions). The main purpose of the MS is to measure the momenta of the muons, as indicated by the name. The technologies of the gas chamber vary based on the usage. The first category consists of fast, coarser-precision trigger chambers and the second of slower readout-speed chambers used for higher precision measurements. The detectors are positioned in such a way that muons reaching the MS system are traversing at least three detector layers, allowing the different detector types to provide the spatial information of the muon trajectory. The three hits along with the deflection from the three toroids allow ATLAS to have a high precision momentum measurement for the muons.

Different chamber technologies are employed for triggering and precision measurements in the MS system. The detectors used for precision momentum and energy measurements are the Monitored Drift Tubes (MDT) and the Cathode Strip Chambers (CSC). For the trigger chambers Resistive-Plate Chambers (RPC) and Thin-Gap Chambers (TGC), are employed. The main difference in each of the detector types in each category is dictated by the installation region. MDT and RPC are placed in the barrel region of the MS (0 < $|\eta|$ < 1), whereas the CSC and the TGC in the end-cap regions (1 < $|\eta|$ < 2.7).

The MDT chambers used for momentum and energy measurements in the barrel are made of aluminum tubes (30mm in diameter and 400µm thick walls). Inside each tube a 50µm diameter central W-Re wire is surrounded by non-flammable $Ar - CH_4 - N_2$ mixture at a pressure of 3 bars. The MDT are constructed with a layout of 2x4 layers of drift tubes for the inner layer and 2x3 layers for the middle and out MDT layers. The reason for having several MDT layers is to increase the resolution on the particle position measurement beyond the capability of the single wire (80µm). MDTs in ATLAS are used only to perform a measurement on the $\eta$ coordinate (as they are placed orthogonal to the beam axis). For the $\phi$ coordinate the RPC and TGC chambers are used.
2.2 The ATLAS detector

Figure 2.8: Layout of the ATLAS MS with all the different detector types. The location of the various sub-systems is shown for both the barrel and the end-cap region [58].

In the same category as the MDT, although placed in the end-cap, are the CSC chambers. CSCs are multiwire proportional chambers with a symmetric cell, with cathode strips positioned above and below the anode wires. The gas used for the CSCs is an admixture of non-inflammable $Ar - CO_2 - CF_2$. For the measurement of the transverse position of the particles, mutually orthogonal chamber pairs are used. One set is placed parallel to the beam axis whereas the other is orthogonal to it.

The RPCs are the muon detector barrel trigger elements. These are detectors based on gas detection methods. Two parallel resistive plates are separated by insulating spacers, causing primarily produced ionization electrons to multiply into avalanches in a highly uniform electric field. The electron avalanche is then detected by aluminum strips separated from the plates with an insulating film.

In ATLAS a large amount of MDTs, CSCs, RPCs and TGCs are used. The positioning of the different type of detectors in the barrel region is in three concentric cylinders at distances of 5, 7.5 and 10 m. For the end-cap regions 2 disks called wheels are used with two layers of detectors each at a position of 7, 10, 14 and 21-23 m from the interaction point. Although the MS covers the whole area in $\eta$ at the appropriate radius, a small opening in the central $r$-$\phi$ plane at $\eta = 0$ is present, called the "MS crack". The reason for the crack position is to allow the readout cables and other services of the ATLAS detector to pass through the MS system.

2.2.6 Trigger and Data acquisition system

The first system of the ATLAS detector that is extending out of the cavern is the Trigger and Data Acquisition System (TDAQ) [84]. In the LHC there is a bunch crossing every $25ns$; a proton-proton collision in the centre of ATLAS is called an event. The ATLAS detector is exposed for every second of stable accelerator operation to 40 million events per bunch crossing. Each of
2.2 The ATLAS detector

Figure 2.9: The ATLAS TDAQ system for Run-2. The arrows indicate the dataflow through the system to allow all the sub-components to contribute in the triggering process [84].

those events in the form of raw data read from the detector is $\approx 1.6$MB. Storing all this raw information without any filtering would lead to a storage requirement for ATLAS of about 6400TB per second. An additional factor affecting data storage if no filter is applied arises from the readout capability of the detector itself along with the recording rate available on state-of-the-art storage modules. Therefore ATLAS deploys the so called TDAQ system which is in charge of performing an online filtering of events based on certain physics requirement in order to reduce the recording rate and maintain the best possible physics data quality (Trigger). Additionally, the TDAQ system monitors continuously the state of the detector, the dataflow and the configuration of certain components (Data Acquisition). The diagram of the ATLAS TDAQ system as it was in Run-2 is shown in fig. 2.9

The TDAQ system is separated into two sub-systems to allow a better data selection by employing data pipelines. The first sub-system in the sequence is a hardware based system called the Level-1 (L1) trigger, performing a fast, coarse selection of events, to reduce the data rate to 100kHz, along with all the services used to monitor the state of the hardware used for this selection. The subsequent sub-system to L1 is the High Level Trigger (HLT) which is based completely on CPUs running software, and performs a more complex and elaborate selection reducing further the rate of data into storage to 1-2 kHz.

The online decision taken by the ATLAS trigger system is based on certain information available from specific parts of the detector. The L1 system uses the information provided mainly by the MS (RPC and TGC) and the calorimeters. Each event that is accepted from the L1 is then allowed to pass at the HLT level (L1-accept). At the HLT level a more elaborate decision is made based on a wider set of information provided from ATLAS. Data that pass the HLT requirements are stored permanently and prepared to be used for analysis purposes. At the level of HLT, similar to L1, there are multiple selection requirements all optimized to serve different physics
2.2 The ATLAS detector

analyses. All the L1 and HLT selections (trigger chains), define the trigger menu of ATLAS which is designed and optimized based on the physics program and luminosity achieved by the LHC. Each trigger chain at the level of the HLT can be prescaled based on its collection rate to ensure that the overall rate remains within nominal values. To achieve the goal of a relatively constant storage rate, trigger chains that exceed their allowed rate in the trigger system are subject to the concept of prescaling, whereby a subset of n out of m triggers of a given type is randomly selected for the next trigger stage, while the remaining m-n are rejected.

On the DAQ side the main system to monitor all individual components in the chain is the Detector Control System (DCS) [85]. The DCS is a service present in most hardware sub-systems providing a uniform interface to monitor certain system parameters during operation. It’s an important feature for all sub-detectors and components since it provides e.g. real-time information for any failure in the cavern when access is not allowed.

Since the hardware component of this thesis is based on TDAQ system, a more detailed explanation of the two trigger levels along with the DCS is provided in the following three sub-sections of the chapter.

Level-1 trigger

The first trigger level completely implemented in hardware is the Level-1 trigger [86]. L1 is installed next to the ATLAS cavern, 100 m subsurface. High speed readout links leave the back-end chips of the different detectors and are connected to custom hardware boards dedicated to extract and package the required information for the first level filtering. L1 employs mainly data from the calorimeters (L1 calorimeter) and the MS (L1 muon), to reduce the rate from 40 MHz to 100 kHz.

The part of L1 responsible for taking the decisions is the Central Trigger Processor (CTP). The CTP decides whether the event will be stored in temporary storing buffers for further HLT processing or discarded. An additional task of the L1 sub-system to provide the HLT with a pointer to certain detector regions where activity giving rise to the L1 selection was detected. Those regions are called Regions of Interest (RoI) and are defined as $\eta - \phi$ detector regions containing physics objects of interest (L1 objects). The key advantage of using RoIs is that the HLT does not need to perform a full event reconstruction and can, at least in early stages of the selection, focus on specific regions, minimizing the required processing time.

During Run-2 ATLAS installed another L1 sub-system called L1-Topological trigger processor (L1-Topo). L1-topo allows the L1 decision to be based on complex combinations of L1 objects rather than just individual energies and object multiplicities to decide not only based on the energy and momentum of physics objects in the calorimeter and the MS systems but also including kinematic information for those objects.
2.3 ATLAS Upgrade

High Level Trigger

The events accepted by the L1 trigger are transferred to the surface of the ATLAS facility. The HLT [87] is a large CPU farm located on the surface close to the main building of the cavern. The rate of data is thereafter reduced from 100kHz to 1-2kHz, which is the recording rate of the experiment. The HLT employs complicated algorithms to select the events. Data are retrieved following a L1 accept, and physics objects start to be reconstructed inside the RoI indicated from L1. The reconstruction of physics objects in HLT and L1 is different. At the HLT level more information is available, allowing the HLT to reconstruct and associate the objects e.g. with their corresponding tracks in the ID, or use the higher granularity calorimetric outputs.

Most of the HLT processing power is used to do tracking in the ID. The main reason is because by construction tracking in ATLAS suffers from a huge number of possible tracks (combinatorics) and a large number of secondary collisions (pile-up). To solve this issue, when the luminosity at the LHC is going to increase, multiple ideas are currently evaluated. The most favorable are, to either increase the size of the HLT-farm or move into a hybrid farm where custom hardware boards dedicated to tracking are going to be used. A better overall explanation about the benefits of hardware based tracking is provided in the following chapters of the thesis.

Detector Control System

The last important system presented in this chapter is going to be the DCS [88]. It’s an important sub-system for maintaining the smooth operation of the detector during data taking. DCS is separated into two sub-systems depending on the modules that are monitored. The first sub-system is the front-end (FE) DCS system which monitors hardware components such as power supplies, cooling systems and environmental sensors. The second sub-system is the back-end (BE) DCS in charge of integrating all the front-end controls. Several standardized electronics protocols (SCADA, OPC, etc.) were deployed to ensure the operation of the DCS system in ATLAS.

2.3 ATLAS Upgrade

The ATLAS detector provided an outstanding performance during Run-1 and Run-2 leading to the discovery of the Higgs boson along with improvements in many other high energy physics sectors. In 2019 the detector was turned-off, to undergo the foreseen upgrade plans to increase the data quality under the new accelerator conditions. The various upgrades planned are outlined in detail in the following chapter of the thesis, where a bigger emphasis is given on the upcoming Phase-II upgrade for the HL-LHC.
3.1 Run-2 performance

Run-2 was the first operation period that the LHC was able to deliver constantly an energy of 13 TeV. In addition to the highest energy ever reached with an accelerator, the LHC provided the highest ever instantaneous luminosity by a human built machine. The ATLAS detector as shown in fig. 3.1 was able to record with a very high efficiency the delivered luminosity from the LHC. The outstanding performance of ATLAS during Run-2 allowed the collaboration to gain access to rare decays such as $B_s^0 \rightarrow \mu^+ \mu^-$. 

Figure 3.1: Total integrated luminosity during Run-2. With the green histogram the luminosity provided from the LHC is shown, with yellow the corresponding fraction that was recorder by ATLAS. Finally with blue the luminosity that after the offline selections was shown to be good for physics [89].
In addition to the energy and luminosity increase, a large number of interactions per bunch crossing was observed. Every time a collision is produced at the center of ATLAS a number of secondary collisions occur as well. Those collisions occur either in parallel to the collision of interest (in-time) or slightly before or after (out-of-time). This is what is called pile-up ($\mu$) in particle physics and affects the performance of the detector TDAQ system. Additional sources of particle/detector background are present during data taking, such as beam halo events, cavern background, etc. In fig. 3.2 the pile-up distribution during Run-2 is shown. The main way that pile-up affects the ATLAS detector is that many sub-systems have a much longer sensitivity than the 25ns between collisions (acquisition window). This effect causes mis-measurements by the TDAQ system, for example more energy could be added in a jet or mis-reconstructed background as high-momentum muons.

Finally, the last factor limiting data storage arises from the need for a decision system (trigger). Designing effectively the selections for the various physics processes (trigger menu) is a key task for ensuring that the collected data have the best possible quality. During Run-2 ATLAS needed to revisit its trigger menu after 2016 since the LHC reached a factor 1.4 higher peak luminosity than its design value. The effect observed was that many trigger chains had to be prescaled reducing the reach of ATLAS in the various physics channels. The menu that was used from 2016 and beyond was designed for an instantaneous luminosity of $2 \times 10^{32} cm^{-2}s^{-1}$. The newly revisited menu proved sufficient for all the $\sim 3000$ trigger chains, currently implemented in ATLAS, to match the hardware and software limitations from the implementation (e.g. data rates, CPU time).

### 3.2 Phase-I upgrade overview

At the time of writing the ATLAS detector is undergoing the first major upgrade (Phase-I) in preparation for Run-3. The LHC upgrade will yield an increase in luminosity of a factor of 1.5,
whilst the average pile-up through Run-3 is expected to be higher than what was measured in Run-2. The increased radiation damage to various sub-detector systems, due to all the years of operation, is another reason to justify the need of the Phase-I upgrade. Certain detector inefficiencies might increase data rates for the corresponding trigger chains, hence to ensure that data rates will be under control the replacement of those systems is needed. These updates account also for the LHC operation at its maximum mode (HL-LHC) which along with the second major upgrade plan (Phase-II) will keep ATLAS at the physics frontier.

3.2.1 New Small Wheel

The first detector sub-system being updated are the two muon small wheels in the end-cap region on each side of the detector. The currently installed wheels are at their limit of lifetime, mainly due to radiation damage. The size of the wheels are 5 m in radius and cover the range of $1.3 < |\eta| < 2.7$. The chambers used in each wheel are made of TGC technology, planned to be used mainly for triggering. The main goal of the New Small Wheel (NSW) [90] is to reduce significantly the muon trigger rates caused mainly by noise or accidental coincidences in the end-cap region.

3.2.2 Liquid Argon

For the LAr system the upgrade concerns only the read-out electronic systems. Both the front-end and the back-end boards are planned to be replaced allowing a higher granularity of the tower information read-out. Increasing the granularity will allow associated trigger chains to perform a better jet reconstruction and hence select events with a higher accuracy. Ensuring that triggers using the LAr information are within nominal rates is important for both the upcoming Run-3 and the HL-LHC runs.

3.2.3 Muon detectors

In ATLAS the NSW along with the bigger wheel cover the range of $1.3 < |\eta| < 2.7$ and $1.0 < |\eta| < 2.7$ respectively. The current geometry, with the two wheels (fig. 2.8), creates a gap at the range $1.0 < |\eta| < 1.3$ between the two wheels, forcing the trigger system to reject useful information. To avoid the issue new RPC and MDT detectors are going to be installed in the small sectors of the muon spectrometer, aiming to recover the losses due to those regions.

3.2.4 TDAQ system

The last ATLAS sub-system to be upgraded during Phase-I, is the TDAQ system. Comparing the expected architecture of Run-3 (fig. 3.3) with the equivalent architecture of the TDAQ system from Run-2 (fig. 2.9), it can be seen that it remains largely the same. Both architectures retain the L1 hardware system followed by the software based HLT system. The only sub-system that
was removed from the actual final implementation is the Fast Tracker (FTK) system \cite{91}, which was absorbed into the HLT farm. The upgrades on certain sub-systems includes also the HL-LHC runs (e.g. eFEX \cite{92}). The main upgraded TDAQ sub-systems can be seen with the red boxes in the following diagram.

Starting from left to right in fig. 3.3 we see that the L1Calo system will include for Run-3 a new Feature EXtractor (FEX), which is a custom board based on Field Programmable Gate Array (FPGA) providing better discrimination of jets, photons, electrons and taus. The inclusion of the new FEX card assists the important task of maintaining trigger rates within nominal. The next system to be upgraded is the L1Topo sub-system. The upgrade concerns the inclusion of a new custom board allowing the implementation of more complicated topological algorithms.

The next system upgraded in fig. 3.3 concerns the electronics processing the information from the newly inserted NSW. Therefore the new Serial Logic board will be included and will reduce the rates of triggers which are based on the muon information from the respected regions of the NSW. In addition the muon CTP \cite{93} will be upgraded to ensure the proper interfacing of the muon data with the rest of the TDAQ system. Finally, the last L1 system that will be included is called Front End LInk eXtractor (FELIX), and will act as a router between the custom front-end links of the various sub-detectors and a commercial multi-gigabit network used for transmitting data to the HLT.

The HLT farm is upgraded during Phase-I, too. The main difference relies on the migration of the currently sequential algorithms into multi-threaded applications. In the current form the HLT farm is processing L1-accepted events with a framework called Athena, that contains a set of sequential algorithms executed in a strict order. Additionally, a strict rule is enforced allowing the algorithms in the flow to depend only on algorithms that have already been executed.
In Phase-I the Athena data-flow will migrate to the so called AthenaMT framework. The introduction of multi-threading allows the better utilization of the HLT farm CPU cores, where multiple events can be handled at the same time on the same physical CPU. The upgrades are expected to maximize efficiency in terms of memory access and keep the thread utilization at the highest possible value. Finally, the HLT farm will upgrade the CPU cores, ensuring the farm longevity.

The Phase-I upgrade is important for the operation of ATLAS during Run-3 and will set the basis for the HL-LHC era. However it’s not enough to handle all the challenges from HL-LHC, hence a further upgrade is planned after the completion of Run-3. The discussion of Phase-II will happen in the following sections of the current chapter below.

3.3 Phase-II motivation

In 2024, the LHC is foreseen to undergo its biggest upgrade, reaching the maximum energy of $\sqrt{s} = 14\,\text{TeV}$ and an instantaneous luminosity of $L_{\text{inst}} = 5 \times 10^{34}\,\text{cm}^{-2}\,\text{s}^{-1}$. During the HL-LHC time the LHC is expected to deliver an integrated luminosity of about $L_{\text{int}} = 3000\,\text{fb}^{-1}$ which in comparison to the $350\,\text{fb}^{-1}$ expected during Run-3 shows a significant increase. The main effect present with the increased instantaneous luminosity is the increase of the pile-up collisions. More pile-up collisions push the detector TDAQ systems to their limits. Dealing with the effects of pile-up is crucial in order to avoid a reduction in the sensitivity of the physics processes observed by ATLAS.

Currently, the main focus for ATLAS is on the precise measurement of the Higgs boson properties. Studying the nature of the Higgs spin and parity, along with it’s couplings to as many final states as possible, constitutes important steps towards future discoveries beyond the SM. An additional goal for ATLAS during the HL-LHC era is to gain access to the measurement of the Higgs self-coupling. Measuring it would allow physicists to establish whether the Higgs mechanism is responsible for electroweak symmetry breaking, one of the biggest current questions in physics. But Higgs physics is not the only reason why the HL-LHC is important for ATLAS. Measuring a deviation from the SM predicted value of the weak boson scattering cross-section can verify or disprove various NP theories (e.g. technicolor, little Higgs, etc.), in addition it will set the ground for resonances at the TeV scale. Finally, analyses on direct searches for NP are another part of the large ATLAS physics program that is going to benefit from the increased luminosity. Currently, the searches of a weak scale supersymmetry remains one of the key goals of the LHC experiments. With the HL-LHC ATLAS will increase its sensitivity to both R-parity violating and conserving SUSY theories, yielding to the potential discovery of the SUSY counterparts of the SM particles.

With all the above physics goals for ATLAS, the Phase-II upgrade, is clearly an important task and a necessary step towards new discoveries. Therefore, the ATLAS detector in 2024 is planned to undergo its largest upgrade since its construction, which will complement the current Phase-I upgrade to ensure the best possible performance.
3.4 Upgraded sub-systems

To achieve a great performance as in Run-2 ATLAS is planning for the complete re-design or replacement of its sub-systems, along with smaller scale modifications. In more detail a complete redesign of the ATLAS TDAQ system is foreseen along with the replacement of the whole inner detector and changes in the LAr and Tile calorimeters. A more detailed discussion on each of these upgraded sub-systems will be provided in the following sub-sections.

3.4.1 Inner Tracker

The largest planned upgrade project within the Phase-II is the replacement of the complete inner tracker. The currently installed ID in ATLAS was designed and inserted during the construction of the detector (with an extra layer a year after - IBL), providing an outstanding performance throughout the years. Providing precise tracking information for so many years in such a high radiation environment is extremely challenging. In the upcoming runs of HL-LHC the number of collisions will increase dramatically, in comparison to the current values, forcing the existing ID to drop its performance due to the accumulating radiation damage.

The plan for ATLAS is to replace the existing detector with a multi-layer system, where the inner most layers are going to be Pixel detectors, surrounded by the outermost Strip layers. The main goal is to obtain a better track resolution than the existing detector in a more demanding environment. To achieve it, more than 10000 pixel modules are required and 18000 strip modules (which are 8000 more pixel modules and 10000 more strip modules than the current Run-2 ID). The coverage of the new Inner Tracker (ITk) system [94] is planned to reach up to $|\eta| = 4$, with four pixel layers and five strip layers in the barrel $|\eta| < 2.7$ region. For the end-cap region the use of pixel and strip disks is planned. The plan for the disk layout is to use a blend of vertical and inclined layers for the pixel detector to allow maximum coverage in the transition region and vertical disks for the strip layers, as can be seen in fig. 3.4. Changes to the initial

![ATLAS Simulation Preliminary ITk Layout - ATLAS-P2-ITK-23-00-00 $\eta = 1.0$](image)

Figure 3.4: ITk layout indicating the tilted pixel layers to allow maximum coverage [72].
3.4 Upgraded sub-systems

plan are mostly driven by the material of the readout links. Since ATLAS decided to readout the ID with copper links the amount of copper allowed within the detector is limited, in order to avoid having material effects. Optimizing the layout of the detector without compromising the physics gain is an important and daunting task for the ATLAS collaboration.

Another important task for the ITk community is to maintain as many as possible of the existing services of the ID and be able to reuse part of the cabling that is currently installed, as well as the existing cooling system. For this reason many optimization studies on the layout of the detector in the end-cap regions are conducted to identify the best trade-off between performance, coverage and installation.

3.4.2 Calorimeters and Muon system

A smaller scale upgrade is foreseen for the calorimeters and the muon system, since both systems are upgraded during Phase-I.

The expected upgrade for the calorimeters concerns mostly the replacements for the on and off detector electronics. The LAr calorimeter off detector part is replaced during Phase-I and therefore only the on detector electronics will be upgraded during Phase-II. New front and back end electronics are planned to be inserted allowing the full granularity data to be read at the maximum rate of 40MHz. For the Tile calorimeter both on/off detector components are planned to be replaced due to the extensive radiation damage (on) and ageing (on/off) during the past operation years. Additionally, the replacement of about 10% of the photomultipliers in the most exposed cells is foreseen. Those changes will allow the Tile calorimeter to endure the radiation from the HL-LHC, be compatible with the new TDAQ architecture and transfer the full detector information at 40MHz rate.

Regarding the muon system the bulk of the detector upgrade was done during Phase-I with the insertion of the NSWs covering the two end-cap regions. Small detector additions in the barrel region were performed to increase the coverage. Similarly during Phase-II a set of MDT, RPC and TGC will be inserted to increase the MS coverage and reduce the rates of muon triggers. The detectors are planned to be installed in the transition regions of the barrel and the end-cap. The RPCs are planned to be positioned parallel to the beam axis, whereas the TGC and the MDT perpendicular.

3.4.3 The TDAQ system upgrade

The second largest ATLAS upgrade project during Phase-II is the complete redesign of the Trigger and Data Acquisition system. The new design of the TDAQ system is composed of three major sub-system. A dedicated hardware based Level-0 Trigger system, the Data Acquisition system and the Event Filter (EF) which will be a hybrid CPU and hardware system. The Level-0 hardware system is planned to reduce the 40MHz detector readout rate to 1MHz and subsequently the EF system to 10kHz rate for storage. A summary of the expected TDAQ architecture can be seen in fig. 3.5.
3.4 Upgraded sub-systems

Figure 3.5: The ATLAS Phase-II baseline architecture with the new hardware based L0 system and the hybrid EF [95].

**Level-0**

The hardware based L0 system is designed to operate at a maximum rate of 1MHz with a latency of 10µs. The decisions at the L0 stage are going to be made based on the information of the calorimeter (L0Calo) and the muon system (L0Muon). Each dedicated sub-detector system, within the TDAQ, will receive the corresponding data and perform a hardware based reconstruction of the received information.

More precisely the L0Calo system will be based on the existing FEX boards from Phase-I. The firmware of those boards is planned to be updated to cope with the more busy HL-LHC environment and the higher granularity readout of the calorimeter. In addition to the currently existing FEX’s the insertion of a new board dedicated to the reconstruction of the forward region jets and electrons (fFEX) is planned. A summary of the the FEX board, coverage, granularity
and trigger objects can be seen in the following table 3.1.

<table>
<thead>
<tr>
<th>Subsystem</th>
<th>Trigger Object</th>
<th>Approximate Granularity</th>
<th>Coverage [η]</th>
</tr>
</thead>
<tbody>
<tr>
<td>eFEX</td>
<td>$e/\gamma, \tau$</td>
<td>Super Cells (10 in 0.1 x 0.1)</td>
<td>&lt; 2.5</td>
</tr>
<tr>
<td>jFEX</td>
<td>$\tau, jet, E_T^{miss}$</td>
<td>0.1 x 0.1</td>
<td>&lt; 2.5</td>
</tr>
<tr>
<td>jFEX</td>
<td>$\tau, jet, E_T^{miss}$</td>
<td>0.2 x 0.2</td>
<td>2.5 - 3.2</td>
</tr>
<tr>
<td>jFEX</td>
<td>$\tau, jet, E_T^{miss}$</td>
<td>0.4 x 0.4</td>
<td>3.2 - 4.9</td>
</tr>
<tr>
<td>gFEX</td>
<td>Large-R jet, $E_T^{miss}$</td>
<td>0.2 x 0.2</td>
<td>&lt; 4.9</td>
</tr>
<tr>
<td>fFEX</td>
<td>$e/\gamma$</td>
<td>Full detector ECAL, HCAL, FCal</td>
<td>2.5 - 4.9</td>
</tr>
<tr>
<td>fFEX</td>
<td>jet</td>
<td>Full detector FCal</td>
<td>3.2 - 4.9</td>
</tr>
</tbody>
</table>

Table 3.1: L0calo performance, coverage-granularity-trigger objects, summary table [95].

For the muon TDAQ sub-system the Phase-II schedule foresees the complete replacement of the entire muon trigger electronics and readout chain. The expected upgrade is planned to provide the complete information from the barrel RPC systems as well as the TGC detectors from the end-caps. In addition with the information from the NSW and TGC’s more refined algorithms can be used to reconstruct more accurate trigger objects in both barrel and end-cap regions. Furthermore with the addition of the barrel and end-cap Sector Logic boards the combined information of the muon detectors and the outermost layers of the Tile calorimeter will be accessible, for more accurate muon measurements. All the information from the different components of L0Muon will be combined in the $\mu$CTP board and propagated to the Global Trigger board. The performance of L0Muon can be seen in the following table 3.2.

The Global Triggers processor’s main task is to apply offline-like algorithms on the received detector data to reach the highest possible granularity. Additionally, the Global Trigger is responsible for applying topological algorithms, replacing the pre HL-LHC L1Topo, for a better reduction in the trigger rate.

The last sub-system of the L0 chain, responsible to provide the accept event signal (L0Accept), is the CTP [96]. The CTP system combines all the information from the Global Trigger, additional information from the $\mu$CTP as well as raw detector information from forward regions. It also applies the required time for synchronization (dead-time) between the various sub-system data as well as the necessary pre-scales, when needed.

### DAQ system

Following the accept decision from the CTP at the rate of 1MHz, a request is issued to all the detector sub-systems to transmit the remaining event detector data. The reception of all the
detector information is left to the Readout sub-system which contains the FELIX cards and a dedicated hardware board called DataHandler. The main task of the FELIX system is to transform all the specific sub-detector interfaces to a uniform standardized network interface. The FELIX cards are planned to be used during Run-3 as described above. The DataHandler will receive the data from FELIX and will apply all the required modification to the data packets to include any extra monitoring information (e.g. dedicated headers/trailers, Cycle Redundancy Check (CRC), etc.) prior to passing the packets to the Dataflow sub-system.

The Dataflow system is the component of the new TDAQ architecture that will pass the incoming data from L0 to the Event Filter system and will receive its result for the permanent storage. The sub-components of the Dataflow system are the Event Builder, Storage Handler and Event Aggregator. The first sub-system is responsible of building the event with all the information received from the Readout system and manages the requested storage in the Storage Handler. The Event Aggregator is responsible for buffering the data until the EF farm reaches a decision; it then formats the data before recording them to the permanent storage.

**Event Filter**

The final step in the new decision system of ATLAS is Event Filter system. It’s dedicated to applying the final decision on the L0 accepted events as quickly as possible and finally reducing the trigger rate to 10kHz. Optimizing the decision process is also crucial for the precedent components of the L0 trigger (e.g. StorageHandler).

The EF system is implemented as a hybrid system consisting of a CPU based farm and the Hardware Tracking for Trigger hardware system. The CPU component of the system is running mostly optimized versions of the algorithms existing in the current HLT. The most time consuming algorithms are those dedicated to tracking. For this reason the EF system (fig. 3.6) makes use of the Hardware Tracking for the Trigger (HTT) system to perform initially high-

Figure 3.6: The EF system dataflow and decision process with the use of the HTT system [95].
speed tracking in certain regions of the detector (regional), to reduce the trigger to 400kHz. Subsequently the EF algorithms perform finer reconstruction of physics objects received from L0, mimicking the offline algorithm behavior, with the difference that the full detector tracking may be requested from the hardware co-processor (global). In practice when a more elaborate track reconstruction is required quickly the full first stage and second stage tracking chain (called global tracking) is requested which introduces a small latency factor but still remains more efficient than a CPU only solution.

The main motivation for employing hardware is the high occupancy observed in the tracking detectors. With these conditions, pattern recognition in software becomes an extremely time consuming task, bringing the trigger system to exceed the limits of its capacity. For this reason the choice to use the HTT system as a co-processor was made in the ATLAS collaboration. However, at the time of this thesis the ATLAS collaboration decided to reassess whether the HTT system should remain in its current form. Various options are discussed, such as performing tracking as it is now on CPU’s (Run 1-3), making use of commodity hardware like GPU’s and/or FPGA accelerator cards, or redesigning the HTT system in a simpler form. More details about the HTT system architecture are going to be given in the next chapter since it’s the project for which most of the hardware work of this thesis was part of.
In this chapter the tracking system at trigger level from ATLAS during Run-4 is discussed. An overview about the motivation and functionality of the system is provided. A technical description of the individual hardware components and their use is given. Finally the system performance is discussed based on the dedicated framework developed within the HTT community for simulation.

4.1 System motivation and functionality

In ATLAS particle tracking at the trigger level for HL-LHC is planned to be performed in the EF. The difference between the HL-LHC system and the one used for Run-2 and Run-3 is that the EF is currently planning to use a dedicated hardware system for tracking rather than executing it completely in software. The custom electronics system, called HTT, is planned to perform very quick pattern recognition and track reconstruction on data retrieved from eight detector layers (at least one pixel layer). Additionally, a second stage of tracking is possible in HTT combining all the inner detector layers, if more precision is requested from the EF.

The two functionalities of HTT are complementary, with their differences arising from the number of hits required for each, as well as the covered detector region. As stated above the first functionality will perform tracking on 8 detector layers at a rate of the L0-trigger (1MHz), yielding to a 10 percent usage of the ITk data. Additionally, the first type of tracking will be applied only around certain regions of the detector and therefore called regional HTT (rHTT). The second tracking stage will use as seeds tracks found from the first stage along with the remaining ITk layers, at a rate of 100kHz. The region for the second stage tracking is the complete ITk coverage, calling the sub-system global HTT (gHTT).[95]

The main motivation for the HTT system relies on the resources required in software to perform track identification with the ATLAS ITk in a busy environment such as HL-LHC. The most prominent limitations in such a busy environment are: the ability of a system to distinguish real tracks from accidental combinations of hits and to perform track fitting for the helix parameters estimation on the identified tracks in a reasonable amount of time. Using dedicated hardware for those tasks is motivated by factors such as power efficiency, short latency, less de-
manding space requirements and cost effectiveness. Due to the advantages of the HTT system over the alternative options ATLAS decided initially to construct the HTT for Phase-II.

4.2 Hardware tracking theory

The HTT system algorithm can be broken down into steps. The first one is called track finding and concerns the ability of the system to identify real tracks in a busy environment of the ATLAS ITk during HL-LHC. The second category is the track fitting which summarizes the ability of the system to provide an estimation of the track parameters. [97]

4.2.1 Track finding

Every charged particle passing through the ITk layers leaves a signature in a form of pixel or super-strip clusters. The centroid of each cluster is used as the best estimate of the particle's coordinates on each given detector layer. Those clusters are called detector hits, where a number of coordinates are known. Each candidate track will contain a set of hits in different layers \( (C \subset R^n, \text{ with } C \text{ indicating the number of hits belonging to a specific track and with } R \text{ the total number of hits and } n \text{ the total number of detector layers used to measure the track}) \). Not all hits in the ITk originate from a real particle passing through the detector. Therefore, track finding is the ability of the system to decide whether a given sub-set of hits \( (T \subset C, \text{ where } T \text{ the set of hits belonging to a real track and } C \text{ the total number of tracks found}) \) belong to a real track.

Any identified track can be described by a set of parameters \( m < n \) (where \( m \) the number of parameters and \( n \) the number of detector layers, in which case \( m \) is a smaller number of parameters than the number of layers needed to measure the track) \( p \) (e.g. \( p_T, \phi_0, d_0 \)). In the ideal case of perfect resolution the tracks belonging to \( T \) will form an \( m \)-dimensional surface existing within the super-set \( C \). The surface can be described with a set of functions called constraint equations, with the form seen:

\[
f_i(x) = 0, \quad i = 1, \ldots, n - m
\]  \hspace{1cm} (4.1)

The calculations of \( f_i \) can be performed either numerically or analytically and depend on the system geometry. To solve the ideal case of track finding the evaluation of the \( f_i(x) \) function is needed. A track is accepted when the evaluation of the constraint functions is zero.

In the non-ideal case of finite resolution the value of the \( f_i \) functions is different than zero. Schematically the finite resolution translates as the introduction of a thickness into the \( m \)-dimensional surface \( T \). To measure the thickness effect it is important to define the covariance matrix \( F_{ij} \) of \( f_i \), and to simplify the calculations a linear approximation is taken:

\[
F_{ij} \approx \sum_{kl} \frac{\partial f_i}{\partial x_k} \frac{\partial f_j}{\partial x_l} S_{kl}
\]  \hspace{1cm} (4.2)

where \( S \) is the covariance matrix of the coordinates \( x \), indicating the correlation of the hits in each layer (assuming that a set of hits originates from the same track the covariance will show
a strong correlation between the hits from the different layers). Using eq. 4.2 the definition of a \( \chi^2 \) variable is allowed, where applying a cut on it would yield on the selection of "good" tracks with a given efficiency.

\[
\chi^2 = \sum_{ij} \tilde{f}_i F^{-1}_{ij} \tilde{f}_j,
\]

where \( \tilde{f} \) are the normalised constraint functions used to identify the track parameters and \( F^{-1}_{ij} \) the covariance matrix indicating the correlations between the different track parameters. A further simplification of eq. 4.3 can be achieved if the matrix \( F^{-1}_{ij} \) is required to be symmetric and therefore can be diagonalized, yielding to the following expression for the constraint functions,

\[
\chi^2 = \sum_i \tilde{f}_i^2
\]

which requires only n-m multiplications and n-m-1 sums. This turns the problem of realistic tracks into an evaluation problem of constraint functions.

Calculating analytically the constraint functions is generally a daunting task. Therefore, linear approximations can be used which schematically result to a tangent hyper-plane to the \( T \) hyper-surface at a given point of the track parameters \( x_0 \). When the \( T \) hyper-plane is flat the tangent hyper-plane approximation holds perfectly.

\[
\tilde{f}_i \approx \frac{\partial \tilde{f}_i}{\partial x}(x - x_0) = u_i x + c_i
\]

The calculation of the constants in the hyper-plane shown in eq. 4.5 can be either performed analytically with the knowledge of the detector geometry or with a use of a more brute force numerical approach based on the concept of principal component analysis [98]. Thus, using Monte Carlo (MC) samples or even data from real collisions, the constants for the determination of the constraint functions can be calculated and subsequently using eq. 4.4 one can decide to accept or reject a track [97].

For the HTT system track identification is absorbed within the pattern matching operation. MC simulations are used with mostly muon tracks in a high pile-up environment. The tracks that are mostly generated are used to form a finite size track bank (pattern bank), along with their parameters. The pattern bank is loaded in dedicated devices called Associative Memory (AM) chips [99] which are used to quickly compare the incoming data hits against the pre-computed simulation tracks. Since the available memory within an Application Specific Integrated Circuit (ASIC) is finite, a pattern compression method is performed on tracks that share a large amount of hits and can be combined with a small penalty on the resolution (implemented with Don't Care bits). If an incoming set of hits matches a pre-computed track in the pattern bank the operation described in the next section is performed: track fitting.

### 4.2.2 Track fitting

The following problem for a hardware tracking system is to identify the track parameters on the set of found tracks. To do this, the same method as the one described in eq. 4.5 for the
4.2 Hardware tracking theory

constraint function constants is applied. Essentially, the parameters of any set of tracks can be expressed as a function of cluster co-ordinates with the following equation:

\[ p_i = p_i(x) \]  \hspace{1cm} (4.6)

where \( x \) are the \( m \) known coordinates of each cluster. By applying a linear approximation of those parameters the following formula can be retrieved

\[ p_i \approx w_i(x - x_0) + p_i(x_0) = w_i x + q_i \]  \hspace{1cm} (4.7)

which reduces the problem of track fitting to the calculation of the constants \( w_i \) and \( q_i \). The main difference in the used methods between track finding and track fitting is that for the latter the parameters are estimated by minimizing the average quadratic terms in eq. 4.7 which yield the minimization of the resolution on the found parameters. The minimization result produces the following equation for the \( w_i \) constants:

\[ w_i = M^{-1} \gamma = M^{-1} \sum_i \tilde{p}_i x - \sum_i \tilde{p}_i \sum x / N \]  \hspace{1cm} (4.8)

where \( \gamma = \langle \tilde{p}_i x \rangle - \langle \tilde{p}_i \rangle \langle x \rangle \), and \( M \) is the covariance matrix. From eq. 4.8 the definition of the remaining constants can be defined with the following equation:

\[ q_i = \langle \tilde{p}_i w_i x \rangle \]  \hspace{1cm} (4.9)

Another difference in the calculations for the two methods is that the constants in the track fitting method require known tracks and therefore cannot be determined on data as with the track finding approach. For this reason a MonteCarlo simulation is performed which produces limitations that are not present when the calculations happens with real data [97].

For the HTT system the required constants for the track fitting method are stored in memory and generated from simulations with muon tracks, with the same distribution as for the pattern bank, for each detector region. A set of incoming clusters that was successfully matched with a specific pattern uses the pre-computed constants \( w_i, q_i \) for each space point. A \( \chi^2 \) variable is used to identify the best set of constants that describe the found track. Track fitting is planned to be performed within an FPGA taking advantage of the parallel processing capabilities of it.

4.2.3 Common issues in hardware tracking

The calculations described above are vulnerable to various effects (e.g. resolution, multiple scattering etc.). On of the most prominent real life effects are detector misalignments that are present. These would affect both the track finding procedure, and track fitting.

On the track finding side, adding misalignments into the equations is effectively translated as perturbations on the kernel matrix of the constraint functions. These perturbations affect the track finding by increasing the amount of "fake" tracks found by the system, since the pre-computed pattern bank will not completely represent the incoming data; This translates to an effect of worsening the resolution on some eigenvalues of the matrix.

On the track fitting side; detector misalignments affect the fitted parameter resolution. The
track parameter calculation on the found tracks is skewed compared to the perfect case (reduced resolution), which means that the $\chi^2$ values will increase (reduced fit quality).

For both algorithms correcting for the misaligned modules produces a better performance for the system; however, it remains a very challenging task that needs to be performed in simulation [97]. Therefore a good understanding of the detector geometry and potential damage factors are required.

In the following section a more detailed discussion will be provided on how hardware tracking is planned to be performed with the use of the HTT system. In addition a discussion of its performance will be given.

## 4.3 HTT Technical description

The HTT is a system completely based on custom electronics boards and organized into units called HTT Units. An HTT Unit corresponds to a fully populated 14-slot Advanced Telecommunications Architecture (ATCA) shelf [100]. The input data of the HTT system are received via optical fibers at 10Gbps from the EF, with a dedicated protocol developed for the system. In total 48 HTT Units are required in order for both the functionalities of rHTT and gHTT to be accommodated. A blend of dedicated boards in each HTT Unit consists of 12 first stage boards and 2 second stage. The expected throughput for rHTT is about 2MB per event at a rate of 1MHz with an addition of 1.5MB for gHTT when needed at a lower rate of 100kHz. A summary of the HTT Units components can be seen in the following fig. 4.1 [95].

During the writing of this thesis, ATLAS started a re-assessment procedure to optimize the layout of the HTT system. Two main alternatives are discussed, with the first one being based on commodity hardware (e.g. FPGA accelerators) and the second one based on a single ATCA card hosting the complete functionality of the HTT system. The main driving force for the re-optimization procedure was initiated due to ATLAS not implementing tracking at the hardware based trigger level, hence most of the latency constraints imposed an the HTT system are lifted. Since the decision process is internal to ATLAS no further discussion on those two alternative systems will be provided, and the focus will remain on the HTT system.

### 4.3.1 HTT Interface

As mentioned in the previous sections (sec. 3.4.3) the use of the HTT system will be requested from the EF, which also provides the necessary data for tracking. For this reason a dedicated interface is developed called HTT Interface (HTTIF) (seen in fig. 4.1) [101]. The mapping between the ITk modules and the HTT Units covering a specific detector region will be available at the EF level allowing the HTTIF to transmit data to the correct unit or units following the EF requests. If the same copy of data needs to be transmitted to more units, the HTTIF will perform it automatically. In addition HTTIF will be responsible for receiving the found tracks from each unit and transmitting them back to the EF for further processing [95].

The implementation of the HTTIF is based on the FELIX hardware. Its functionality will be to
4.3 HTT Technical description

Figure 4.1: HTT System organization in HTT Units with the data flow via the HTTIF. With the orange colour the Track Processor (TP) boards which depending on whether they host a Pattern Recognition Mezzanine (PRM) or a Track Fitting Mezzanine (TFM) are called either Associative Memory Track Processor (AMTP) or Second Stage Track Processor (SSTP) respectively. In addition with the green colour the ATCA crate is shown which is hosting a specific number of AMTP and SSTP defining the HTT Unit. [95]

receive the data over a Peripheral Component Interconnect Express (PCIe) interface from the EF and it will have to buffer and aggregate the received data on an event by event basis. Additionally, the transmission of the received data towards the HTT Units will be over a dedicated fibre link and formatted into a low level protocol acceptable by the FPGAs used in HTT. [102]

4.3.2 Track Processor

The board responsible for receiving the incoming data from HTTIF will be the TP [103]. The TP board will be an ATCA card hosting a large FPGA chip. The HTT system will deploy mezzanine cards (attachable cards on top of host board) for processing the data, hence the TP will contain the connectors towards those mezzanine cards. Depending on the type of mezzanine the TP can either be an AMTP (seen in fig. 4.1), focused on the first stage tracking functionality of HTT, or SSTP (seen in fig. 4.1) dedicated to the second stage tracking.

The FPGA on both type of TPs will interface, via HTTIF, with the EF CPUs, perform clustering on the incoming data and identify the hits for the track finding and fitting algorithms (implemented on the mezzanines). In addition the TPs will be responsible for reliable data distribution within the HTT-Unit either via the high-speed ATCA back-plane or via the Rear-Transition Module (RTM) to other HTT-Units. A summary of the TP FPGA implementation can be seen in fig. 4.2 [103].
4.3.3 Pattern Recognition Mezzanine

As mentioned above the HTT system will perform tracking in two stages. Distinguishing the two cases is based on the amount of information used for the track finding and fitting. The first stage which will be used always as part of both rHTT and gHTT, will be performed in a dedicated mezzanine card called the PRM [104].

The PRM will perform track finding based on the theoretical description (4.2.1, 4.2.2), with the use of dedicated ASICs and track fitting in a firmware implementation. In more detail the PRM will receive data from the 8 ITk layers, will store the constraint function constants derived from the simulation, will perform the track fitting on the candidate tracks (Roads), using the equivalent of eq. 4.7 and return the track parameters on the found tracks that pass a $\chi^2$ cut, which evaluates the quality of the fit. Since the PRM is the most time consuming part of the system an efficient loading of data and organization of the required components for the fits is required [102]. The mount of a PRM in a TP is shown in fig. 4.1 as an AMTP. A summary of the internal functionality of each PRM is shown in the following fig. 4.3.

Figure 4.3: The PRM FPGA implementation along with the dedicated interfaces for all the required functionalities. With the term SSID the HTTs found by the TP are described; with RoadID the resulting found tracks on a given set of SSIDs are described. [104].
4.3.4 Associative Memory chip

Track finding in HTT will be based on custom ASIC which are called AM chips [99]. The AM chips are based on the Content-Addressable Memory (CAM) technology [105] allowing high-throughput, parallel pattern matching at a very high rate. In more detail the AM chip stores in layers (8 for the PRM) pre-computed hits generated in simulations. Those tracks are called patterns and are matched in real-time with the incoming hit data. The AM chip differs from typical memory access: instead of giving the memory location of a certain pattern, it gives a binary response accord to whether a pattern exists in a specific set or not. This exact difference between CAM and RAM makes the AM being a perfect candidate for high-speed pattern matching to identify the required tracks for the track finding algorithm in HTT. The outgoing data words from the AM chips are called Roads and are used as address pointers for a High Bandwidth Memory (HBM) which will store the incoming clusters, identifying whether a set of clusters corresponds to a real track or not. Subsequently the found track will be fitted as described in the theory subsection above (4.2.2).

The first AM chips were developed in the early 80s and put into operation at the SVT project of CDF [97]. The HTT system will use the 9th version of the AM chip which will be able to store a maximum of 384k patterns per chip, in order for HTT to have the required efficiency in a given region (efficiency derived from simulations). Each PRM will host 12 AM chips covering in total in each HTT Unit a 0.2x0.2 $\eta - \phi$ region of the detector [95].

4.3.5 Track Fitting Mezzanine

The last remaining hardware component of the HTT system is the TFM [106]. The TFM is a mezzanine card responsible for the second stage tracking with the remaining ITk layer of the detector. Each TP dedicated to second stage tracking in an HTT Unit will mount two TFM’s, and is thus called SSTP.

The TFM functionality differs from the PRM because instead of using the AM chips to perform pattern matching on raw clustered data (Hits), it receives fitted 8 layer tracks from the first stage tracking performed on the AMTPs and extrapolates into the remaining layers. This is the main reason why the TFM does not require any custom hardware component; the number of hit combinations for the remaining layers given the 8 layer track is much smaller, hence its functionality is completely implemented in an FPGA.

The two main blocks of the firmware are the Extrapolator which will extrapolate the 8 layer tracks found from the PRM into the remaining ITk layers using a linear approximation. The second main firmware block is shared between the two mezzanines and is the track fitting block that will apply the same track fitting algorithm as for the PRM in order to identify the track parameters for a candidate track. Finally, a $\chi^2$ cut is applied to identify the fit with the best quality [106].
4.4 Performance using simulation

An important step in the system implementation is the ability to test its performance prior to construction, in addition to optimising the various aspects of it (e.g. pattern bank, track constants). In HTT the simulation of the system was implemented with the development of a new framework called HTT Simulation (HTTSim) that is based on the Athena [107] framework. Currently HTTSim is a standalone software package derived from Athena and is not part of the official release used in ATLAS. The main goals for HTTSim are to define the performance of the different system blocks prior to implementation, produce the patterns for the AM chips, calculate the fit constants for the track fitting step and provide the expected resolution of the track parameters calculated by the HTT system. In addition, the simulation framework will be used to estimate the system performance on the various analyses for which performance depends extensively on tracking (e.g. Higgs signatures).

4.4.1 HTTSim overall use

For both track fitting stages implemented in hardware, HTTSim is an important tool for their configuration. The first use of the HTTSim is to optimize the patterns stored in the AM chips due to the limitation from the memory size. Since AM chips have a capacity limit, a $p_T$ cut is applied on the patterns to reduce their numbers. This process yields to the following two categories:

1. **1GeV-banks**: $p_T > 1\text{GeV}$, 3.75M patterns (gHTT)

2. **2GeV-banks**: $p_T > 2\text{GeV}$, 2M patterns (rHTT)

Since gHTT depends on patterns found from the first stage an implicit overlap in the number of patterns in the AM chips is present [102]. Additionally, the HTTSim is used to define the coverage of each HTT Unit, as well as the total system coverage in terms of $\eta$. An HTT Unit covers a specific region of the detector $\eta \times \phi$, with the corresponding detector modules, $p_T$ above a certain threshold and $|z_0| < 15cm$ and $|d_0| < 2mm$ (where as $z_0$ the difference in the z-coordinate between a track and the Primary Vertex (PV) is defined and with $d_0$ the track impact parameter with respect to the PV). These requirements allow HTT regions differ from standard RoI but also overlap in terms of detector elements used in each. The complete coverage of HTT is defined based on the specific ITk layout. Since the studies of the ITk layout are still progressing HTT is planned to provide a coverage up to $|\eta| < 4$ for a specific geometry [102]. Due to the changes in the ITk geometry and the available detector modules, studies with the HTTSim are on-going in specific $\eta$ regions which are representative enough to study the overall performance.
4.4 Performance using simulation

4.4.2 HTT estimated performance

The studies on the performance of the HTT system with the HTTSim require some well understood physics samples. To study the system performance on signal-like events, single muon, electron and pion samples are used with a $p_T$ range of 1-400GeV. For the background samples minimum-bias events are used with a pile-up of $<\mu>=200$ and filtered di-jets events with a parton $p_T>30\text{GeV}$ in the $\eta-\phi$ region under study [102].

Since the HTTSim is a software simulation of the HTT system but not an emulation of all the hardware components, a set of assumptions is made for these studies. Those assumptions mostly concern the simulated event generation itself, the ITk layout and the pile-up description. For this reason all the results provided by the simulation include an extra uncertainty factor of 30% (conventionally used at CERN for unknown uncertainties during a system design) to account for any discrepancy between simulation and actual hardware. When the HTTSim modules are included in the full ATLAS simulation framework the contingency margin of the quoted results will drop to 10% with additional margins derived from extrapolations and unexpected contributions, since certain assumptions needed for the initial design will be lifted from the more elaborate simulation within the Athena framework.

In the following two sub-sections the performance of each HTT component is given along with the resolution obtained from HTT after the first stage tracking.

Pattern-matching

Pattern matching is one of the key tasks performed by HTT. Accurately simulating its performance and efficiency is crucial for the functionality of the full system.

To measure the performance in a given $\eta-\phi$ region two main factors need to be calculated. The first one is the ability of the system to match a "real" single muon signal event, within a busy environment. The second is the average number of hit combinations produced. The former metric mainly describes the discriminating power of tracking applied in hardware, while the latter the average load expected for the subsequent system algorithms. To measure both factors two different type of samples are required. For the efficiency on single muon events, a signal sample with single muons overlaid with pile-up 200 events is used. For the average number of hits produced, a minimum bias or a jet event sample is used.

In the following table 4.4 a summary of the pattern matching performance is provided. Along with the numbers representing the performance of pattern matching in various regions, the size of the Super-Strips are shown, since this is a factor affecting the discriminating power of the system. Additionally, the information on the Don’t Care bits [99] is shown, indicating the level of compression applied on the pattern bank.

Track parameter resolution post first-stage tracking

The second important task for HTT is the determination of the resolution of the track parameters on the found tracks. The best possible resolution on those parameters provides more
accurate measurements of the particle properties and hence allows the trigger to select events with a better quality. To determine the resolution on the track parameters calculated by the system, the residual distribution is calculated for the parameters between the best $\chi^2$ fit and the truth candidate in a 10k single lepton sample (Generated with the Pythia8 generator, with single muons in the final state). The pile-up in the sample was set to 200 as expected during HL-LHC. What is quoted as resolution on the track parameters is the 95% interval on the Root Mean Square (RMS) distribution normalized to the expected width of a unit Gaussian. Unfortunately, at the time of these studies no official results were obtained by HTTSim on the track parameter resolution. Therefore to estimate the performance of the system two different frameworks were used, developed for similar purposes in past ATLAS projects. The first one is the FTKSim which was developed for the FTK system and the later one is the L1TrackSim which was developed for the possible inclusion of tracking with the same hardware at the stage of L1 in the trigger [102]. Both results were modified according to the changes and assumptions imposed by the HTT system.

### Figure 4.5: Track parameter resolutions estimated with two existing frameworks prior to HTT-Sim.

<table>
<thead>
<tr>
<th>$\eta$ range</th>
<th>muon eff.</th>
<th>mean roads in pile-up</th>
<th>99% interval roads in pile-up</th>
<th>super-strip width</th>
<th>DC bits</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.1 &lt; \eta &lt; 0.3$</td>
<td>99.6%</td>
<td>166</td>
<td>590</td>
<td>33×402</td>
<td>40</td>
</tr>
<tr>
<td>$2.0 &lt; \eta &lt; 2.2$</td>
<td>99.2%</td>
<td>40</td>
<td>131</td>
<td>16×200</td>
<td>10</td>
</tr>
</tbody>
</table>

**Figure 4.4:** Pattern matching performance estimated from HTTSim [102].

### Figure 4.5: Pattern matching performance estimated from HTTSim [102].

**Table**: Parameters used to calculate the resolution on the track parameters.

<table>
<thead>
<tr>
<th>$\eta$ range</th>
<th>$\eta$</th>
<th>$\phi$</th>
<th>$q/P_T$ [GeV$^{-1}$]</th>
<th>$d_{\eta}$ [mm]</th>
<th>$z_0$ [mm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.1 &lt; \eta &lt; 0.3$</td>
<td>0.003</td>
<td>0.002</td>
<td>0.006</td>
<td>0.42</td>
<td>0.9</td>
</tr>
<tr>
<td>$2.0 &lt; \eta &lt; 2.2$</td>
<td>0.003</td>
<td>0.002</td>
<td>0.011</td>
<td>0.39</td>
<td>6.7</td>
</tr>
<tr>
<td>$3.0 &lt; \eta &lt; 3.2$</td>
<td>0.004</td>
<td>0.004</td>
<td>0.065</td>
<td>0.17</td>
<td>2.2</td>
</tr>
</tbody>
</table>

**Table**: Parameters used to calculate the resolution on the track parameters.

### 4.5 HTT operation services

In addition to the hardware components and the HTTSim framework the HTT system includes some extra services that will be used during commissioning and operation. Since the HTT system is going to be a hardware component in the ATLAS TDAQ architecture the use of DCS is foreseen. Additionally from the studies performed in this thesis the system status monitoring interface was chosen as IPbus [108] (discussed in detail in the following chapter) via a standard network.

The DCS system in HTT will be used for hardware-related controls and will pass through a dedicated System on Chip (SoC) that will be mounted on the TP board. In practice all hardware related information like temperature sensors, DC/DC converter voltage will be read via dedic-
4.5 HTT operation services

ated lines from the SoC and then published over standard network to the DCS infrastructure. To monitor the system data-flow and configuration the IPbus interface was chosen to be used in HTT. Each, firmware block in the system will be separated by a common block between all boards called SpyBuffer. In practice the SpyBuffer will be a minimal latency memory block which will be accessible via the IPbus interface and will contain a real-time copy of the system data. In case of an error the content of the SpyBuffer can be frozen with certain conditions and the data causing the issue can be read out and investigated by the software. Since the SpyBuffer will be a combination of a block memory and a FiFo (interfacing to IPbus) an example implementation of the standard Xilinx FiFo as an IPbus slave was performed as part of this thesis.

The software dedicated to handle the monitoring data (and in general for the HTT system during operation) is the so-called HTT online software, which will be part of the larger Run Control software used by ATLAS. It’s a modular C++ based framework which will contain all the required information about the system during operation and also allow the Run Control shifter to configure and operate HTT remotely during data taking. It will contain the software part of IPbus and will provide debugging information to the user as part of the general ATLAS Run Control software tree application.

Since the monitoring and configuration will be performed via the IPbus protocol, a detailed outline of the interface is given in the next chapter. In addition IPbus was one of the key ingredients for the hardware tests performed for this thesis.
The IPbus protocol was chosen to be the main monitoring protocol for the HTT system. In this chapter a general overview of the interface is provided. Initially, the description of the IPbus protocol is given, followed by the individual system components implemented in software and firmware. The description of a specific IPbus packet is discussed followed by the allowed transactions described in the specifications. Finally, some applications of the IPbus protocol are shown, with a back-of-the-envelope calculation of the IPbus usage in the HTT system.

5.1 Protocol overview

Modern High Energy Physics TDAQ systems are transitioning from the VME [109] technology towards systems that use ATCA and micro-TCA (µTCA - smaller crate size ATCA) technologies. The main difference between the VME and ATCA technologies is that for VME, dedicated specifications are provided as a hardware access protocol for board monitoring. For ATCA based technologies only the serial links are clearly defined, but nothing specific on the protocol side, allowing plenty of room to define custom protocols based on existing ones inspired from industry.

The IPbus protocol is a 32-bit data transfer protocol based on the general IP protocol created specifically for monitoring ATCA based boards. It consists of a firmware implementation in VHSIC Hardware Descriptive Language (VHDL) and a corresponding software implemented based on uHAL [110] libraries. The main implementation provided by the IPbus developers concerns only Xilinx FPGAs; however Intel/Altera implementations are possible. The protocol assumes a 32-bit wide address and data virtual bus to transfer data reliably from a PC to a FPGA based board. Each board that contains an implementation of IPbus will have a unique IP-address in the network, which the software uses to address it. Additionally, a clearly defined set of operations handled by the protocol are provided: Read, Write, Read-Modify-Write bits and Read-Modify-Write sums [111], which are discussed in detail below.

The original implementation of IPbus transfers data between a PC and the hardware board over the standard Ethernet [112] protocol. Therefore, a simple User Datagram Protocol (UDP)/IP [113] core is provided fully developed in VHDL, without any dependence on Xilinx specific IP-
cores. A simpler UDP implementation is preferred over the overhead of a Transmission Control Protocol (TCP) one. In simple network topologies UDP is a reliable solution if a sequential packet ID method is included in the packet header to recover dropped packets [108]. The structure of IPbus does not strictly depend on Ethernet. Any transfer protocol can be used to pass data into the hardware boards via a specific port. For instance, a second implementation of IPbus is provided which utilizes the PCIe interface. In contrast to the Ethernet version, the PCIe depends heavily on Xilinx specific IP blocks for encoding and decoding the frames.

5.2 Firmware and software components

IPbus consists of the following three components, the first one is the firmware block for the end-user hardware of the protocol, which is completely developed in VHDL. The second component of the system is implemented completely in software called the ControlHub and is dedicated to include the reliability mechanism information in the UDP packet in addition to any IPbus specific information. The last system component is the \( \mu \)HAL based software implemented either in C++ or Python and contains specific end-user commands to perform the IPbus allowed transactions.

The system modularity allows easy integration of IPbus as part of any project e.g. a monitoring interface, since the user has only to interface any custom hardware firmware block with a simple 32-bit data word on the firmware side and develop dedicated software on top of specific IPbus libraries on standard software languages.

5.2.1 IPbus firmware block

The firmware component was designed to provide a simple and low-resources usage block that can be integrated along with any user-logic. The block can be separated into two sub-blocks that are dedicated to different tasks.

The first firmware sub-block concerns the stripping of the Ethernet/PCIe packet payload in order to extract the IPbus packet. As a key aspect is to maintain low-resources, the transmission protocol cannot be resource demanding. For this reason on the Ethernet based implementation the use of the UDP protocol was chosen over the TCP (which by construction enhances features that are not strictly needed). On the PCIe side the complication of the frame is much simpler than the corresponding Ethernet frame, therefore no concern for the resources is needed.

The second firmware block contains the IPbus specific decoding. After extracting the payload from the incoming packet, further decoding is needed in order to extract the required 32-bit address and data words provided by the user. The 32-bit address and data words are subsequently used to access any slave implemented in the firmware.

Finally, for the Ethernet version there are another set of services optionally included and accessed through dedicated protocols. For instance an ICMP [114] block is implemented to allow the ping operation between the PC and the FPGA. The extra services are part of what is called
5.2 Firmware and software components

the full IPbus implementation and of course require more resources in comparison to the minimal implementation, which consists just of the two blocks described above [108]. A summary of the resources required is given in fig. 5.1 which shows the number of flip/flops, LookUp Tables (LUT) and BRAMs in absolute numbers estimated on a Virtex-7 FPGA from Xilinx.

<table>
<thead>
<tr>
<th>Element type</th>
<th>Minimal configuration</th>
<th>Fully-featured</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flip flops</td>
<td>2000</td>
<td>3500 (0.4 %)</td>
</tr>
<tr>
<td>Slices</td>
<td>1000</td>
<td>1078 (1.0 %)</td>
</tr>
<tr>
<td>Block RAMs</td>
<td>5</td>
<td>17 (0.6 %)</td>
</tr>
</tbody>
</table>

Figure 5.1: IPbus core resources in a Xilinx Virtex-7 (XC7VX690T) FPGA for both the minimal and full implementation [108].

5.2.2 IPbus software component

On the software side the IPbus protocol requires two separate components which are dedicated to different operations of the interface. Both components are implemented completely in software and assume 32/64bit Linux based machines.

**ControlHub**

The ControlHub is a software application that serves as a single point of access for IPbus to control each device. Via the ControlHub multiple applications are able to communicate with devices that contain the IPbus firmware block. In addition it implements an extra reliability mechanism to avoid packet dropping which is not covered by the UDP interface. The smooth operation of the ControlHub is key for the reliability of the IPbus interface. Therefore, it has to be equivalently reliable, transparent (as the VME crate controller, in terms of how packages are formed and encoded) and avoid crashes and failures. A further functionality for the ControlHub, is to implement simultaneous multi-client to multi-device communication with an efficient method.

For achieving those constraints, the ControlHub communicates with any client monitoring application via TCP/IP, for the best possible reliability. To reach a high level of transparency and efficiency an Erlang [115] based implementation was chosen, which allows the concurrent processing of multiple requests.

**µHAL**

The highest level component of the IPbus interface is the µHAL library. This is a Hardware Access Library (HAL) providing an end-user C++/Python API for the IPbus operations. µHAL is responsible for implementing any software based application defined by the user and for queuing the requested transactions for specific IPbus slaves to be sent towards the ControlHub based on a dispatch system.
In practice the µHAL application maintains a set of xml files that define the specific 32-bit addresses of the hardware implemented slaves and the IP address for the different devices. Each xml entry has a unique address as well as a unique string to be identified by the software, for each IP address. Any user-application gains access to the xml entries, and by referencing the specific string the user can define the desired IPbus operation for the selected slave.

5.3  IPbus packet structure

For the HTT system IPbus is planned to be used over the standard network with the use of the Ethernet protocol. Thus the protocol structure described in this section will be assuming the Ethernet frame structure for a UDP/IP based communication.

All the IPbus specific information is included in the payload of the UDP packet (application level). The UDP packet is wrapped subsequently in the IP packet level which is finally included in the Ethernet frame, shown in fig. 5.2.

In standard UDP based communication the maximum size of a packet is 1.5 kB. Removing the

Figure 5.2: Ethernet packet structure with the IPbus packet as a part of the UDP protocol payload [111].

IP specific header (20 bytes) and the UDP header (8 bytes), yields to a maximum IPbus packet size of 1472 bytes, which is the equivalent of 368x32-bit words. The first 32-bits of these are dedicated to the so called IPbus header which is responsible for the protocol reliability and all the remaining words are assigned to the specific IPbus operation. Performing larger size transfers is possible, however a split into multiple Ethernet packets in software is required, decreasing the protocol throughput.

5.3.1  Packet header

The IPbus header is a quick method introduced in order to ensure the reliability of the protocol and keep the firmware resources under control (UDP over TCP decision). The header is added as the first 32-bit word in the UDP payload from the ControlHub and cross-checked for its validity in the firmware. Any packet that is not confirmed by the decoder part in the firmware is dropped silently.
The breakdown of each bit’s functionality in the 32-bit word can be seen in fig. 5.3. Starting from the most significant byte, IPbus checks if the version of the firmware and the software are matching and assumes the same packet structure.

The next byte is reserved and then followed by a two byte word called the PacketID. These bytes are used by the system to ensure that packets within the same transaction are received and sent properly. An example of the use of PacketID can be seen in fig. 5.4, showing how enforcing both firmware and software to maintain PacketIDs allows the recovery of any potential losses.

The next to last byte of the header is called the Byte-order qualifier. The use of this byte must be combined with the most significant byte, allowing the system to define the endian-ness of the 32-bit words in the IPbus packets. The IPbus protocol can handle both type of endian-ness (big/little) but it is important for the PC and the firmware block to agree prior to any data transmission.

The last byte inside the header is dedicated to defining the type of packet that is followed. There

![Figure 5.3](image1.jpg)  
Figure 5.3: IPbus packet header with a byte breakdown, starting from left bit-31 to right bit-0 [111].

![Figure 5.4](image2.jpg)  
Figure 5.4: IPbus reliability mechanism for single packet in flight [111].
are several options for packets supported by the IPbus protocol and therefore it is important for
the firmware to know how to handle the 32-bit words received in the payload. For example, with
the use of 0x0 the firmware handles the 32-bit words as transaction specific words and decodes
them as slave addresses and slave data words (Control packet).

5.3.2 Control packets

Depending on the value of the least significant byte in the IPbus header, various options for
packets are allowed. The most used packet type is the IPbus Control Packet which contains the
transaction requests from software.

Each IPbus Control Packet consists of a 32-bit word, called the transaction header and a num-
ber of requested IPbus operations, from software. The transaction header is used as prefix in
all requests with a non-zero value and followed by incremental values to ensure the transaction
order. If a header is not matched between the target and the client, the packet is dropped.

**Transaction Header**

The transaction header has a breakdown of bits as shown in fig. 5.5. Starting from the most
significant bits of the header, the use of these bits is explained.

The most significant half-byte is used to define the version of the protocol. The reason why
those bits are still in place are mostly historical between the current and older versions of the
IPbus protocol. In the most recent implementation (v2.0), IPbus does the version matching as
explained above at the level of the IPbus Packet Header. However, a secondary cross check is
performed in each transaction header.

The following 12 bits in the header are called the Transaction ID. It’s used by both the client
and the target to verify that a specific request is connected to a given reply and vice versa. The
main motivation for having 12-bits for the Transaction ID is to provide IPbus with the potential
to include Jumbo Frames (extensive packet size). In this way no transaction will end up with a
wrapped around Transaction ID.

The next byte is used as a counter to the number of 32-bit words that are needed to perform the
given transaction. In practice, it describes the number of words that have either to be written
in a given slave or read from it. Any write/read operation that exceeds the maximum packet
size of 255 words (without Jumbo Frames) needs to be split into multiple similar transactions
on the same slave.

Finally the last upper 4-bits of the least significant byte are used to select the IPbus operation

<table>
<thead>
<tr>
<th>31</th>
<th>28</th>
<th>27</th>
<th>16</th>
<th>15</th>
<th>8</th>
<th>7</th>
<th>4</th>
<th>3</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protocol Version (4 bits)</td>
<td>Transaction ID (12 bits)</td>
<td>Words (8 bits)</td>
<td>Type ID (4 bits)</td>
<td>Info Code (4 bits)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 5.5: IPbus transaction header in a control packet [111].
(read/write/rmw) and the lower 4-bits are used to provide information about potential failures. The allowed values are summarised in the following fig. 5.6.

<table>
<thead>
<tr>
<th>Info Code</th>
<th>Direction</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>0x0</td>
<td>Response</td>
<td>Request handled successfully by target</td>
</tr>
<tr>
<td>0x1</td>
<td>Response</td>
<td>Bad header</td>
</tr>
<tr>
<td>0x2</td>
<td>n/a</td>
<td>Rsvd.</td>
</tr>
<tr>
<td>0x3</td>
<td>n/a</td>
<td>Rsvd.</td>
</tr>
<tr>
<td>0x4</td>
<td>Response</td>
<td>Bus error on read</td>
</tr>
<tr>
<td>0x5</td>
<td>Response</td>
<td>Bus error on write</td>
</tr>
<tr>
<td>0x6</td>
<td>Response</td>
<td>Bus timeout on read (256 IPbus clock cycles)</td>
</tr>
<tr>
<td>0x7</td>
<td>Response</td>
<td>Bus timeout on write (256 IPbus clock cycles)</td>
</tr>
<tr>
<td>0x8</td>
<td>n/a</td>
<td>Rsvd.</td>
</tr>
<tr>
<td>0x9</td>
<td>n/a</td>
<td>Rsvd.</td>
</tr>
<tr>
<td>0xa</td>
<td>n/a</td>
<td>Rsvd.</td>
</tr>
<tr>
<td>0xb</td>
<td>n/a</td>
<td>Rsvd.</td>
</tr>
<tr>
<td>0xc</td>
<td>n/a</td>
<td>Rsvd.</td>
</tr>
<tr>
<td>0xd</td>
<td>n/a</td>
<td>Rsvd.</td>
</tr>
<tr>
<td>0xe</td>
<td>n/a</td>
<td>Rsvd.</td>
</tr>
<tr>
<td>0xf</td>
<td>Request</td>
<td>Outbound request</td>
</tr>
</tbody>
</table>

Figure 5.6: IPbus info code values within the transaction header [111].

**Read transaction**

The simplest transaction for IPbus is the read operation. For the read operation on a single 32-bit register, what is required for the successful transaction is a packet that contains a 32-bit transaction header and 32-bits of the address word to indicate the accessed register to be read. On the response side the same transaction header is copied, followed by the value of the register that was requested. The summary of this operation with the individual bit breakdown can be seen in fig. 5.7.

The BASE_ADDRESS is the 32-bit address of a given slave as defined for the software in the xml files that are accessed by the μHAL libraries. If multiple addresses in the same slave need to be accessed then the BASE_ADDRESS is incremented automatically by the software. This type of operation is normally called a block read operation and is applied when accessing data stored in a block RAM memory. If data are stored in a FIFO or in a single register, the same

---

**Read transaction**

The simplest transaction for IPbus is the read operation. For the read operation on a single 32-bit register, what is required for the successful transaction is a packet that contains a 32-bit transaction header and 32-bits of the address word to indicate the accessed register to be read. On the response side the same transaction header is copied, followed by the value of the register that was requested. The summary of this operation with the individual bit breakdown can be seen in fig. 5.7.

The BASE_ADDRESS is the 32-bit address of a given slave as defined for the software in the xml files that are accessed by the μHAL libraries. If multiple addresses in the same slave need to be accessed then the BASE_ADDRESS is incremented automatically by the software. This type of operation is normally called a block read operation and is applied when accessing data stored in a block RAM memory. If data are stored in a FIFO or in a single register, the same
5.3 IPbus packet structure

operation is called a non-incremental read since only a single BASE_ADDRESS is accessed in each transaction.

Write transaction

The second transaction performed with IPbus is the write operation. The methodology of this operation is largely the same as the read operation, but as seen in fig. 5.8 the request and response formats are swapped along with the choice of the BASE_ADDRESS.

When the user needs to write a specific location then a pair of 32-bit words is to identify which slave will be accessed and a second 32-bit word for the content that will be written on the chosen slave. For this reason, following the transaction header word is the BASE_ADDRESS chosen by the user, followed also by the specific value required to be written.

If multiple sequential locations need to be written then the equivalent of a block read can be performed. In practice this translates to having each second 32-bit word in the Control Packet to be treated as the BASE_ADDRESS value incremented automatically. If a single write needs to be performed only a single pair of 32-bit words is required.

Since this operation is only writing data to hardware, only the transaction header is provided as a response, to validate that the operation was completed successfully.

Read/Modify/Write transaction

Along with the write/read operations, IPbus foresees a set of more complicated transactions called Read/Modify/Write (RMW) bits and sums transactions. The two operations are effectively used for setting/cleaning bits on a register and for adding a specific value to the existing content, with the use of single transaction. The equivalent of both RMW operations can be performed with the use of the standard read/write transactions but that would then translate to passing two response/requests and effectively doubling the amount of transaction IDs.

The packet for the request/reply of both operations can be seen in fig. 5.9 and fig. 5.10 which show, how the IPbus packet is organised. The BASE_ADDRESS of the accessed register is provided to identify the accessed register and a set of commands in order to perform AND/OR or AP-
5.4 IPbus topologies

PEND operations to the register value. In both cases the outcome of the operation is added to
the reply packet to prove that the value of the register was updated accordingly.

\[
\begin{array}{c|c|c|c|c|c}
31 & 24 & 23 & 16 & 15 & 8 & 7 & 0 \\
\hline
\text{Word 0} & \text{Version} = 2 & \text{Transaction ID} & \text{Words} = 1 & \text{Type ID} = 4 & \text{InfoCode} = 0x0f \\
\text{Word 1} & \text{BASE_ADDRESS} & & & & \\
\text{Word 2} & \text{AND term} & & & & \\
\text{Word 3} & \text{OR term} & & & & \\
\end{array}
\]

**Request**

\[
\begin{array}{c|c|c|c|c|c}
31 & 24 & 23 & 16 & 15 & 8 & 7 & 0 \\
\hline
\text{Word 0} & \text{Version} = 2 & \text{Transaction ID} & \text{Words} = 1 & \text{Type ID} = 4 & \text{InfoCode} = 0x0f \\
\text{Word 1} & \text{Content of BASE_ADDRESS as read before the modify/write} & & & & \\
\end{array}
\]

**Response**

\[
\begin{array}{c|c|c|c|c|c}
31 & 24 & 23 & 16 & 15 & 8 & 7 & 0 \\
\hline
\text{Word 0} & \text{Version} = 2 & \text{Transaction ID} & \text{Words} = 1 & \text{Type ID} = 4 & \text{InfoCode} = 0x0f \\
\text{Word 1} & \text{Content of BASE_ADDRESS as read before the summation} & & & & \\
\end{array}
\]

Figure 5.9: The IPbus Read-Modify-Write bits operation which performs in one transaction
\(X <= (X & A) | B\), where \(X\) is the value of the register and \(A, B\) the new values applied on \(X\) [111].

\[
\begin{array}{c|c|c|c|c|c}
31 & 24 & 23 & 16 & 15 & 8 & 7 & 0 \\
\hline
\text{Word 0} & \text{Version} = 2 & \text{Transaction ID} & \text{Words} = 1 & \text{Type ID} = 5 & \text{InfoCode} = 0x0f \\
\text{Word 1} & \text{BASE_ADDRESS} & & & & \\
\text{Word 2} & \text{ADDEND} & & & & \\
\end{array}
\]

**Request**

\[
\begin{array}{c|c|c|c|c|c}
31 & 24 & 23 & 16 & 15 & 8 & 7 & 0 \\
\hline
\text{Word 0} & \text{Version} = 2 & \text{Transaction ID} & \text{Words} = 1 & \text{Type ID} = 5 & \text{InfoCode} = 0x0f \\
\text{Word 1} & \text{Content of BASE_ADDRESS as read before the summation} & & & & \\
\end{array}
\]

**Response**

Figure 5.10: The IPbus Read-Modify-Write sums operation which performs in one transaction
\(X <= (X + A)\), where \(X\) is the value of the register and \(A\) the new value added on \(X\) [111].

5.4 IPbus topologies

Having the full IPbus chain split to all the three different components (firmware, ControlHub,
\(\mu\)HAL) allows the implementation of IPbus in various topologies, from small point to point
configurations where the communication is between a single PC and an FPGA based hardware
card, up to large scale topologies with multiple boards and control applications. A set of ex-
ample topologies can be seen in fig. 5.11.

The HTT system consists of 48 ATCA crates fully populated with 14 boards. For this reason the
large-scale system shown in fig. 5.11 shows a setup which is as close as possible to the imple-
mentation for board monitoring in an HTT like topology. In this case the ControlHub will be
implemented in separate dedicated PCs that will handle all the traffic generated by the mon-
itoring applications. In addition, the ControlHub will ensure 100% reliability by calculating
the appropriate IPbus headers accurately and concurrently. The HTT topology foresees also a
number of switches that will allow the correct handling of the IP addresses of each board and
the reduction of needed cabling.
5.5 Performance studies

IPbus is foreseen to be used as a monitoring system for ATCA cards. Therefore, measuring its throughput and latency is a very important aspect for any system that is planning to use it. What is measured as latency in the studies below is the complete round-trip of an IPbus transaction as measured by the monitoring client application. Throughput is the amount of user data transferred/received as a function of unit time [108].

The system can be implemented in various topologies each affecting latency, since multiple components intervene between the hardware component and the monitoring application. For this reason, in all the cases where the latency was measured, it’s extremely important to understand all the intermediate components of the set-up, as discussed in sec. 5.4.

In the first scenario the system latency and throughput measurement was performed for a single block transfer, when a single µHAL client is used, the ControlHub runs on a different PC and the access is on a single hardware board (one IP address).

For a single word transaction a latency of 250 $\mu$s was measured which increases as a function of the number of transferred words in fig. 5.12. The behaviour observed is consistent with the expecting since increasing the number of words means that the maximum size of the IPbus packet is reached and therefore the breakdown into multiple transaction needs to be performed. Hence, the resulting throughput for 1 MB of data is measured to be larger than 0.5Gbps, due to the number of used Ethernet packets.

The latency measurement was repeated including multiple µHAL clients, simultaneously addressing the same register via the same ControlHub. The main reason for this test was to ensure the reliability of the protocol within the ControlHub can be achieved. The summary for 1,2 and
5.5 Performance studies

Figure 5.12: IPbus latency and average throughput on single point-to-point transactions as a function of the number of words written/read [108]. Added in the plot can be seen the value measured in Sussex with the tests described in the next chapter, with a fixed number of words send over time (hence only one point in the graph).

4 clients can be seen in fig. 5.13, which is consistent with the expected behaviour since by increasing the number of handled requests, more CPU interrupts are issued by the processor on which the ControlHub is running for different applications.

The last set-up in which the performance of IPbus was measured, is a large scale system with multiple hardware components and a 10Gbps switch included all in an ATCA shelf. Multiple block read/writes were performed yielding the result shown in fig. 5.14.

Figure 5.13: IPbus core latency with the total system polling frequency for n clients each simultaneously accessing the same register on different boards [108].

A measurement of the IPbus latency was also performed by the HTT project described in the chapter below (ch. 6), to ensure that the quoted values from the developers fit the required
5.6 Application in the HTT

With the latency and throughput value measured above [108], an initial back-of-the-envelope calculation was performed to assure that the IPbus protocol fits the budget of the HTT system. As already mentioned, an independent measurement was made in addition to the one shown by the IPbus developers to ensure that the numbers are accurate.

The HTT system will fully populate an ATCA shelf, hence the value with the 12 ATCA card configuration can be used as the expected latency (0.8Gbps throughput 5.5). The HTT system will contain 48 ATCA shelves with 12 AMTPs in each crate, therefore it will need 0.8Gbps per HTT Unit. Each AMTP will contain a PRM with 12 AM chips holding 384k patterns each.

\[
\text{conf} = \frac{HTTUnitsize}{0.8Gb} \times 1s = \frac{12 \times 12 \times 2 \times 12 \times 384 \times 18}{0.8Gb} \times 1s \approx 78.6s \quad (5.1)
\]

By performing the calculation it can be seen that each HTT Unit can update its patterns every 78.6 seconds. Consequently, the full system can have, in the ideal scenario, all the patterns changed within 1h. The shown calculation ensures that the IPbus protocol can be used both as a monitoring and configuration protocol for the HTT system. Normally during runs of the accelerator, more than 1h is available and hence in the worst case that the whole system needs to be re-configured the IPbus protocol provides sufficient bandwidth.

Another independent measurement of the latency and throughput values will be performed below by the HTT collaboration and compared to the designers values, as a validation cross-check before the final decision.
6 Testing legacy hardware

The University of Sussex is one of the leading contributors to the ATLAS HTT system. The main hardware contributions from Sussex are presented in the current chapter. Initially an overview of the individual goals is provided along with the creation of the Sussex ATLAS lab that fulfills those goals. Subsequently, a description of the used hardware is given along with the required firmware development. Finally, the test results using the hardware and the facility present at Sussex are discussed, in alignment with the agreed deliverable within the HTT project.

6.1 Sussex HTT contribution

The University of Sussex, as a member of the HTT projects, is responsible for the Quality & Assurance (QA) of the TP boards produced in the UK. For this reason the deployment of a dedicated lab facility within the University’s premises was important.

The facility at Sussex was used for hardware developed in past high energy physics projects (legacy) in order to establish a solid foundation in custom board evaluation with hardware similar to that expected to be produced for HTT. The process is aimed at gaining experience with the operation of the hardware and diagnostic tools in view of the mass production of HTT boards. Additionally, exploration of key HTT elements can be performed prior to the arrival of the final hardware:

- IPbus protocol latency measurement
- ATCA back-plane characterization
- Operation of Associative Memory chips

A detailed description of the tests listed above will be provided in the following sections of this chapter.
6.2 Legacy hardware

To achieve the various goals of the Sussex Lab, the procurement of existing hardware was necessary to populate the test-bed facility. The main aim was to prepare a legacy system identify any potential issues early on.

Since HTT is a scalable system organized in ATCA crates the purchase of an ATCA shelf was necessary, along with motherboards that would resemble the TP boards. The closest existing components are the Pulsar-IIb boards that were used in the FTK system. In addition, to bring the legacy setup much closer to the final HTT system, the deployment of a mezzanine card hosting AM chips was needed. All the components procured for the test system allow for HTT like operations to be performed, but not at their full capability.

In the following sub-sections a detailed explanation of each of the needed components in the final system is given. Some of the items were introduced in the previous chapter, however since they were used under the premise of the legacy setup, they are discussed in more detail in this part of the thesis.

6.2.1 ATCA shelf

The ATCA standard was introduced by a consortium of telecommunication companies, to allow reliable high availability operations. The high reliability aspect along with the specification document of ATCA \[100\], with clear constraints on all components involved, makes the ATCA standard an appealing option for HEP applications.

In the specification document a clear design of the controller boards, power supplies, communication buses and hot swap operations are clearly defined. Additionally, the interconnection of all slots in both available flavors of 14U and 8U crates, with increased dimensions in comparison to VME based crates, allows front boards to have a better airflow and taller heatsinks. The lower temperatures on the front-boards enables the longer lifetime of the designed hardware.

The board communication through the back-plane can be classified into two different types. The first type concerns the exchange of data between the ATCA cards using any custom serial interface (Fabric). The second type of connection is called the base channel and is used to allow the shelf monitoring information via the Intelligent Platform Management Interface (IPMI) protocol \[117\]. Each Fabric connection is called a lane and consists of 4x serial lines, with receiving and transmitting capabilities. The speed grade of each lane depends on the characteristics of the back-plane. At Sussex two available technologies are present \[118\].

1. 40Gbps per lane, full mesh
2. 100Gbps per lane Air-/Plane (model name) full mesh

The final important components of an ATCA shelf is the Hardware Platform Manager. All component of the ATCA shelf are designed to enable instrumentation and telemetry. Monitoring packages pass through the base channel (IPMB-0/1) of the back-plane via IPMI and are
6.2 Legacy hardware

Figure 6.1: ATCA Shelf in Sussex with 10Gbps switch included

Figure 6.2: Full mesh backplane connections in the Sussex ATCA back-plane [119]

handled by a central single board computer called Shelf Manager (ShMM). The ShMM can be either a separate component of the ATCA shelf, or it can occupy a physical slot in the crate. In the set-up at Sussex both options are available, the 10Gbps switch that can be seen in the centre of the shelf in fig. 6.1 is mostly used for monitoring purposes and the dedicated ShMM mount on the back for power up and cooling.

The IPMI protocol and the ShMM require all crate components to include a ROM that can be accessed via IPMI and provide identification information to the ShMM. Since the IPMI bus is designed to communicate through the I2C protocol, the components identification information is received from the ROM using standard I2C operations [120]. Any other information passed to the ShMM via IPMI is handled by each board through a dedicated device called the
6.2 Legacy hardware

6.2.2 Pulsar-IIb

The legacy motherboards used in Sussex are the Pulsar-IIb boards (fig. 6.4) developed by Fermilab and used in various CERN projects, such as the FTK system. It’s effectively a full mesh enabled FPGA based carrier card, which allows hosting of up to 4x mezzanine cards connected through a Field Mezzanine Connector (FMC). By providing a full mesh capability to the card, it allows FPGA designers to share data via the Fabric at very high speeds across the shelf boards.

The Pulsar-IIb is essentially a multi-I/O card containing a large FPGA to perform any required processing on the incoming/outgoing data (7th series Virtex Xilinx FPGA (XC7VX960T) [122]).
The chosen FPGA is the larger option in the given family providing access to both back-plane serial transceivers as well as to the mezzanine card connector, due to its high pin count. Additionally, a large number of logic cells in the FPGA Fabric is provided to the firmware developers for implementing various algorithms. Finally, the maximum communication rate via any serial link (back-plane, mezzanine) is 12.5Gbps due to the specific type of Xilinx transceivers included in the FPGA package [123].

To handle the power-up and all monitoring information the Pulsar cards include a dedicated slot for an IPMC board. This is a small form factor connector allowing the plugin on a dedicated card which gains access to the IPMI interface in the base channel. The IPMC can be seen in fig. 6.3 is interacting with the ShMM via a MicroController which implements all the necessary protocols for the Pulsar-IIb sensors.

Finally, the Pulsar cards provide further I/O capabilities through the top backplane connector (a.k.a. Zone-3) that is connected directly to the RTM. The RTM (managed by the IPMC plug on the Pulsar card for powering and monitoring purposes) provides a set of Quad Form-Factor Pluggable Transceiver (QSFP) modules for further serial communication with the external world.

A schematic overview of all the components and connections on the Pulsar-IIb card can be seen in fig. 6.5. For the Sussex setup the Pulsar cards are used to mimic the HTT TP boards.

![Figure 6.5: Pulsar-IIb board block diagram. All mezzanines can be seen in light blue, and the mounted components in solid blue. Communication lines are noted with black and red [119].](image)

### 6.2.3 Xilinx evaluation cards

To avoid the complication of the ATCA system and the use of the Pulsar board during the early testing phases, a commercial alternative was used. The use of the multi-purpose evaluation cards produced by Xilinx allows use of the same FPGA family as the Pulsar-IIb on a simpler card that can be accessed directly via the Joint Test Action Group (JTAG) protocol [124] from a e.g. PC.
6.2 Legacy hardware

However, since the FPGA on the Pulsar-IIb card is the largest in the family, an evaluation card hosting the same exact model did not exist at the time of purchase. Therefore a simpler FPGA is used which contains the same type of transceiver and allows use of the same Intellectual Property (IP) cores as the one mounted on the Pulsar card. The card used can be seen in fig. 6.6 (VC707) which is hosting a Virtex-7 FPGA [125]. In addition, the evaluation card contains the same type of mezzanine connector as the Pulsar card which allows the plugging of the same type of mezzanine card.

6.2.4 Pattern Recognition Mezzanine 06

The PRM06 is a custom hardware board, hosting a medium-sized FPGA and an older generation of the AM ASIC (6.2.4), designed by the CMS collaboration during their R&D program to implement tracking at the L1 stage [126]. It's a 14.9x14.9cm² Printed Circuit Board (PCB) designed to be mounted on two mezzanine slots of the Pulsar-IIb card.

The connection with the carrier card is accomplished via the two FMCs [127] at the lower part of the PCB. They provide a maximum serial rate of 12.5Gbps in a total of 8 dedicated serial pin pairs and 160 LVCMOS [128] pins for parallel data exchange. In addition via the FMC connector 3.3 and 12.0 volt power lines are passed from the carrier card allowing the PRM power regulator to power up successfully.

The PRM06 mounts a Xilinx Kintex Ultrascale [129] FPGA (KCU060) which receives from the on board power regulators 1.0, 1.2, 1.8 and 2.5 volt to power up. The KCU060 FPGA receives the incoming data from the carrier card, and performs any user defined logic and interface with the mounted AM chips. Communication with the AM chips is established over serial transceivers. With the use of scalable fan-out the FPGA requires only 16x serial links to interface with all the chips, since each group of 6x AM chips shares the same input lines [126].

The PRM06 can be used both with the Pulsar-IIb card (directly) or with some Xilinx evaluation cards. In the latter case a passive electromagnetic adaptor board is needed in order to provide the appropriate supply to the PRM06 and ensure the spacing of the connectors is appropriate.
The front layout of the PRM06 mezzanine card can be seen in fig. 6.7.

![Figure 6.7: Pattern Recognition Mezzanine 06, with 12xAM06, the KCU060 and the 2xFMC connectors (Front side)](image)

### Associative Memory chips 06

Similarly to HTT, the L1Track CMS R&D activity [126] performs pattern matching with the use of the Associative Memory chip technology. The PRM06 mounts 12xAM06 grouped in four different groups (2x2x4x4).

The AM06 in contrast to the one planned to be used in HTT (AM09) can hold up to 126k patterns per chip, yielding a total capacity of 1.5M patterns per mezzanine card. Each of those patterns consists of eight independent 16-bit words corresponding to eight detector layers.

Incoming data are received over 8x serial transceivers from the FPGA at the rate of 2Gbps and streamed out at the same rate over a single serial line per chip. Each AM chip is clocked at 100MHz. The total latency of the ASIC is calculated by the following function

\[
t_{\text{latency}} = N_{\text{matches}} \times t_{\text{clk}} + t_{\text{SERDES}} + t_{\text{datadist}}
\]  

(6.1)

where, the \(N_{\text{matches}}\) describes the number of matched patterns per event, \(t_{\text{clk}}\) the period of the clock in nanoseconds, \(t_{\text{SERDES}}\) the latency introduced by the serial transceivers for the in/outcoming data and \(t_{\text{datadist}}\) the internal latency for distributing the data within the ASIC after decoded from the serial line, which is estimated to be between 10 and 20 clock cycles [130].

The AM chip can change its behaviour based on a specific set of words sent via a dedicated control line. Those words are called OPCODEs and can have various effects on the behaviour of the ASIC. The list of allowed OPCODE values can be seen in table 6.1.

### 6.3 Hardware setup

With the hardware described in the previous section (6.2) two main set-ups are planned. Both set-ups contribute to the final goal of jumpstarting the test-bed facility which will evolve through
6.3 Hardware setup

<table>
<thead>
<tr>
<th>OPCODE value</th>
<th>Operation</th>
</tr>
</thead>
<tbody>
<tr>
<td>4'b(0001)</td>
<td>Matching threshold 8/8</td>
</tr>
<tr>
<td>4'b(0002)</td>
<td>Matching threshold 7/8</td>
</tr>
<tr>
<td>4'b(0003)</td>
<td>Matching threshold 6/8</td>
</tr>
<tr>
<td>4'b(0005)</td>
<td>Event separator</td>
</tr>
<tr>
<td>4'b(0007)</td>
<td>Decrease threshold by one level</td>
</tr>
<tr>
<td>4'b(000a)</td>
<td>Set the threshold to 0, all patterns match</td>
</tr>
<tr>
<td>4'b(000b)</td>
<td>Set the threshold over 8, no patterns match</td>
</tr>
<tr>
<td>4'b(000e)</td>
<td>Toggle the request for layer 0 to match</td>
</tr>
<tr>
<td>4'b(0000)</td>
<td>NOP</td>
</tr>
</tbody>
</table>

Table 6.1: Allowed OPCODE value with the different actions that are performed.

the test of the early HTT TP board prototypes.
Ultimately, these set-ups will converge into one single ATCA based set-up which will make use of the Pulsar-IIb boards mounted with the PRM06 mezzanines.

6.3.1 Workbench

The first set-up is what is called the workbench set-up. The main goal of this setup is to have an early simple and reliable interface towards the PRM06 mezzanine card, allowing the development of its firmware. Therefore, it uses the Xilinx evaluation card as a motherboard to allow the mezzanine communication with the external world. The workbench setup also provides access to the PRM and evaluation card FPGAs through JTAG to USB connections.

The advantage of cooling provided by the crate is not present in the workbench setup. For this reason a mechanical support was designed with the use of Autocad [131]. This support includes a cooling fan installed to ensure that the AM chips do not overheat. The fan is mounted on the side of the PRM06, the test set-up mounted on the support can be seen in fig. 6.7.

For powering the mezzanine card the adapter card was used which is connected to a 3.3V and a 12.0V power supply with an ammeter to check the current drain of the mezzanine during operation. For the evaluation card the standard 12.0V power supply [125] from Xilinx was used.
Additionally, the mechanical support isolated the mezzanine card from the 3.3V pins on the FMC connector of the evaluation card. A summary of the setup along with all the connectors can be seen in fig. 6.8 below.

6.3.2 Crate

The second more versatile set-up is based on the ATCA shelf. With the use of Pulsar-IIb cards high speed I/O capabilities of many components can be tested in the shelf. In the Sussex shelf, sending data into the system is done via the 10Gbps ATCA F-125 switch [132]. The switch front-panel contains a set of standard Ethernet connectors (RJ-45), which can be interfaced with the back-plane lines.

The ATCA shelf in Sussex contains in practice two Shelf Managers. The first one is external to the shelf (mounted on the back), which is automatically negotiating with the fan tray, any in-
6.3 Hardware setup

Figure 6.8: Workbench setup at Sussex ATLAS lab. On the left the mezzanine card with the JTAG connector, in the middle the adapter card allowing power-up and on the right the VC707 Xilinx evaluation card.

inserted boards, and power supplies. This is a standard pigeon point (vendor) ShMM-500R which can be accessed either via a serial connection or ssh. The second Shelf Manager is the switch itself, which is mostly used to enable the connections in the back-plane for various tests between the inserted ATCA cards.

A complication with the ATCA set-up which is not present in the Workbench is accessing the FPGA on the Pulsar-IIb boards. The Pulsar cards contain a JTAG connector where the same JTAG to USB interface can be used. However, populating the shelf as shown in fig. 6.9 makes the access physically very challenging. For this reason, all the Pulsar-IIb IPMC mezzanines implement a version of the Xilinx Virtual Cable (XVC) [133]. The XVC converts the JTAG packages from the FPGA into a 10Mbps Ethernet connection via the IPMB lines on the back-plane. Since the ShMM on the switch has access to those lines it can map the Ethernet frames to the front-panel of the switch allowing a standard PC to gain access into the IP and MAC addresses of each Pulsar-IIb board. In this way implementing the firmware to any Pulsar card can be achieved via the Xilinx Vivado Software [134].

With the ATCA setup a large amount of tests can be performed for all the ATCA components. In

Figure 6.9: ATCA Shelf at Sussex, populated with Pulsar-IIb boards and the 10Gbps switch.
6.4 Firmware development for PRM06

All custom hardware boards for the lab set-up use an FPGA as the main processing unit. Therefore, to enable them and gain access to all the components it was necessary to develop firmware at Register-Transfer Level (RTL) with the use of VHDL and/or Verilog [135].

The IPbus protocol was used to exchange data in both set-ups. Effectively, standard Ethernet packets are sent from a PC and are decoded in the FPGA into 32-bit words that can be then stored in memory components.

The whole system is based on the communication of two FPGA (carrier card-mezzanine) with functional blocks distributed across two FPGAs as follows: The Ethernet packets would be minimally decoded in the carrier card and passed through a parallel interface via the FMC connector to the mezzanine. In the mezzanine FPGA all remaining decoding is performed, in addition to any desirable user logic to allow operation of the AM chip ASICs.

This organization greatly simplifies the firmware porting to the Pulsar-IIb + PRM06 set-up, since the FPGA families are equivalent, and the most complex part of the firmware is implemented directly in the mezzanine. A high level overview of how the firmware is organized can be seen in fig. 6.10.

Since all the FPGAs are from the same vendor (Xilinx), the development of the firmware employs Vivado 2017.3 [134] in all cases. The core logic was designed to be as vendor independent as possible, however the decoding of the Ethernet packets uses some Xilinx dependent pre-defined firmware blocks (IP-cores). The simulation of all the critical components was per-

![Figure 6.10: Legacy hardware high level block diagram. With red indicating the carrier card, and blue the mezzanine. The double arrows show serial communication and the black lines parallel interfaces. Dark blue indicates the different implemented firmware blocks.](image-url)
formed with the use of the integrated to Vivado Xilinx Simulator, with all test-benches written in VHDL.

6.4.1 Carrier card block

With the organization of the firmware as shown in fig. 6.10 the carrier card acts as a bridge passing the Ethernet frame to the mezzanine FPGA, which contains the IPbus firmware. The Pulsar-IIb FPGA receives the incoming packets via the back-plane and implements the Xilinx 1G Ethernet PCS/PMA or Serial Gigabit Medium Independent Interface (SGMII) IP core [136], with the firmware implementation of the 1000Base-X PHY layer [112]. The output of the IP-core is a standard 125MHz 8-bit data-path Dual-Data Rate (DDR) interface. The complete data-path in the ATCA setup can be seen in fig. 6.11. Despite aiming at a common design, the data-path of the Evaluation Board and the Pulsar-IIb in the ATCA differ slightly.

In the ATCA crate sending data happens via the 10Gbps switch which acts as an interface towards the Ethernet network. In early tests the data was routed through the switch front panel, through a 1 Gbps line.

For the Workbench set-up (shown in fig. 6.12), instead of the data going through an extra hard-

![Figure 6.11: The PC to PRM data-path in the ATCA setup. Black double arrows the serial connections, black lines the parallel DDR, dark blue the firmware blocks in the IP core and light blue the different hardware components.](image)

ware component (switch) the Evaluation Board has a direct connection to an RJ-45 port. The first device retrieving the data from the connectors magnetics is a Marvel Alaska PHY (88E1111) layer chip [137], which converts the incoming data to SGMII lines connected to the Virtex-7 FPGA. There a standard Xilinx IP-core decodes the SGMII protocol and converts it to a 125MHz DDR interface.

The parallel DDR interface sent to the PRM06 mezzanine, via the FMC, is the Gigabit Medium Independent Interface (GMII) defined in the IEEE 802.3 standard [112]. In total GMII requires 16-bits to exchange data and 4-bits as control signals, in addition to one clock line. To pass those signals via the FMC the LVDS lines clocked at 125 MHz were used.

The DDR interface through the FMC connector is source-synchronous, meaning that the data
Figure 6.12: The PC to PRM data-path in the Workbench setup. Black double arrows indicate the serial connections, black lines the parallel DDR, dark blue the firmware blocks in the IP core and the light blue the different hardware components.

are transmitted with a clock from the carrier card to the mezzanine and sampled with the same clock. This required the use of specific timing constraints \[138\] to maintain aligned package transmission and reception with the source clock.

The firmware block is extremely light-weight for both carrier cards, using only few FPGA resources. A post-implementation resources utilization for the Xilinx Evaluation card case is shown in fig. 6.13 generated with Vivado. For the ATCA set-up the percentage resources are even smaller, as the FPGA is much larger.

Figure 6.13: Carrier card firmware block resources utilization from Vivado 2017.3 for the Xilinx VC707 Evaluation Card

### 6.4.2 Mezzanine card block

The developed firmware for the PRM06 is more complicated than the carrier card firmware. The board layout described above foresees the interaction of the PRM06 FPGA with both the carrier card FPGA and the AM chips. To allow for full functionality of the AM chips various different interfaces need to be set in place. A block diagram overview of the various interfaces needed can be seen in fig. 6.14. The red and black lines interface directly with the ASICs.
6.4 Firmware development for PRM06

Figure 6.14: Top level PRM06 firmware block diagram with all the individual blocks interfacing with the AM chips. The orange colour indicates the AXI4-Stream [139] interface, red is the JTAG, dual arrows indicate the high speed serial links and finally the green indicate the I2C lines towards the temperature sensors.

I2C lines are connected to temperature sensors mounted on top of the AM chips (for the temperature tests described below). The post implementation utilization of the firmware block can be seen in fig. 6.15. Apart from the number of high speed serial transceivers, all the remaining resources are below or at the limit of the traditionally expected 30% occupancy. A detailed description of each block is provided in the following parts of this section below.

MAC layer

The first block receiving the data from the carrier card in the form of the GMII interface is the MAC layer. The MAC layer corresponds to the standard MAC sub-layer from the Ethernet protocol link layer [112] implemented with the use of the Tri-Mode Ethernet MAC [140] IP core.
from Xilinx. The main functionality of the MAC core is to strip the the extra information from
the Ethernet packets and convert the GMII interface into the AXI4-stream [139] interface mostly
used by Xilinx to interface various firmware blocks.

**IPbus core**

The next block after the MAC IP core is the so called IPbus core. In practice this block is per-
forms the operations described in ch. 5. The incoming data are stripped from all the Ethernet
overhead data (e.g. MAC address, IP address) and converted into the 32-bit address and 32-
bit data words for the IPbus slaves. The IPbus slave list consists of single registers which are
used for configuring the other blocks of the firmware, custom made dual clock block memories
to hold data that need to be converted into a specific protocol (e.g. JTAG) in a different clock
domain and Xilinx FiFo memories to allow an automatic data flow without user intervention.
The last two IPbus slaves exceed the standard set by slaves provided by the IPbus developers
and therefore after being implemented were shared with the IPbus developers and the HTT
community.

**I2C master**

The next set of blocks implement functionalities required to operate the AM chips. The first
block from the left of fig. 6.14 is the I2C master, controlled through IPbus. The block consists
of two sub-blocks, where the first one is responsible for retrieving the 32-bit words from the
FiFo IPbus slave and decode the bits to define the I2C operation to be performed. Basically, it’s
an implementation of a data driven Finite State Machine (FSM) with the following states listed
below and seen in fig. 6.16.

- **Reset state**: Brings the whole I2C module to reset, which can be triggered either by cer-
tain bits in the data word from the FiFo or a global asynchronous reset to the whole FPGA
firmware.

- **Idle state**: The default state where the trigger is expected in order to begin the opera-
tion. The idle state performs a minimal decoding of the received 32-bit word to identify
whether the operation is a write or read and how many 8-bit words will be transmitted
via the dedicated I2C lines. This is the state where the system returns when a complete
operation was performed.

- **Write state**: The write state is triggered via the idle state and breaks the received data
word corresponding to I2C commands into 8-bit words to be passed into the second sub-
block. In addition, it issues the command to start the generation of the I2C clock.

- **Read state**: This state is accessed via the idle state when a read I2C operation needs to
be performed on a given slave. Additionally, it issues the command to free the ports on
the FPGA in order to be driven by the slave expecting to send the data. The received
Figure 6.16: I2C FSM diagram, with the signals triggering a state change

Information is organized in 8-bit words and finally appended into a 32-bit word sent to the dedicated IPbus slave storing the output.

The FSM is set to be synchronous with the IPbus core at a clock rate of 31.25MHz. Since the I2C operation is much slower than the IPbus core frequency, the clock domain crossing of the words can be performed with the use of asynchronous registers accessible from both clock domains.

The second sub-block within the I2C master implements the actual bi-directional converter in the 8-bit words required by the FSM to the serial commands expected by the I2C protocol. When the converter receives the start clock command from the write/read states it starts the generation of a 100kHz clock driving the SCL signal. In both situations of write and read the converter needs to transmit a 7-bit address word appended with a 1-bit parity which informs the slave of a read or write operation. If a write operation is performed, the next 8-bit word received from the FSM is transmitted over the I2C SDA line. In the read operation after the address was set, both SDA and SCL ports are set to high impedance to allow the slave to drive the lines. In both operations the converter is aware in advance of the length of the words expected to be transmitted or received, from the decoded data words in the FSM. The last operation of the converter foresees the situation of an observed error which issues a reply to the FSM forcing it to move into idle and prepare the default output word for an error 32'h(deadbeef) (which is an arbitrary value used by the designer of the firmware) to notify the user.
6.4 Firmware development for PRM06

**JTAG master**

The next block implemented in firmware is the JTAG master. The JTAG master’s main purpose is to interface with the AM chips and bring them into a certain operational state. Additionally, it is used to load the patterns that each chip will contain along with the number of hits required to fire a specific pattern.

The logic of this block is largely the same as in the I2C case. There is a sub-block called a controller that implements an FSM to identify the operation that will be performed, and prepare the data for the actual JTAG protocol signals. The operation and configuration are handled by a single IPbus register, whereas the data sent and received are stored in the custom dual port dual clock block RAM slave. The states for the controller FSM (fig. 6.17) are the following:

- **Reset state (R1):** This state sets all the values to their default 0x0 value stopping any operation in the system.
- **Prepare state (S1):** This state is automatically accessed after the reset state. It sets internal signals to the values expected for a given operation, and moves directly to the next state. This state is implemented mostly for timing purposes, since the direct transition from the reset to the operation state might have otherwise caused erroneous information to be transmitted.
- **Wait for trigger (S2):** This is the state which corresponds to the idle state for the I2C FSM. The system waits for the user to set the configuration register trigger bit to 1'b(1) in order for the transactions to start. When the trigger arrives the FSM transitions to the next state.
- **Process commands (S3):** When the system arrives in this state the number of written words in the block RAM IPbus slave are checked. Based on this number the amount of

![JTAG FSM controller states](image-url)

Figure 6.17: JTAG FSM controller states.
operations is automatically calculated to inform the block how long the JTAG transaction will be active. If no words are stored in the memory the system returns to the prepare state since no operation can be performed.

- **Get commands (S4)**: In this state the FSM receives the first word from the IPbus memory and proceeds to the next state for decoding. The main reason for this state is again mostly due to timing since immediate memory access and decoding of a word can cause errors.

- **Send command (S5)**: When accessing for the first time this state the first command is prepared to be passed into the sub-block that handles the JTAG protocol signals. When the command was fully prepared and passed an acknowledgement signal is raised transitioning the FSM into the next state.

- **Receive result (S7)**: The outcome of the JTAG operation is received and a decision is made whether more data exist in the memory that need to be sent or if the system should lower the trigger and return to the prepare state. The output received is transferred to the IPbus memory slave, ready to be retrieved by the user when needed.

- **Increase address (S6)**: This state is reached from Receive result if there are IPbus slave commands to be processed. Therefore the system increments the memory address and returns to the Process commands state to further process the new data word.

The JTAG protocol is more complex than I2C and hence a more careful decoding of the information is needed. Knowing just the slave address is insufficient for the user to perform an operation and therefore an a-priori agreement on the data format is needed between user and firmware developer. In this implementation it was decided to offload the complexity to the software rather than firmware for the economy of time. Hence, a software routine was prepared that is fully aligned with the operation of the JTAG interface preparing the commands handled by the controller with the required bit aligned organization (bit-accurate).

The second sub-block in the JTAG master is the block responsible for actually driving the JTAG signals. Every JTAG slave implements an FSM foreseen by the specifications [124]. Transitions in this FSM are performed in combination of the TCK (clock) and TMS signals. For this reason the protocol handling block receives the commands from the controller and depending on the values enables the TCK clock signal and drives the values of TMS to move the slave FSM to the desired state where the register values can be updated or read with the TDI and TDO lines. If a read operation is performed the output is bundled in a 32-bit word and the acknowledgement signal is raised for the controller to store the value in the IPbus memory.

### Data preparation logic

The data preparation block acts as a data interface for the AM chips, formatting, sending and retrieving data with the AM chips. It is the most complicated block developed for the test set-up, since it is implemented across two comparably fast clock domains: the IPbus one (with
6.4 Firmware development for PRM06

the 31.25MHz clock allowing I/O at the rate of 1Gbps) and the serial transceiver I/F domain (which operates at a line rate of 2Gbps exchanging data on the parallel I/F side at 62.5MHz for 32-bit words). The co-existence of both clock domains made possible by a set of IPbus slaves and the appropriate clock domain crossing synchronizers. For the data words the standard dual clock dual memory IPbus slave [108] was used, whereas for the configuration bits dual flip/flops were added to ensure that the system will not end in a metastable state due to the clock domain crossing.

This logic block is implemented again as a FSM. Each AM chip has 8 serial inputs called hit lines and 1 serial output called road line. The same 8 data lines are fan-out to the input of all the 6 AM chips in the same group whereas each of AM chip has a unique output line connected to the KCU060 FPGA. Hence to cover both group of 6 ASICs, 16 IPbus memories are used to buffer the data that need to be sent from the serial I/F and 12 memories for the output lines from each chip. In addition a configuration register is used to drive the block through the various states and finally two registers for transmitting a special type of data word called OPCODE, needed by the AM chips for special configuration modes (e.g. Stream data out with 7/8 successful hits).

The FSM states (fig. 6.18) that describe all the potential functionalities needed for the AM are:

![Figure 6.18: FSM for transmitting data towards the AM chips.](image)

- **Reset state (R1):** In this state the FSM issues a reset towards the transceiver I/F along with setting all the operational parameters to default values to avoid any error. The state can be reached via the value of the configuration register or via the global asynchronous reset to the whole firmware.

- **Decode state (S1):** Transfers the value of the configuration register to the serial transceiver clock domain and based on the register value decides which data sending operation mode and which serial lines will be employed. The various operations modes are
discussed below and arise from the fact that data cannot be sent continuously to the AM chip due to the mismatch between the IPbus and AM serial lines data line rate. The system will leave this state once a dedicated trigger line is raised from the IPbus control interface, going either to *set switching values* state or to *increase data address* state or directly to *_OPCODE send* state, depending on the decoded configuration information.

- **Increase data address state (S2):** Once the transmission options was determined in addition to which set of chips will the data sent, the controller reaches this state from the *decode* state. Here the address of the chosen memories will be incremented to the next data location to be sent to the AM chips, based on the current configuration. Once completed, the controller will either transition to the *end* state or to the *_OPCODE send* state based on the operation.

- **_OPCODE send state (S3):** The controller will reach this state if the option in the configuration register requires the transmission of the value set in the OPCODE IPbus slave. A list of the available OPCODE values, consistent with the AM06 specifications, can be seen below tab. 6.1.

- **Set switching values state (S4):** In this state the FSM can arrive only from the *decode* state. One of the firmware block’s operation modes foresees the transmission to the AM chips of data interleaved with idle words. Achieving this operation mode requires the generation of a clock signal with variable duty cycle. The number of idle and data words sent is defined through the value of an IPbus register which in this state here are synchronized into the serial transceiver clock domain. The FSM will transition, after the values were set, to the *decode* state to start data transmission with the set ratio of data/idle words.

- **End state (S5):** This is the final state of the system and will be reached before completing its operation. A safety signal is raised to indicate the end of operation in this state effectively taking the FSM to the *reset* state, so that the user can define another operation mode.

Because of the line rate mismatch between IPbus and the ASIC transceivers, data cannot flow continuously in the system therefore buffering is required. The FSM implemented in firmware allows a versatile use of the input data, mimicking as close as possible the operation of the AM chips in the HTT system:

- **Send a single stream:** The content of the IPbus memories is streamed to the AM inputs continuously until one loop over the memory address counter is completed.

- **Send OPCODE word:** The value of the register holding the special OPCODE character is sent to the AM chip.
• **Combined data with OPCODE:** An OPCODE character is sent after the data memory content is streamed out.

• **Mix data with idle words:** Every m data words from the memory n serial transceiver IDLE words [141] will be sent, allowing to control the rate of data processed by the AM chip.

• **Loop over memory content:** Each of the operation modes described above can be used in a loop mode where the memory content is looped over and over until the user manually de-asserts the trigger bit in the IPbus configuration word.

In order for the AM chip to stream the matched roads based on incoming hits the use of special OPCODE characters is needed which translates to the end of event operation. The allowed OPCODE values can be seen in table 6.1 and can be sent via a separate control line to each group of 6 AM chips.

Once an OPCODE was issued the AM chips stream out from the output serial line the buffered patterns that were matched based on the instructions provided. The output lines are connected to dual clock dual port RAMs (Road RAMs) that store the road data stream received from the AM chips. Since no processing on the data itself is needed there are three simple operation modes for those memories. Before the AM chip transceivers are enabled the serial lines cannot be tight to 0 [141] and thereafter exchange electrical idle words. The road rams have a filter where those words can be ignored and never stored in the content. In addition when operating the transceivers to exchange protocol specific IDLE words the road memories can either store or exclude this information. Finally, if both filters are enabled only patterns are stored in the road memories which can be read then directly from IPbus. Since the depth of the RAMs is fixed, a flag in the configuration register is raised (informing the user that the memory content was overwritten) if the AM output words to be stored exceed the road RAM size.

**High speed serial I/F and In-System IBERT**

The last block of the system is a Xilinx IP configured with the requirements for the AM chip protocol. The block is implemented with the use of the GT Wizard for Ultrascale Architecture [142] IP. The AM protocol is a standard 2Gbps 8b/10b encoded serial protocol, with IDLE words 32'h(1c1cbcbc) or 32'h(bcbc1c1c). The protocol has no header or trailers to avoid decoding complexity at the level of the AM. A set of registers were implemented in the IPbus core block to allow the dynamic configuration of some serial transceiver specific parameters like the differential swing and so on.

In addition the integration of the Xilinx In-System IBERT core was provided with the serial interface to allow signal integrity tests. Since the AM chips implement only receivers the Xilinx IBERT infrastructure cannot be used to produce eye scans however internal counters are implemented in AM06 which count the number of errors in a specific pseudo-random input pattern (PRBS).
Simulation

The two key blocks for the PRM06 firmware were simulated in detail to ensure operation at RTL level before the firmware is loaded in the real hardware to minimize risk of damage. The first block simulated was the data preparation logic since it's the block with the more critical operations for the usage of the AM chips. The data preparation block allows as shown above multiple operational modes to send data. All modes were simulated individually, however for the economy of space only the most combined mode is shown in fig. 6.19. The signals in fig. 6.19 show the 66.5MHz clock required for the serial transceiver interface in addition to the state signal indicating when the controller is sending data and when idle. Additionally, the data lines connected to the receivers of all AM chip sending the word 32'h(00000001) for 32 times and then followed by a set of idle words. Finally, the OPCODE sent via dedicated serial lines (bus-0) to both AM groups can be seen after the transmission of 32 data words. The frame is repeated since the loop mode is selected and it will repeat until the trigger line from the IPbus slave is lowered. The equivalent frame decoded by the oscilloscope on the actual serial line can be seen in fig. 6.20, showing an example of the validation between simulation and hardware.

The next block that was simulated before implementation, is the I2C master. Although the master is a very slow component in comparison to the data processing logic block, simulating it's actual functionality was very important to show the accurate behaviour and ensure reliable temperature readouts. In addition the master was modified to allow an arbitrary amount of bytes to be sent or read over the I2C line rather than the default value of 8-bits. The simulation can be seen in fig. 6.21. As expected the master does not drive the lines to high or low but rather to high impedance since according to the I2C specification [120] the bus is allowed to be driven by either the master or the slaves. The readout word was chosen to be 16-bit since the output value of the temperature sensors is 16-bit as well.

6.5 Lab tests

The firmware described above was developed to allow a range of tests that could be performed on the existing hardware. The tests were chosen as prototypical of what would need to be performed on the HTT demonstrator and production boards. This exercise serves the purpose of better identifying the needs of the test facility, and practice for the tests themselves. Three type of tests were performed described in detail in the remainder of this section below.

The first tests concerned the performance of the IPbus protocol in its standard 1Gbps implementation. Since the IPbus was chosen to be the main communication protocol for monitoring the HTT system, verifying its performance prior to the decision was an important aspect. For this reason a similar setup to the HTT demonstrator had to be used where two FPGA were communicating with each other via a similar type of connector as the one planned to be used for the HTT. The legacy hardware with it's firmware that exists in Sussex proved to be the best ready to be used setup for this measurement.

The second important test for the HTT system is the measurement of the AM chip tempera-
Figure 6.19: Simulation frame of 2.6us for the data preparation logic. The combined operation of sending into both groups of AM chips a number of data words and a number of idle followed by a specific OPCODE in a loop can be seen.
6.5 Lab tests

Figure 6.20: Oscilloscope decoded frame on 8b/10b encoded serial line on PRM06.

Figure 6.21: 100kHz I2C sim frame with a single 8-bit write operation for register selection followed by a 16-bit (same as temp. readout) data word.

ure while in operation. HTT is planned to use the 9th version of the ASIC which will not be available for testing until the end of 2023; thereafter the use of the existing version of the chip was the best practice available in view of the future tests. For this reason the PRM06 board was used, with the addition of the dedicated temperature sensors coupled to the ASICs and read-out through the board's FPGA.

The last tests performed in Sussex with the existing hardware concerns the testing of the ATCA shelf. The HTT system will be organized as explained above in ATCA shelves called HTT-units. The communication of those boards will be performed via the 100Gbps back-plane. The back-plane characterization is a key aspect of the system behavior, hence the existing ATCA shelf [118] was used to derive the qualification procedure for the equivalent shelf planned for HTT, once available. For this reason the link quality of all the connections was validated with the usage of Pulsar-IIb cards and the Xilinx Integrated Bit Error Rate (IBERT) [123] firmware configured appropriately.

In the following sub-sections the results obtained from the different tests are going to be presented, along with the description of the exact setup employed.

6.5.1 IPbus latency and throughput measurement

The first set of measurements with the legacy hardware concerned the IPbus latency and throughput in a system with two FPGAs. The main effects affecting the performance are expected to be related to the UDP packet decoding and the Ethernet connection, with a second order effect to be the connector between the two FPGAs. This organization emulates the setup of the HTT system where a carrier card FPGA (e.g. TP) interfaces the external network to the mezzanine FPGA (e.g. PRM/TFM). For the latency tests the Xilinx evaluation card was used along with the
PRM mounted on the top.

On the firmware side two different firmware components were used on the different FPGA.

The first firmware block loaded in the FPGA of the evaluation card is the block described above dedicated to minimally decode the incoming Ethernet link into the GMII parallel DDR interface. The output packages are sent over the FMC Low Voltage Differential Signaling (LVDS) lines to the mezzanine FPGA, where the second firmware block is loaded. The second block is a standard implementation of the IPbus firmware block provided by the developers. In practice the firmware implements the standard UDP core which accesses the payload information where the IPbus package is held, followed by the IPbus logic to decode the package and perform the requested operation on a dedicated slave.

To perform the measurement of the throughput and latency a method to measure the required time was developed, in order to ignore any overhead added by the PC issuing the request. For this reason the use of an oscilloscope and signal coming from the FPGA were used. In practice the procedure was the following. The PC issues a write packet for a single register on the mezzanine FPGA. When the packet arrives in the evaluation card a flag is raised from the FPGA and sent to the oscilloscope via the SubMiniature version A (SMA) connector. The package at the same time passes to the mezzanine card where the IPbus core decodes the information and

Figure 6.22: Test-bed setup with hardware triggers for measuring the IPbus latency between two FPGA, without the software overhead.
updates the register value, replying to the software the success of operation. At the moment the register is updated by the IPbus write request, a flag is raised and sent via the second evaluation card SMA connector to the oscilloscope. The time difference between the two signals is used to measure the latency. In addition a software stopwatch is implemented in Python to measure the equivalent value folding in all the latency from the cable and CPU. Fig. 6.22 reports a picture of the set-up, highlighting the connectors used for the hardware measuring method.

The latency found on a single write operation with the hardware and software stopwatches is:

\[ \mu_{HW} = 1.64 \pm 0.04 \mu s \quad \text{and} \quad \mu_{SW} = 0.2 \pm 0.05 ms \] (6.2)

The measurement of the latency was compared with the values provided by the IPbus developers [108], which found to be consistent within uncertainty although the definition of latency in ref. [108] is not entirely the same (more components are included/excluded in the setup).

To measure the throughput of IPbus with the current setup an incremental amount of writes was performed on an IPbus slave memory taking the average time differences from the hardware and the software values over twenty operations. Initially, one hundred and twenty eight (128) words were written in a block RAM slave, increasing all the way to 128000 words of 32-bits per request. The average values for the transfer rate found over twenty repeats, can be seen in fig. 6.23. The error bars are mostly dominated by statistical fluctuations rather than systematic uncertainties on the measurement. The overall summary for the throughput was found to be \( O(0.5) \text{Gb/s} \) or better for any transfer that exceeds 40kB.\(^1\)

Comparing the measured numbers with the defaults provided by the IPbus ref. [108] it can be seen that the IPbus protocol is within the margins of the HTT system for monitoring purposes. An additional usage of IPbus could be for configuring the system. For that the worst case scenario was taken where the HTT-unit starts from no configuration to fully filled with patterns. Given the measured rates of IPbus this operation would take between 1 – 2 minutes, which is

---

\(^1\) The main reason why the maximum size was determined to be 128000 words of 32-bits was due to the maximum length of a standard Ethernet frame. Larger values could were picked however that would lead to breaking the Ethernet packet in two or more packets and thereby affecting the transfer rate measurement.
far from a limiting factor. Based on those measurements the IPbus protocol was chosen to be the monitoring protocol of the HTT system, with the possibility to be used as configuration protocol as well.

6.5.2 AM thermal tests

The main functionality of the PRM06 firmware was developed to measure the temperature of the AM chips on the test-bed setup. The goal was to ensure that the chip temperature under basic operations remains under control before inserting them in the ATCA environment. To achieve this measurement a set of I2C sensors were mounted on-top of all the AM chips. The I2C cables were driven by a 2.5 volt FPGA bank via a 1 inch small form factor connector present on the PRM06. The picture of the temperature sensors can be seen in fig. 6.24. The sensors were initially functionally validated with an Arduino board.

The sensors were then mounted with thermal paste on-top of the ASIC and connected to the dedicated connector. For the work-bench set-up accessing and reading the temperature from the FPGA, happened with the IPbus protocol. Since, the IPbus protocol foresees multiple requests at a time, measuring the temperature from the AM chips was decided to be performed in a continuously way along the configuration requests for the data transmission.

To read the temperatures a dedicated software component was developed to monitor each sensor online while the data are transmitted. The output of the sensors is plotted online during the tests and a record is kept in form of text files to allow playing older data back through the plotting system. The software output can be seen in fig. 6.25, showing the temperature of the chips when 100 data sent to the AM chip were matched to pre-loaded patterns.
Multiple factors drive the temperature behavior of the ASICs. The main two are the serial transceivers receiving the data, and the pattern matching rate in the AM core. The bare temperature of the transceiver when only the internal core was powered up was found to be 21 degrees. Once, the serial lines are powered up the package temperature increases between 24-25 degrees Celsius, declared as the average operational temperature for the ASIC.

From past studies [91] it was found that a critical temperature of 70 degrees Celsius exists, where the AM chip starts to misbehave. The most common symptom close to the critical temperature is to lose the stored patterns. Performing various different operations with the setup at Sussex the temperature of the ASIC never exceeded the average of 26 degrees with the maximum amount of matching tested to be 1000 hits without any spacing. However, these preliminary exercises described above served mostly the purpose of establishing the firmware and software thermal testing framework.

6.5.3 ATCA back-plane evaluation

The ATCA shelf as described above is currently one of the highest throughput communication options for digital electronics boards available. The HTT system will need to allow board to board communication within a shelf to provide data sharing. The transmission rate for these transactions was set to 10Gbps per communication line, which is well within the margin of the currently used ATCA back-planes [118]. However, ensuring the integrity of those transactions is needed to avoid any data loss.

To measure the back-plane signal integrity of the HTT test facility ATCA shelf the legacy Pulsar-IIb boards were used. All the tests were performed at a line rate of 10Gbps with the use of the 100Gbps Air-plane back-pane from COMTEL. The main goal for those tests was to characterize the physical connections in the back-plane with minimal tuning on the FPGA serial transceiver. The main reason for the minimal tuning constraint arises from the fact that the per-line and
per-board tuning across the full scale HTT system is an onerous step, which can be avoided since the target bandwidth is well below the design bandwidth of the back-plane.

**Bit Error Rate theory**

Measuring the signal integrity on serial communications was historically performed with the so called eye scans measuring the number of bits in a given pattern that were decoded wrongly over time. This method is called bit error rate and is nominally performed with the use of oscilloscopes attached to a given transmission line. However, since line rates increase their speed rating and PCB layouts become more complicated. Accessing and performing eye scans with oscilloscopes becomes more and more difficult, due to probing difficulties. Additionally, the signal amplitudes were decreased, making the probing with an oscilloscope disruptive to the waveform itself, due to the capacitance introduced by the probes. Solution to the issue is provided by the FPGA vendors with integrated tools in form of firmware blocks that can perform measurement of the signal integrity on the FPGA ports allowing more accurate results, due to lower jitter and noise.

The procedure to measure the signal integrity with an eye scan is performed always on the receiver side of a transceiver with an agreed pseudo-random pattern between transmitter and receiver. The circuit performing the measurements of the bit error rate can be seen in fig. 6.26. In more detail an incoming serial link is captured by the receiver with certain characteristics set on the transmitter side. Since any data sent over a link are subject to attenuation and distortion a filter is applied on the receiving side to clean the incoming signal. This stage is what can be seen as the equalization box in fig. 6.26. After the equalization step the incoming pattern is compared to the expected one. The comparison to the pattern will provide a certain amount of received bits being different than the pattern expectation. The ratio of erroneous bits over the total amount of bits received, defines the so called bit error rate. Following the calculation of the bit error rate a modification is applied to the data sampling time relative to the incoming

![Figure 6.26: Receiver architecture for eye scan generation based on pre-agreed pseudo random pattern between transmitter and receiver [123].](image-url)
stream effectively changing the conditions and performing another bit error rate calculation. In addition, the differential voltage threshold in the receiver pattern is altered, creating in practice a two dimension grid where each point is a bit error rate calculation for a given sampling time window and voltage threshold [143].

An example of the resulting "eye" scan can be seen in fig. 6.27, comparing what can be measured with the oscilloscope and what the IBERT firmware allows to measure. The blue area on the right hand side eye scan is a result of the binning of the z-axis. The signal integrity is gauged based on the blue area according to the threshold value. Nominally bit error rates of $10^{-9}$ are required, meaning that the blue area defines a space of less than one bit error over one billion sampled bits. In the HTT system the bit error rate is defined around $10^{-6}$ to ensure safe data transmission.

In a transmitting/receiving pair the transmitter is always driving the link conditions. Improving the integrity of a link depends on various aspects like the certain conditions set on the transmitter, the transmission line itself and the receiver. An example of such a parameter affecting the line integrity is the differential voltage of transmission, normally called differential swing.

**Back-plane comparisons**

To measure the signal integrity in the ATCA back-plane the Xilinx IBERT IP core was used. In Xilinx FPGA nominally the signal integrity is measured on a single transceiver (RX/TX pair), by looping the exiting transmitter trace to the incoming receiver trace. However, in an ATCA setup such configuration is not possible since each transmitter on a given board drives the link which connects a receiver on a different in-shelf board. Additionally, the signal integrity tests need to be performed on conditions that are similar to the actual HTT system. With the following setup each RX/TX pair accesses two physical connections on the back-plane and there after two eye scans need to be performed, for a full transceiver pair on the FPGA. To fully cover a full mesh back-plane fourteen Pulsar-Ilb boards need to be used in order to minimize moving boards in
the shelf. In Sussex only 5 boards were available meaning that a moving procedure between boards in certain slots had to be in place.

In more detail the measuring procedure foresees two boards communicating at a time in order to cover a certain link on the back-plane. A snapshot of one of the board configurations can be seen in fig. 6.28. Both Pulsar-IIb contributing on the measurement of the signal integrity need to be configured with the IBERT firmware block from Xilinx [143]. Since a serial link is asynchronous only four sets of parameters need to be a-priori agreed between the two boards. The first one concerns the pseudo-random pattern transmitted, which needs to be the same between transmitter and receiver in order for the receiver to measure the eye scan. The second parameter is the line rate of the transmitter and the receiver used and the line polarity. Finally, the last parameter concerns the transmitter and receiver conditions such as swing, termination and emphasis. Handling line polarity is achieved via the Dynamic Reconfiguration Port (DRP) [142] of the transceiver, which allows either the receiver to invert the incoming data or the transmitter to invert the outgoing. All lines in the characterization procedure were set to a 10Gbps link speed, with a pseudo-random pattern of 7-bits (PRBS-7) and polarities were set consistent with the board design [119].

Regarding the configuration of the transmitter in the current case the parameter that proved to give an increased performance is the swing. In Vivado its called TX Diff Swing and takes the units of mili-Volt (mV). The swing in a given link is defined based on the length of the trace, and thereafter its value depends on the board positioning in the crate. Given the trace length of the Fabric links in the COMTEL back-plane being used, a value of 756mV swing has found to provide an improved performance from the extreme trace length case all the way to the closest possible connection. From here on this swing value will be referred to as the default value, in contrast to the default value set by Vivado which is of the order of 100mV.

Fig. 6.29 illustrates a typical eye scan obtained with the parameter values identified above. The horizontal axis is parameterized in sampling time steps, and the vertical axis in voltage steps. A step in sampling time corresponds to 30ps whereas a voltage code corresponds to 1.5mV. The blue central area corresponds to the minimum desirable bit error rate of $10^{-9}$.

Due to concerns about potential cross-talk effects, the IBERT firmware was prepared so that individual links could be driven with the other kept at electrical idle state. In addition to the
6.5 Lab tests

Fig. 6.29: Eye scan on ATCA back-plane link at 10Gbps.

firmware flavour with all the connections transmitting PRBS-7 data. Fig. 6.30 summarizes the signal integrity tests: the 2D grid is organized by slot positions (x and y axes) and each plot corresponds to the eye scan for the link connecting those two positions. A few remarks can be made on the outcome of the characterization. The first one is about the missing scan in the grid. The line connecting those boards was not able to establish a stable link with neither of the two firmware configurations. Further investigation pointed to a damaged pin in the back-plane and hence this connection has not been evaluated. The second remark is about the number of scan pairs seen in the grid. In a 14 slot full mesh back-plane a 14x14 table would be expected however the table is 12x12, since two slots are occupied by the ATCA switch described in sec. 6.3.2. For this reason those slots were considered special and removed from the study.

Figure 6.30: Full-mesh ATCA back-plane signal integrity tests at 10Gbps.
6.5 Lab tests

The scans were compared with the reference scans provided by the Pulsar-IIb designers in order to identify links that deviate a lot from the reference and further investigate. A deviation compared to what obtained in the Pulsar boards qualification would most likely point at a back-plane issue. The reference plots obtained from the validation in Fermilab can be seen in fig. 6.31 [119]. No substantial differences were found between the two tests effectively proving that by just changing the differential voltage from the transmitter an acceptable bit error rate can be achieved.

The characterization study illustrated in this section allowed for the familiarization with the ATCA technology, as well as allowed the progress towards the quality control procedure for the HTT TP boards planned to be produced in the UK. Pathologies of the Sussex ATCA shelf were found and thereafter allowed for a better understanding of the set-up, in view of the QA. In addition, it showed that only a single configuration of the serial transceivers can be used in the TP firmware up to a bit error rate of $10^{-9}$. Further improvement to the link integrity requires more tuning of the transceiver parameters.

6.5.4 Integration into HTT online software

The last contributions by Sussex HTT group, during the time of this thesis concerns the integration of the test setup with the HTT online software under development [102]. The online software is made from a set of classes that are inspired from the current Run Control [107] infrastructure of ATLAS: a set of standardized tools and interface classes providing access to the various detector sub systems during data taking. The HTT operation will be monitored and configured as a sub module of the Run Control tree and thereafter all its monitoring data need to be interfaced with it. Given the choice of using IPbus as the main slow control mechanism for HTT, part of the HTT online software development work required its integration with the uHAL libraries.
The test facility provided sufficient experience with the IPbus protocol and its use along with its firmware. For this reason a test configuration was provided to the HTT community to allow gain experience on testing the developed firmware with IPbus. The test setup provided two components, the software handling a set of operations and the firmware component which was developed to be loaded on a standard Xilinx Kintex-7 Evaluation card. An overview of both components is presented in the following two sub section along with some successful examples of their operation.

Sussex software

The Sussex software for the HTT test-bed was developed mainly in Python. The online software consists, however of C++ classes with specific interfaces. The Python software was translated into C++ in order to be integrated with the online software infrastructure. The decision of following a modular approach in the development made the integration straight forward, following the structure of the example design seen in fig. 6.32 below.

The various sub modules of the software are linked to specific functions in the firmware that communicates with the AM chip. The command manager is one of main sub-modules it converts all the data into the required format for the JTAG master control to access and update registers in the JTAG chain. The JTAG sub-module encodes the JTAG specific data into 32-bit words for the IPbus slaves. Dedicated unpacking is performed to the data retrieved from the JTAG slaves in the firmware. The SERDES class configures and enables the transceivers on both the AM side and the FPGA. The chip manager class is dedicated to define the group of AM chips and append the required amount of data so that a specific chip in a given location on the board can be accessed through the JTAG daisy chain. Finally the two remaining classes are the AM chip class and the utils class. The AM chip class is responsible of holding the unique information for all the chips in a daisy chain to help the chip manager, and the utils class is responsible on providing a higher level interface for the lower lever IPbus specific function like dispatch.
write or read. For the example design integrated into the HTT online software only the JTAG class was used, for simplicity and compatibility with the example firmware prepared.

**Example firmware**

The example firmware prepared was based on the standard IPbus implementation and an in FPGA Fabric implemented JTAG slave. In practice a look-up table was connected to the output of the JTAG master in the FPGA which was mimicking the behavior of the standard JTAG register. The main purpose was to provide a complete test-bed setup with both the software infrastructure and a firmware block equivalent to the AM chip. The operation of the look-up table JTAG slave is summarized in table 6.2 where the IR [124] column written value and DR [124] column value returned after a successful read.

<table>
<thead>
<tr>
<th>IR (write)</th>
<th>DR (read)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0x01</td>
<td>0xca</td>
</tr>
<tr>
<td>0x02</td>
<td>0xfe</td>
</tr>
<tr>
<td>0x03</td>
<td>0xba</td>
</tr>
<tr>
<td>0x04</td>
<td>0xbe</td>
</tr>
</tbody>
</table>

Table 6.2: Allowed look up table values and returned information from in-Fabric JTAG slave.

Vivado Simulator. The required resources for the target Kintex-7 evaluation card were obtained via the post-implementation estimation method integrated in Vivado and can be seen in fig. 6.33. The correct behavior of the system is exercised in fig. 6.34 where all the internal signals of the JTAG slave can be seen. The simulation contains the whole IPbus chain after removing all the Ethernet information from the incoming data.

**6.6 Contributions**

In this chapter an overview of the Sussex contribution to the HTT project was provided. I contributed to the project as part of my qualification task which then was extended to preparatory work for the UK HTT test-bed facility. The qualification task contribution concerned the operation of the legacy boards (PRM06, Pulsar-IIB). During this period, extensive experience was
obtained on operating the IPbus protocol and culminating with the measurement of the protocol's latency and throughput. The developed firmware allowed the temperature characterization and operation of the PRM06 and more precisely the AM ASIC mounted on it. I supported the characterization of the ATCA shelf back-plane link performance in view of the upcoming HTT boards. Finally, the test environment for the HTT online software also came as a product of the experience obtained from the use of the legacy hardware, indicating the importance of its existence for the HTT project. The remaining chapters in the thesis will focus on the analysis work I were performed, and more precisely for the first effective lifetime measurement of the $B_s^0 \rightarrow \mu^+\mu^-$ with the ATLAS detector.
Particle physics processes in nature are observed with the use of detectors like ATLAS which was described in the previous chapters. Upgrades on those detectors increase the access to rare processes which can hold hints of NP phenomena. Such a process is $B^0_s \rightarrow \mu^+\mu^-$, which according to ch. 1 provides a handle to NP, via both the branching fraction measurement and the effective lifetime. This chapter will give an overview of the most recent experimental measurements of $B^0_s \rightarrow \mu^+\mu^-$ decays publicly available. More precisely, the state of the art measurement from CMS and LHCb are going to be discussed briefly followed by a deeper discussion on the equivalent ATLAS partial Run-2 branching fractions measurement. The recently published partial Run-2 analysis is of high importance for this thesis physics work, since it was based on the ground work set by the branching fraction analysis.

7.1 CMS measurements

The latest $B^0_s \rightarrow \mu^+\mu^-$ result from the CMS experiment is a combined measurement of both the branching fraction and the effective lifetime of the $B^0_s$. The measurement was based on the partial Run-2 and full Run-1 data-sets to maximize the available statistics [144]. The collected data samples had an integrated luminosity of 5 and 20 $fb^{-1}$ during the 2011 and 2012 periods with an energy of $\sqrt{s} = 7 – 8 TeV$ and 36 $fb^{-1}$ recorded luminosity during 2016 at an energy of $\sqrt{s} = 13 TeV$. The analysis provided an improved branching fraction result compared to the Run-1 only measurement [145] due to the improved muon identification method and the usage of a Boosted Decision Tree (BDT) discriminator.

7.1.1 Branching fraction

The branching fraction analysis methodology was based on separating the data set in multiple BDT bins, on which a simultaneous maximum likelihood fit was performed on the reconstructed invariant mass of the di-muon pair. The BDT has been used in order to minimise the combinatorial background of the analysis. Since this background is produced from muons not originating from the same parent, training a discriminator based on variables (decay vertex position, lifetime value, etc.) that can distinguish the differences between the two decays allows
for event selection where the two muons originate from a B-meson rather than random combinations of muons. The muons were required to have opposite charges and their combined mass to fall in the range of $4.8 < m_{\mu^+\mu^-} < 6.0 \text{GeV}$ [144]. The sub-region within the fit range of $5.20 < m_{\mu^+\mu^-} < 5.45 \text{GeV}$ was declared as the signal region and thereafter blinded to avoid any biases.

The signal models and configurations were developed and studied on MC samples based on the SM values. After the fit model was developed and tested thoroughly, the signal region was un-blinded and fitted with the developed functional templates to extract the number of signal and background. The first term ($P_i(m_{\mu^+\mu^-},\sigma(m_{\mu^+\mu^-}))$) corresponds to a binary observable that discriminates muons bending towards each other $(\text{cowboys})$ or away from each other $(\text{sailors})$. The inclusion of $C$ is due to the possible underestimation of the $B \rightarrow hh$ background decays when the two hadrons are misidentified as muons.

The equation showing the extraction of the branching fraction can be seen in eq. 7.1,

$$\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+\mu^-) = \frac{N_{(s)}^0 f_u}{N_{obs}^{B^+} f_{sid}} \epsilon_{tot} \mathcal{B}(B^+ \rightarrow J/\psi K^+) \mathcal{B}(J/\psi \rightarrow \mu^+\mu^-)$$

where $N_{(s)/obs}^{B^+}$ are the number of reconstructed $B_{(s)}$ and $B^+$ decays, $\epsilon_{tot}$ the efficiencies of the two channels with both values assuming the SM expectations [41], and $f_u/f_{sid}$ is the ratio of the probabilities to create a $B^+$ or a $B_{sid}$ meson, for which the value was measured in CMS independently and then combined with the results from the other experiments ([146], [40]).

The extraction of the number of reconstructed signal events was achieved with a simultaneous fit on the four BDT bins of the invariant mass distribution. The form of the models used for the unbinned maximum likelihood fit can be seen in eq. 7.2.

$$P_i(m_{\mu^+\mu^-},\sigma(m_{\mu^+\mu^-}),C) = P_i(m_{\mu^+\mu^-};\sigma(m_{\mu^+\mu^-}))P_i(\sigma(m_{\mu^+\mu^-})/m_{\mu^+\mu^-})P_i(C)$$

With $P_i$ noting the individual analytical Probability Density Function (PDF) models for both signal and background. The first term ($P_i(m_{\mu^+\mu^-};\sigma(m_{\mu^+\mu^-}))$) contains the analytical model for the mass distribution, the second term ($P_i(\sigma(m_{\mu^+\mu^-})/m_{\mu^+\mu^-})$) indicates the PDF for the mass resolution, which was included separately. The final term ($P_i(C)$) (sometimes also referred as $C$) corresponds to a binary observable that discriminates muons bending towards each other $(\text{cowboys})$ or away from each other $(\text{sailors})$. The inclusion of $C$ is due to the possible underestimation of the $B \rightarrow hh$ background decays when the two hadrons are misidentified as muons.
7.1 CMS measurements

The analytical functional models used for the fit in the invariant mass distribution are:

1. **$B_{(s)}$ signal**: Crystal Ball function (A sum of a double Gaussian core in addition to an exponential tail) [147] superimposed with Gaussian kernels, for both $B^0_s$ and $B^0_d$.

2. **Peaking background**: Sum of a Gaussian and a Crystal Ball function with common mean.

3. **Semi-leptonic background**: Exponential function.

4. **Combinatorial background**: Non-negative Bernstein polynomials [148] of the first degree.

In the fit only the parameters of interest to the $B_{(s)}(s) \to \mu^+\mu^-$ measurement are left free. All remaining parameters are subject to Gaussian constraints introduced in the likelihood [144].

The fit result on the highest BDT bin can be seen in fig. 7.1. Combining all the results obtained for all the different sub-sets (forward/central), the data taking periods and the different BDT bins yielding the measurement of the branching fraction seen in eq. 7.3.

$$\mathcal{B}(B_{(s)} \to \mu^+\mu^-) = [2.9^{+0.7}_{-0.6}(exp.) \pm 0.2(frag.)] \times 10^{-9}$$ (7.3)

As experimental uncertainty the combination of the statistical uncertainty and the dominant systematic sources is noted. The leading uncertainty contribution, called fragmented, is arising from the fraction of $f_s$ and $f_u$ used in the branching fraction eq. 7.1 [144]. For the $B^0$ an insignificant result was obtained since the available statistics in the studied data set were low, yielding an upper limit seen in eq. 7.4.

$$\mathcal{B}(B^0 \to \mu^+\mu^-) < 3.6 \times 10^{-10}$$ (7.4)

at a 95% confidence level. The obtained results are consistent with the results extrapolated from the CMS Run-1 data only analysis [145] and the theory predictions. In addition, the obtained value is consistent with the SM expectation [41] including the one order of suppression more for the $B^0_d$ branching fraction in compared to the $B^0_s$, shown in fig. 7.2.
7.1 CMS measurements

119

Figure 7.2: Likelihood contours from the branching fractions fits of $B^0_s$ and $B^0_d$ along with the SM predicted value for the CMS collaboration [144].

7.1.2 Effective lifetime

For the CMS effective lifetime measurement the same selection criteria to the branching fraction was applied to data. However, instead of using various BDT bins only one bin was used to increase the available statistics, and thus the sensitivity on the lifetime parameter. The result was extracted with two independent methods in order to allow for validation and a better study of the systematic uncertainties. The first method that was used as the primary method due to its better median performance of the parameter of interest (e.g., lifetime), validated prior to unblinding, was based on a 2D unbinned maximum likelihood fit in both the invariant mass and decay time distributions [149]. The second approach contains a 1D binned fit on the background-subtracted, signal-only distribution. The subtraction was achieved with the use of the statistical tool called sPlot [150].

The fitting range in decay time for both approaches was defined as a region of $1 < t < 11 \text{ps}$ since for low decay times the CMS detector was found to suffer from low reconstruction efficiency due to the flight-length significance and isolation requirements applied on the muons [145]. The limiting factor in both measurement approaches for the effective lifetime arises from the low number of signal events, effectively making the statistical uncertainty the dominant uncertainty, with the systematics having a limited impact.

Both methods yield results that are consistent with the SM expectation. However, exclusion of any BSM scenario due to the presence of light state decays (sec. 1.3.2) is impossible, since the uncertainty in both cases is very large.

Two-dimensional unbinned maximum likelihood method

For the first fitting approach a similar method as for the branching fraction part was used, with analytical templates for both invariant mass and decay time. However, the models describing the invariant mass were simplified to align with the fit being performed in a single BDT bin rather than multiple. The overall PDF in the two dimensions (inv. mass and decay time) can be
seen in eq. 7.5.

\[ P(m_{\mu^+\mu^-}, t; \sigma_t) = N_{\text{sig}} P_{\text{sig}}(m_{\mu^+\mu^-}) T_{\text{sig}}(t; \sigma_t) \epsilon_{\text{sig}}(t) + N_{\text{comb}} P_{\text{comb}}(m_{\mu^+\mu^-}) T_{\text{comb}}(t; \sigma_t) \]

\[ + N_{\text{peak}} P_{\text{peak}}(m_{\mu^+\mu^-}) T_{\text{peak}}(t; \sigma_t) \epsilon_{\text{peak}}(t) + N_{\text{semi}} P_{\text{semi}}(m_{\mu^+\mu^-}) T_{\text{semi}}(t; \sigma_t) \epsilon_{\text{semi}}(t) \]  

With \( N_{\text{sig/comb/peak/semi}} \) the corresponding yields obtained from the fit. \( P \) and \( T \) define the analytical PDF models in mass and proper time and finally the \( \epsilon(t) \) are the efficiencies for the various components estimated with simulated events [144].

The models used to fit the mass distribution are slightly different from the corresponding ones from the branching fraction analysis and can be seen along with their decay time counterparts in table 7.1. All decay time models were corrected with their corresponding efficiency factors.

<table>
<thead>
<tr>
<th>Event source</th>
<th>Inv. mass</th>
<th>Decay time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal</td>
<td>Crystal Ball function</td>
<td>Convoluted exponential with Gaussian</td>
</tr>
<tr>
<td>Combinatorial</td>
<td>Bernstein polynomial O(1)</td>
<td>Convoluted exponential with Gaussian</td>
</tr>
<tr>
<td>Peaking</td>
<td>Crystal Ball function</td>
<td>Convoluted exponential with Gaussian</td>
</tr>
<tr>
<td>Semi-leptonic</td>
<td>Bernstein polynomial O(1)</td>
<td>Convoluted exponential with Gaussian</td>
</tr>
</tbody>
</table>

Table 7.1: Signal and background analytical PDFs for 2D unbinned maximum likelihood fit for both invariant mass and decay time.

except for the combinatorial background because its shape arises directly from data. The signal yield, the lifetime parameter and the combinatorial background are allowed to vary freely in the fit. In contrast, all the other remaining parameters were either constrained or fixed to the MC expectation. Fixing and constraining parameters provides a further source of systematics that was studied [144].

The effective lifetime obtained can be seen in eq. 7.6, with the uncertainty being the combined (statistical and systematic).

\[ \tau_{\mu^+\mu^-} = 1.70^{+0.61}_{-0.44} \text{ps} \]  

(7.6)

The systematic uncertainty only was found to be \( \sigma_{\text{syst}} = 0.09 \text{ps} \) [144]. The result of the fit can be seen in fig. 7.3 with the corresponding PDF projections plotted on top of binned data, in both invariant mass and decay time.

**One-dimensional binned fit on sPlot signal projection**

The alternative method used to measure the lifetime is based on the sPlot method. The details of sPlot can be found in app. B, however a coarse overview of the method is given here. On a pair of uncorrelated variables an unbinned maximum likelihood fit is performed on the variable where the analytical PDFs are known (e.g. invariant mass). From the fit result, a set of per event weights is produced for each event source including the correlations existing between the sources. The weights of a particular source are applied on the variable of interest (e.g. decay time) to extract the subtracted distribution.

In the case of the CMS experiment the invariant mass is the independent variable with the analytical PDFs taken from the branching fraction analysis. The background subtracted decay time distribution, is fitted with a binned exponential model convoluted with a Gaussian, to
7.1 CMS measurements

Figure 7.3: The result of the 2D unbinned maximum likelihood for the extraction of the $B_s^0$ effective lifetime \cite{144}. Blue indicates the full PDF for the fit, red the signal model for the $B_s^0$, purple the signal model for the $B^0_{d\ell}$, green the semi-leptonic background (Separate semi-leptonic B decays wrongly associating the muons with a single B) and finally dark blue the combinatorial background (prompt muons reconstructing the same mass as the B candidate) within the mass window.

include the resolution and efficiency terms arising from the selection and detector resolution. Applying the standard maximum likelihood fitting method is not trivial for weighted events and some modification to the likelihood parameterization is required. For this reason a custom algorithm was implemented to perform a one dimensional weighted histogram fit based on ref. \cite{151}.

The fit result obtained with the custom algorithm can be seen in eq. 7.7 with both the statistical and systematic uncertainty combined in quadrature.

$$\tau_{\mu^+\mu^-} = 1.55^{+0.52}_{-0.33}\, ps$$ (7.7)

The contribution of the systematic uncertainty in the combined uncertainty is as expected smaller than the statistical ($\sigma_{\text{syst}} = 0.12\, ps$), with the main source arising from the negative bins that are not allowed in the weighted likelihood fit. To avoid negative bin contents a variable binning approach was followed to ensure that in every fitted data set all bins are positive. The fit result projection on the binned data set can be seen in fig. 7.4.

The CMS analysis performed the first measurement of the $B_s$ to two muons effective lifetime from a general purpose detector at CERN. The resulting value hinted the major limitation of the analysis originating from the limited statistics and hence not capable to provide a strong preference of any of the mass eigenstates. However, combining the results from all major experiments at CERN (LHCb, ATLAS, CMS) will give a better handle for a 5 $\sigma$ measurement on the effective lifetime measurement. For this reason it is highly important for both ATLAS and LHCb to provide their measurements for a potential final combination which will hint whether NP is affecting the process or not.
7.2 LHCb result

Prior to CMS the first result containing both the branching fraction and the effective lifetime measurement for the $B_{(s)}$ di-muon decays was from the LHCb collaboration. The first publication from LHCb contained the result obtained with the partial Run-2 data combined with the data collected during Run-1, to increase as much as possible the sensitivity [146]. More recently, the collaboration updated their result with the full Run-2 data set, however an official publication is still pending at the time of this thesis.

The integrated luminosity and energy for the Run-1 data was 1 fb$^{-1}$ of pp collisions at $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV. For the partial Run-2 the integrated luminosity was 1.4 fb$^{-1}$ at $\sqrt{s} = 14$ TeV. The analysis provided a more than 5$\sigma$ measurement on the branching fraction of the $B^0_s$ meson, which was the first observation of the decay from a particle physics experiment in the LHC. For the $B^0_d$ an observation was not achieved. Finally, the analysis provided a measurement on the effective lifetime for the $B^0_s \rightarrow \mu^+ \mu^-$ [146].

7.2.1 Branching fraction

The methodology for the branching fraction is largely the same as in the CMS analysis. The $\mathcal{B}(B_{(s)} \rightarrow \mu^+ \mu^-)$ is measured precisely with the use of the branching fraction of some reference channels, seen in eq. 7.8.

$$\mathcal{B}(B_{(s)} \rightarrow \mu^+ \mu^-) = \mathcal{B}_{\text{norm}} \frac{f_{\text{norm}}}{f(s)} \epsilon_{\text{norm}} \frac{N_{\text{obs}}^{(s)}}{N_{\text{obs}}^{\text{norm}}}$$  (7.8)

Eq. 7.8 is very similar to 7.3 discussed for the CMS case. The main difference relies in the decays used in the reference channel. LHCb is by nature more sensitive to B-meson decays than ATLAS and CMS. For this reason the normalization channel used in LHCb contains in addition to the $B^+ \rightarrow J/\psi K^+$, the $B^0 \rightarrow K^+ \pi^-$ decay, which is kinematically very similar to the signal channel. Having very similar decays in the normalization and in the analysis channel yields cancellation of the systematics on the observation efficiencies and the uncertainties on the $f_s/f_d$ ratio.

The goal of measuring the branching fraction was achieved with a very similar approach to
7.2 LHCb result

Figure 7.5: Simultaneous fit result projection in the highest BDT bin plotted on binned data [153]. The red line notes the $B^0_s$ signal, the green the $B^0_d$ ($B^0$ in LHCb notation) the signal model, the dashed blue shows the combinatorial background (equivalently defined as CMS and ATLAS), magenta indicates the peaking background from two hadrons faking muons, blue, light blue indicate the partially reconstructed background sources and finally the purple and orange show the $B^+_c$ and $\Lambda^0_b$ background sources.

CMS. A BDT discriminator was trained and applied on data to reduce the background contribution in the analysis mass window. The data set is split into 4 bins in BDT on which a simultaneous extended maximum likelihood fit [151] was performed on the invariant mass distribution. The data-set of Run-1 and Run-2 are kept distinguished, however BDT binning remains the same between the two sets, hence 8 total categories are included in the invariant mass fit. The analytical PDF models used to describe the invariant mass distribution can be found in the list below.

1. $B_{(s)}$ **Signal**: Crystal Ball function.

2. **Peaking background**: Double Crystal Ball function with common mean.

3. **Semi-leptonic background**: Argus function [152] convoluted with a Gaussian.

4. **Combinatorial background**: Exponential function.

The shape parameters of the PDF components are all constrained with Gaussians with their values extracted from fits on the MC samples. The slope of the exponential describing the combinatorial background includes a constant dependence across the BDT bins [153]. The fit result on the highest BDT bin in Run-1 and partial Run-2 data can be seen in fig. 7.5. The resulting branching fraction measurement can be seen in eq. 7.9 for both $B^0_s$ and $B^0_d$.

$$\mathcal{B}(B^0_s \rightarrow \mu^+ \mu^-) = (3.0 \pm 0.6^{+0.3}_{-0.2}) \times 10^{-9}$$

$$\mathcal{B}(B^0_d \rightarrow \mu^+ \mu^-) = (1.5^{+1.2+0.2}_{-1.0-0.1}) \times 10^{-10}$$ (7.9)

The first uncertainty is quoted as the statistical and the second as the combined systematic uncertainty. The resulting values yield to a 7.8$\sigma$ measurement on the $B^0_s$ branching fraction.

*Figure 7.5: Simultaneous fit result projection in the highest BDT bin plotted on binned data [153]. The red line notes the $B^0_s$ signal, the green the $B^0_d$ ($B^0$ in LHCb notation) the signal model, the dashed blue shows the combinatorial background (equivalently defined as CMS and ATLAS), magenta indicates the peaking background from two hadrons faking muons, blue, light blue indicate the partially reconstructed background sources and finally the purple and orange show the $B^+_c$ and $\Lambda^0_b$ background sources.*
and an upper limit for the $B_d$ [146]. The likelihood contours for the $B_s^0 - B_d^0$ branching fractions, with the SM expectation value can be seen in fig. 7.6.

![Figure 7.6: The likelihood contours obtained from the branching fraction analysis with the central value next to the SM expectation. The error bars on the SM prediction indicate the theory uncertainty pointed out in ch. 1 [153].](image)

### 7.2.2 Effective lifetime

In addition to the branching fraction measurement, LHCb published the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime analysis. The overall procedure is similar to the second approach of CMS (sec. 7.1.2), summarized as following. The lifetime measurement is performed on the background subtracted decay time distribution. For the subtraction, the sPlot method with an unbinned maximum likelihood fit is performed on the invariant mass to extract the required subtraction weights. The resulting signal projection in decay time is then fitted with an unbinned maximum likelihood fit where the $sWeights$ are included with the use of the $sFit$ [154] method.

The main limitation for the effective lifetime measurement arises from the number of $B_s^0$ candidates observed in the data set. For this reason, the Run-1 and partial Run-2 data sets were combined to maximize the total number of $B_s^0$ mesons. In addition, the muon identification criteria were loosened and a single BDT cut was applied to create a single bin rather than multiple. Finally, an optimization on the configuration of the mass fit was performed to determine the best choices in PDF models and fit range [153].

To optimize the fit range and models a large number of pseudoexperiments were used to identify the best result. For this study, the best configuration was identified to be a fit on a reduced invariant mass range compared to the branching fraction analysis fit. The range is $5200 < m_{\mu^+\mu^-} < 6000$ MeV, which maximizes the signal yield in addition to significantly reducing the background contribution. Reducing the number of background sources allows the simplification of the combined mass model to only contain two components seen in the list below.

1. **$B_s^0$ signal**: Crystall ball function, same as in the branching fraction.

2. **Combinatorial background**: Exponential function, same as in the branching fraction.
For both models the shape parameters were constrained with Gaussians with values obtained from fits on MC samples.

For the lifetime distribution fit the used PDF model was an exponential convoluted with a Gaussian and can be seen in eq. 7.10.

\[ P(t) = \epsilon(t) \times e^{-t/\tau} \] (7.10)

The modeling of the efficiency curve for the \( B^0_s \) signal was based on the \( B^0 \to K^+ \pi^+ \) normalization, since they are kinematically very similar. For the background a combination of two PDFs were used to model the efficiency term. The two components were summarized by a long-lived and a short-lived acceptance function. The exact parameterization of those components was achieved with the use of the \( B \to h^+ h^- \) combinatorial background decays [153].

Following the determination of the various fit models in both invariant mass and decay time the application of the sPlot procedure was used for the background subtraction. At this stage a major difference exists between the CMS and the LHCb results. As discussed above the CMS result relies on the standard parametrization of a weighted maximum likelihood fit on a binned data-set [151], excluding negative weight events. The LHCb measurement in contrast relies on the sFit tool which allows the use of the \textit{sWeight} as created by the statistical tool itself, however the resulted fit to the decay time leads to a mis-estimation of the statistical uncertainty due to the sPlot uncertainty being related to the sum of weights rather the number of actual observed candidates [146]. For this reason, a correction factor was applied on the derived \textit{sWeights} seen in eq. 7.11,

\[ w'_i = w_i \frac{\sum_j w_j}{\sum_j w_j^2} \] (7.11)

where \( w_i \) are the \textit{sWeights} per candidate and \( w_j \) are the \textit{sWeights} summed over all decays [154].

The correction applied on the weights, leads to a correct estimation of the statistical uncertainty returned by the fit.

The result of the one dimensional fit on the signal projection can be seen in eq. 7.12.

\[ \tau_{\mu^+\mu^-} = 2.04 \pm 0.44 \text{ps} \] (7.12)

The quoted uncertainty contains both the statistical and the systematic uncertainty (\( \sigma_{\text{syst}} = 0.09 \text{ps} \)). The fit result on the lifetime with the model projection super-imposed along with the invariant mass fit required to calculate the \textit{sWeights} can be seen in fig. 7.7.

The final result was found to be consistent with the SM expectation, with a \( 1.0 \sigma \) distance from the theoretical expectation, and consistent with the most extreme deviation for BSM with \( 1.4 \sigma \) distance.

### 7.2.3 Full Run-2 improvements

In 2021 the LHCb collaboration announced their preliminary results based on the full Run-2 data set. Although an official publication describing the exact changes in the procedure with respect to the previous version described above was not available at the time of this thesis, the
7.2 LHCb result

Figure 7.7: The fit result on the combined Run-1 and Run-2 data-set from LHCb. (a) Is the invariant mass fit required to generate the sWeights for the background subtraction, (b) the fit on the $B_s^0 \to \mu^+ \mu^-$ signal projection[146].

mass plots and results are publicly available.

On the branching fraction measurement the LHCb Run-1 and partial Run-2 result yielded a remarkable 7.6σ observation for the $B_s^0 \to \mu^+ \mu^-$, and an upper limit for the corresponding $B_d^0$ di-muon decay. The likelihood contour in the two branching fraction plane showed a 1σ deviation from the predicted SM (fig. 7.6). Integrating the full Run-2 data set into the analysis yields a 10.6σ observation of the $B_s^0 \to \mu^+ \mu^-$ decay branching fraction, which can be seen in eq. 7.13.

$$\mathcal{B}(B_s^0 \to \mu^+ \mu^-) = (3.09^{+0.46+0.15}_{-0.43-0.11}) \times 10^{-9}$$ (7.13)

Compared to the previous version of the analysis, a significant improvement in both the systematic and statistical uncertainty was achieved. The result of the mass fit in the highest BDT bin, needed to extract the number of observed $B_s^0$ mesons, can be seen in fig. 7.8.

For the $B_d \to \mu^+ \mu^-$ decay an observation was again not successful, thus setting a more stringent upper limit seen in eq. 7.14.

$$\mathcal{B}(B_d \to \mu^+ \mu^-) < 2.6 \times 10^{-10} \quad (95\% CL)$$ (7.14)

The corresponding likelihood contour plot plotted on top of the equivalent from the previous version of the analysis can be seen in fig. 7.9, showing a consistent result with the partial Run-2 analysis.
In parallel with the branching fraction analysis the $B^0_s$ effective lifetime measurement was also improved, due to the improved statistical strength of the full Run-2 sample. The analysis procedure was largely the same as in the partial Run-2 round, however the cut applied in the invariant mass distribution was optimized to include the highest possible value of the $B^0_s$ yield, on which the lifetime statistical uncertainty depends. In addition, a simultaneous fit on two BDT bins was used, instead of one. The new invariant mass window containing the signal and the combinatorial background only was adjusted to $5320 < m_{\mu^+\mu^-} < 6000 \text{MeV}$. The fit to extract the required $sWeights$ for the background subtraction can be seen in fig. 7.10 for the two different BDT bins.

The resulting background subtracted decay time distribution obtained by sPlot can be seen in fig. 7.11 for both BDT bins. The analytical model used to fit the decay time distributions were the same models used in the previous version of the analysis. The lifetime obtained from the
Figure 7.10: Invariant mass fit for signal sWeight extraction in two different BDT bins.

Figure 7.11: Decay time distribution after sPlot weights applied for two analysis BDT bins.

The full Run-2 result proved to be 1.5\(\sigma\) compatible with the SM prediction and 2.2\(\sigma\) with the extreme scenario \((A_{\Delta r} = +1)\) of BSM prediction.

All the results in this sub-section were obtained from the values presented during Moriond 2021 by the LHCb collaboration. As mentioned previously, at the time of the thesis the official paper was not yet released, therefore an overview is given rather than an in-depth discussion.

### 7.3 ATLAS result

The last LHC experiment discussed in this chapter is ATLAS, which was first to published the branching fraction result [40]. However, the branching fraction result came with the cost of not publishing a result on the effective lifetime as done by the two experiments on the partial Run-2 data-sets.

The analysis work of this thesis is heavily dependent on the work done for the branching fraction analysis and therefore, a greater detail was given. The discussion is mostly limited to the partial Run-2 result, and when necessary will refer to the first ever version of the analysis with only the Run-1 data. In addition, the current state of the art of the full Run-2 analysis will be briefly discussed. This is expected to be published late in 2022.
7.3 ATLAS result

Samples and triggers

The samples used to perform the analysis were collected during the 2015 and 2016 data taking periods of the LHC. During this time, the energy of the accelerator was $\sqrt{s} = 13\text{ TeV}$ and the collected luminosity from ATLAS was recorded at $36.2\ f\text{b}^{-1}$ with an uncertainty of 2.1%, mostly taken during 2016 [155]. The evolution of the integrated luminosity in both 2015 and 2016 can be seen in fig. 7.12.

The trigger chains employed for the $B_s \to \mu^+ \mu^-$ analysis are based on the existence of two muons in the final state that are captured by the MS. Depending on the instantaneous luminosity, a set of different low-$p_T$ thresholds are available. All the available options have in common a full track reconstruction performed at the level of the HLT on the muon candidate. In addition to the track reconstruction, the HLT applies also a loose di-muon invariant mass cut in the range of $4000 < m_{\mu^+ \mu^-} < 8500\text{ MeV}$ [157].

The chains providing the best available luminosity for the branching fraction analysis required a leading muon with a $p_T > 6\text{ GeV}$ and a sub-leading muon with a $p_T > 4\text{ GeV}$. The available triggers at the HLT level can be seen in the following list, where all trigger items are seeded by the same L1-trigger item (L1_MU6_2MU4; checking that $\mu_1 > 6\text{ GeV}$ and $\mu_2 > 4\text{ GeV}$) at the level of the hardware based trigger system.

1. **HLT_mu6_mu4_bBmumu**: Di-muon trigger with $p_T$ cut on the muons at 4 and 6 GeV. During 2015 the trigger was almost completely unprescaled, however in 2016 it was heavily prescaled in favor of the next item in the list.

2. **HLT_mu6_mu4_bBmumu_Lxy0**: Di-muon trigger with same $p_T$ cuts as the item above with an additional requirement that $L_{xy} > 0$, where $L_{xy}$ is the projection of the distance between the PV and the Secondary Vertex (SV) on the B-meson transverse momentum evaluated at the HLT level. To optimise the vertex identification the analysis performed an extensive study to cross check various methods used by ATLAS, and finally concluded that the approach used in Run-1 is the most suitable. The vertex is defined based on the
distance of the tracks in the z-axis direction [155]. The trigger was not active at all during 2015 but largely un-prescaled during 2016.

In both cases a loose inv. mass cut was applied at trigger level around the B-meson mass (4.6 < $m_{\mu\mu}$ < 5.9 GeV) Both trigger's effective luminosity, during all the periods, sums to 26.3 $fb^{-1}$, which compared to the total available luminosity during the same periods results in a reduction of the available statistics due to prescaling by a factor of 1.4 [155]. Including another trigger item (e.g. HLT_2mu6_bBmumu_Lxy0) would result in recovering the lost luminosity from the prescales, however studies performed on MC yielded a limited improvement in the result. Therefore, to keep the analysis complexity under control no additional trigger selection was added in this version, and left as a potential improvement when the full Run-2 data is included.

For the control ($B \to J/\psi \phi$) and reference channel the trigger strategy was the same with the only difference applied on the invariant mass cut imposed by the trigger arising from the fact that the mass of the muons (2.4 < $m_{\mu\mu}$ < 4.3 GeV) has to be equal to the mass of the $J/\psi$, rather than the B meson. Both channels (control, reference) were used to minimize systematic uncertainties and improve the measurement on the $B(s) \to \mu^+\mu^-$ branching fraction. For this reason, the same L1 selection was used in the trigger selection (L1_MU6_2MU4) and a very close strategy for the corresponding HLT chain, seen in the list below, with the difference on the invariant mass of the muons as discussed.

- **HLT_mu6_mu4_bJpsimumu**: Equivalent to the first $B(s)$ trigger chains with muon $p_T$ cuts at 4 and 6 GeV respectively. This trigger was active only during 2015.

- **HLT_mu6_mu4_bJpsimumu_Lxy0**: Trigger that introduces the cut on the $L_{x y} > 0$ variable on top of the selections used for the previous object. The trigger was prescaled in the first periods of 2016 but more towards the end, since it was covered by the next trigger.

- **HLT_mu6_mu4_bJpsimumu_Lxy0_delayed**: Same trigger as the previous item; however, data were re-directed from the main stream to the so called delayed stream.

The total effective integrated luminosity collected for the reference and control channel results to 15.1 $fb^{-1}$, which translates to a prescale factor of 2.4. Following the trigger strategy presented above, in both the reference/control and signal channel, a two-fold increase in statistical power with respect to the equivalent Run-1 analysis was estimated [158].

A detailed description about the MC samples used along with further selections applied on them was provided in app. A. Since the effective lifetime analysis, which is the main topic of the physics contribution of this thesis, is based on the samples produced for the 15/16 branching fraction analysis, decisions outlined in this section affect various decisions and results seen in the effective lifetime chapter below.
Analysis strategy

The strategy followed for the ATLAS analysis is mostly based on the previous round that concerned the Run-1 data [158] and is very similar to the CMS and LHCb analysis procedures described above. The formula used to extract the result of the branching fraction for both the $B_s \rightarrow \mu^+\mu^-$ and $B_d \rightarrow \mu^+\mu^-$ can be seen in eq. 7.16, and is very similar to eq. 7.1 used in the CMS experiment [155].

\[
\mathcal{B}(B_{(s)} \rightarrow \mu^+\mu^-) = \frac{f_u}{f_{s(d)}} \times \mathcal{B}(B^+ \rightarrow J/\psi K^+) \times \mathcal{B}(J/\psi \rightarrow \mu^+\mu^-) \times \frac{N_{B_{(s)}}^{obs}}{N_{J/\psi K^+}^{obs}} \times R_{Ac} \tag{7.16}
\]

where the variables have the following meaning:

- $f_u/f_{s(d)}$: The probability ratio for a $b$-quark to hadronize into a $B^+$ or a $B_{(s)}^0$, by interacting with a sea quark. In $pp$ collisions $b$-quarks are generated in $b\bar{b}$ pairs, due to colour confinement quarks cannot travel as free particles and hence have to hadronize into mesons or baryons. The most likely environment to obtain the addition quark is from the pair annihilation from sea quarks, leading to the creation of jets in the detector. The ratio is hinting exactly those probabilities for the different quark flavours in the sea. For this reason The value was not measured in the analysing the value obtained by the most recent HFLAV average [34] was used. For the $f_u/f_s$ this value is $0.256 \pm 0.013$, $f_u/f_d$ is assumed to be 1. In order to calculate the ratio certain $b$-kinematics are assumed thereafter a correction factor may be applied in order to make the value from HFLAV directly usable by the analysis. The approach to estimate the correction used was discussed below.

- $\mathcal{B}(B^+ \rightarrow J/\psi K^+) \times \mathcal{B}(J/\psi \rightarrow \mu^+\mu^-)$: The branching fraction for the control channel ($B^+ \rightarrow J/\psi (\rightarrow \mu^+\mu^-)K^+$). The value was obtained from the world average of the two branching fractions: $\mathcal{B}(B^+) = (1.010 \pm 0.029) \times 10^{-3}$ and $\mathcal{B}(J/\psi) = (5.961 \pm 0.033)\%$ [10].

- $N_{J/\psi K^+}^{obs}$: The observed control channel yield extracted from data. The value was calculated with half the data sample, since the other half was used to tune the kinematic distributions of simulated events, effectively avoiding any potential correlation effects [155].

- $R_{Ac}$: The combined term describing the ratio between the acceptances and the efficiencies to detect either the reference or the signal channel. The analytical equation is the following $\frac{A_{ref,\psi K^+}}{A_{sig,\psi K^+}}$, which was derived from MC samples and corrected based on corrections derived from the control and reference channels. A notes the acceptance of the detector and $\epsilon$ the detection efficiency.

- $N_{B_{(s)}}^{obs}$: The number of observed $B_{(s)}$ extracted from data with the use of an invariant mass fit described below.

The analysis used the reference channel since its branching fraction is well measured and allowed the cancellation of the several systematic uncertainties present also in the signal chan-
In addition, the control channel provided access to kinematic variables due to the common $B^{0}_{s}$-meson parent as the signal. An example use of the $B^{0}_{s} \rightarrow J/\psi \phi$ control decay is for the possible correction applied on the $f_u/f_s$ ratio, by studying the momentum and $\eta$ dependence with the yield ratio $N_{B^{0}_{s} \rightarrow J/\psi \phi}/N_{B^{0}_{s} \rightarrow K^{*} J/\psi}$ (with $N$ noting the yield of each decay). Using the results of the Run-1 analysis [159], the value for the probability ratio was found to be: $0.240 \pm 0.0004^{\text{stat}} \pm 0.010^{\text{syst}} \pm 0.017^{\text{theo}}$, which if compared with the average value from HFLAV indicated that no modification was required and hence the HFLAV was used.

An important aspect of the analysis was to avoid inclusions of wrongly identified hadrons as muons. Therefore, an extensive study was performed to minimize the presence of the so called fake muons from $B \rightarrow hh'$ decays. Since decays like this have also a very similar topology and a branching fraction of $O(4)$ higher than the signal, the mis-identification would severely reduce the sensitivity of the analysis.

Fully hadronic $B$-meson decays are not the only source of background that affects the signal. Due to the small predicted branching fraction of $B_{(s)} \rightarrow \mu^+ \mu^-$ the invariant mass distribution without any signal specific selection is overwhelmed with background events. To reduce the background obtained from random muon pairs originating from $b$-quarks from $b\bar{b}$ pairs (called continuum), a BDT based classifier was used. Variables with high discriminating power were identified originating mostly from the kinematics and topology information of the signal candidates.

To utilize the recorded data set, a split in the higher $S/B$ region of the BDT in 4 bins with the same simulated signal efficiency was applied. The number of observed $B_{(s)}$ candidates was extracted with the use of a simultaneous fit on the invariant mass distribution, as was done in both the LHCb and CMS analysis. The models for the invariant mass fit were studied on MC samples and then verified on data [155].

The final branching fraction result was extracted with the use of the frequentest Neyman belt construction [160]. This approach provides a statistical method to extract the confidence intervals on the branching fraction, by allowing the switch between upper limits and measurements to happen naturally. The main reason for choosing the Neyman belt construction arises from the fact that it does not require any assumptions on the distribution of the measured quantity for the extraction of the confidence intervals.

The complete analysis was performed on the blinded data as for the Run-1 version [158], allowing the minimization of any potential biases that might arise due to the specific analyzed sample. Therefore, the following three main steps can be identified before the extraction of the final result:

- Sample/Trigger and event selection optimization.
- Tuning of selections and ingredients to eq. 7.16.
- Blinded region un-blinding and Neyman belt construction for final result.
7.3 ATLAS result

Invariant mass fit

There are two main event categories present in the mass window for the signal yield extraction. The first event category is the background which consists of three different sub categories listed:

- Continuum background.
- Peaking background.
- Partially reconstructed background.

The second category is the signal which contains both the $B^0 \rightarrow \mu^+ \mu^-$ and the $B^0_d \rightarrow \mu^+ \mu^-$. For all the different event sources, an analytical functional model was identified describing best the data and tested when possible on MC or directly in data (e.g. side-bands).

The dominant source of background is the so-called continuum background, which is present in the whole mass window. The continuum background is from real di-muon pairs originating from distinct b-quarks in a $b\bar{b}$ pair [155]. The two muons originating from different b-quarks (different vertices) can be mis-measured to form a common SV vertex and thereafter misidentified as a possible B-meson candidate.

The topology of such a process can be seen in fig. 7.13. Such processes make 99% of the B candidates identified in the mass window at the level of reconstruction, therefore reducing their contributions with the use of the BDT yields an increased sensitivity for the signal. The model used to describe the continuum background is a Chebychev polynomial of the 1st order, with the slope being constrained to have a linear dependence across all the BDT bins. A fit projection on the 4 different BDT bins of the continuum background fit result plotted on top of the MC sample can be seen in fig. 7.14. The slope dependence across the BDT bins can be seen in fig. 7.15, which was validated for its shape on data side bands. A point in fig. 7.15 is the
Figure 7.14: Continuum background fits on MC sample with O(1) Chebychev model. The four BDT bins with their limits can be seen with the resulted fit projection. The data to model compatibility was established with the use of the $\chi^2$ statistical tool [151], although the fits were unbinned maximum likelihood [155].

Figure 7.15: O(1) Chebychev polynomial slope dependence across the 4 analysis BDT bins [155].

average of the BDT distribution in a given bin, and the errors indicate the corresponding root mean square. Using such a constraint across the bins does not contain any direct physics motivation and hence its systematic impact on the final result was studied, and included as an uncertainty.

The second source of background is the so-called partially reconstructed background which contains mostly processes with a real muon and an additional track in the final state coming from the same b-quark. Since the muons identified originate from the same b-quark, the B candidate reconstruction will yield a mass that is smaller than the $B_c(s)$ signal and thus those
7.3 ATLAS result

(a) Same-side decay.

(b) Same-vertex decay.

(c) $B_c$ decays.

(d) Semileptonic decays.

Figure 7.16: Partially reconstructed background visual sketches [155].

As mentioned above, categorizing the partially reconstructed background allows for a better functional description to be derived in order to fit the data. The first two categories of same side and vertex were modeled with an exponential function of the form $f(x) = e^{ax}$ in all 4 BDT.
7.3 ATLAS result

(a) BDT bin 0 (0.1439 < BDT < 0.2455)

(b) BDT bin 1 (0.2455 < BDT < 0.3312)

(c) BDT bin 2 (0.3312 < BDT < 0.4163)

(d) BDT bin 3 (0.4163 < BDT < 1.00)

Figure 7.17: Fits on the two sub-categories of the partially reconstructed background performed on MC samples. Since, the fit is performed on MC weights need to be applied and thereafter the standard $\chi^2$ does not provide good agreement. The use of the Pearson $\chi^2$ [151] was used to test if the models follow the data faithfully [155].

Bins. The projection of the fit in all the different bins can be seen in fig. 7.17. For the $\alpha$ parameter a similar approach was followed as for the Chebychev slope. The dependence across the BDT bins was studied and found to be constant. The equivalent evolution graph showing the fitted slope value as function to the average BDT bin value can be seen in fig. 7.18. Since the study of the exponential slope dependence was based on MC, the observed constant dependence across the bins was included in the study of the systematics for the final signal yield extraction.

On the side of the two remaining categories, similar studies can be performed to extract the best possible distribution. However, for different reasons their contributions were found to be small in the mass window of the analysis, hence adding analytical PDFs would complicate the overall fit. The choice of not modelling the sources analytically produces a systematic effect, which was studied and discussed below. Hence, the final model used to fit the background in the mass window was identified to be a O(1) Chebychev polynomial with a linear dependence.
Figure 7.18: Exponential slope dependence across BDT bins evaluated on MC. The x-axis values are the mean value of the BDT variable in that given bin and the error bars represent the RMS of the distribution in the x-axis and the error returned by the fitter on the y-axis [155].

across the BDT bins and an exponential with a constant dependence. The fit result of those models in the 4 BDT bins data side-bands can be seen in fig. 7.19. The dependence found from the MC studies was checked by performing the fits on the data side-bands independently, followed by a simultaneous fit. The result of those studies can be seen in fig. 7.20, which proves that the MC dependence is compatible with the observed dependence, when the models are fitted independently on the data side-bands. In addition, it can be seen as validation that the simultaneous fit algorithm assumes the functional form of the dependence correctly providing the expected linear and constant functions for the Chebychev and exponential slopes respectively. The compatibility between the result on data and MC was studied as well leading to a good agreement for the Chebychev slope; however, a discrepancy was observed between the exponential slope observed in data and MC. The discrepancy was interpreted as residual differences between data and MC samples and therefore assessed not to be a factor affecting the final result when the data are unblinded. In any case, a study of the discrepancy was performed as part of the systematics assessments.

The last components to complete the mass model for the signal yield extraction are the peaking background and the signal model itself. Peaking background is the $B \rightarrow hh'$ decays for which both hadrons are mis-identified as muons. The available statistics in the MC sample is limited to identify the shape of this background component with the nominal muon requirements and the fake rate. Thus, the ID based calculation of the vertex mass was used without requiring any fake-muons in the final state. Choosing the ID based mass required a correction to be applied in order to represent the mass distribution faithfully.

The corrections needed for $B \rightarrow hh'$ arise from the fact that the ID plus MS mass distribution, used to extract the signal yield has an offset with respect to the ID only mass distribution and a different resolution. Additionally, the di-muon final state affects the mass distribution with effects that arise from the trigger selection and due to the extra preselections applied to the samples discussed in app. A. Such an effect could be the different muon working points (con-
7.3 ATLAS result

Figure 7.19: Simultaneous fit projections performed on the 4 BDT bins with the combined background model on data side-bands. Red shows the combinatorial background model and green indicates the Same-Side Same Vertex (SSSV) model, both derived from the MC fits [40].

Figure 7.20: Exponential and O(1) Chebychev slopes (right/left) as function of the average BDT values in each bin. The red points show the independent fit result on data side-bands, the black points the equivalent fit on MC and finally the green and blue the simultaneous fits on data and MC respectively [155].
ditions met to associate detector object as a muon) which would affect the number of events and hence affect the mass model derived in MC. All effects were carefully studied and thereafter correction factors were derived and applied on the invariant mass distribution. The PDF model used for the fit was a double Gaussian (equivalent to the $B_{s}$ signal). No BDT dependence was found for any of the parameters, as in the case of the background model shape parameters. The list of mean and RMS of the Gaussians as well as the relative fraction of them that create the final signal shape are listed below.

- $\mu_1$: 5239.5 $\pm$ 1.3 MeV
- $\sigma_1$: 85.90 $\pm$ 0.2 MeV
- $\mu_2$: 5216 $\pm$ 10 MeV
- $\sigma_2$: 188.7 $\pm$ 3.2 MeV
- fraction between Gaussians: 0.85 $\pm$ 0.02

The last models discussed concern the PDF functions for the two signal distributions. For these, two superpositioned Gaussian PDFs were chosen to describe the core of the distribution as well as the radiative tails appearing on the left of the distribution. The shape parameters were extracted from MC studies; however, the signal MC sample for both signal signatures has more events than the expected signal in data. For this reason, attempting to model the full distribution will produce a very complicated PDF model, since it will have to take into account shape effects appearing only in the regime of high statistics. Hence, a toy-MC study was performed where 1000 samples were bootstrapped [151] from the full sample according to the expected number of events in data. The resulting samples in all the 4 BDT bins were fitted with a simultaneous extended unbinned maximum likelihood fit and the average value of the shape parameter was calculated. The toy study was performed on both the combined ID and MS mass and the ID only, indicating that the first one has a better resolution in compare to the second one, as expected. This was also the decisive cross check to define which of the two available invariant mass calculations would be used for the analysis. Fitting the signal also indicated a shift between the Particle Data Group (PDG) foreseen mass value and the value measured in the toy study. The discrepancy was found to be of the order of 9 MeV, which was not found to be present in the ID only mass. The offset source is most likely due to calibration effects on the combined muon tracking [155]. To cross-check the bias the control sample was used, indicating a 5MeV discrepancy between the PDG and the measured value. Observing a similar discrepancy concluded that the source is due to the combined muon requirements and are present in both data and MC samples. Therefore, the means and sigmas of the Gaussians were fixed to the MC values and any shifts were studied as systematic uncertainties separately.

The resulting values for both signal models can be seen in table 7.2. The resulting fits on the bootstrap extracted data set can be seen in fig. 7.21. The combined model with all the optimizations performed was identified to contain 7 sources of systematic uncertainty, listed below.
### 7.3 ATLAS result

<table>
<thead>
<tr>
<th></th>
<th>$B^0 \rightarrow \mu^+ \mu^-$</th>
<th>$B^0_d \rightarrow \mu^+ \mu^-$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_1$</td>
<td>$5357.7 \pm 1.5$ MeV</td>
<td>$5270.6 \pm 1.5$ MeV</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>$83.0 \pm 1.9$ MeV</td>
<td>$79.2 \pm 2.2$ MeV</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>$5257.0 \pm 18.3$ MeV</td>
<td>$5197.1 \pm 14.5$ MeV</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>$193.8 \pm 10.5$ MeV</td>
<td>$172.9 \pm 9.6$ MeV</td>
</tr>
<tr>
<td>fraction</td>
<td>$0.88 \pm 0.02$</td>
<td>$0.83 \pm 0.03$</td>
</tr>
</tbody>
</table>

Table 7.2: Signal model shape parameters extracted from toy-MC study.

![Di-muon invariant mass fit for $B_s$ signal extracted from signal MC assuming the SM prediction on the branching fraction.]

Figure 7.21: Di-muon invariant mass fit for $B_s$ signal extracted from signal MC assuming the SM prediction on the branching fraction. With the red dots the full $B_s$ signal MC sample scaled to the expected number of events in data were noted, and with the magenta triangles the equivalent $B_d$ signal MC sample. [155].

- **Mass scale ±5 MeV**: When the $Y$ mass was measured a momentum scale uncertainty was observed of the order of ±5% on the di-muon mass [161]. The same effect in the $B^0_s$ case is of the order of 2.7 MeV and equivalently for the $J/\psi$ mass it’s 2 MeV. Thus, the systematic effect was studied by varying the mass peak position by ±5 MeV to provide a conservative estimate.

- **Mass resolution ±5%**: The resolution of the invariant mass signal was allowed to vary by 5%. This systematic uncertainty was estimated from studies on the $J/\psi$ and $Z$ decays into di-muons [157].

- **Chebychev BDT dependence**: Estimates the effect of the assumption that the slope of the Chebychev is linear across the BDT bins.

- **Exponential slope BDT dependence**: Estimates the effect of the assumption that the exponential slope is constant across the BDT bins.

- **Background analytical PDF models**: Changing the identified models for the background sources to check the impact on the final result. Nominally all possible PDF could be tested, however in practice only the most likely alternatives are used.
• **Inclusion of semi-leptonic background model**: Including the PDF in the fit model to check impact on observed signal events. The background events are included only in the toy generation and not in the fit model.

• **Inclusion of \(B_s\) background**: Same as above. The events are included in the toy generation but not in the fitting of the invariant mass distribution.

All the systematic sources were tested for their impact on the signal yield. Hence, the observed dependence of the signal yield due to systematic variations was included in the final fit as a Gaussian smearing on the signal yield parameters for both the \(B_s^0\) and the \(B_d^0\).

**Result extraction**

Setting the mass fit procedure for both the signal and reference channel, allowed for the direct extraction of the branching fraction for both the \(B_s^0\) and the \(B_d^0\), from the mass fit itself. To perform the extraction of the measured value the following modification seen in eq. 7.17 was performed based on eq. 7.16.

\[
N_{B_s^0}^{obs} \rightarrow \left( \frac{\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) \times N_{J/\psi K^+}}{f_{ud} \times \mathcal{B}(B^+ \rightarrow J/\psi K^+) \times \mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-) \times R_{AC}} \right)
\]  

(7.17)

The fit includes all the systematic uncertainties obtained from tuning the mass fit, the studies on the efficiency function and the fragmentation ratio, hence the final result provided both the statistical and systematic uncertainty. Additional smearing was applied to all the parameters in the mass fit except the branching fractions which are the parameters of interest. The values of all the different parameters in the fit were taken as discussed above.

Unblinding the data set and applying the fit procedure with all the systematics yielding the following fit projections seen in fig. 7.22. According to the SM prediction the expected number of signal events was \(N_{B_s^0}^{exp} = 91\) and \(N_{B_d^0}^{exp} = 10\). Running the unbinned simultaneous likelihood fit yielded to the observed values seen in eq. 7.18.

\[
N_{B_s^0}^{obs} = 80 \pm 22 \\
N_{B_d^0}^{obs} = -12 \pm 20
\]  

(7.18)

which is fully compatible with the SM expectation. In addition, a correlation of \(\rho = 0.82\)\% is observed mostly arising from the limited resolution of the ATLAS detector on muons. The uncertainties quoted represent the total systematic plus statistical uncertainty obtained from a Gaussian approximation around the minimum of the likelihood fit [40].

The non-physical result on the \(B_d^0\) number of observed events did not prove to be an issue for the analysis validity. Since it’s an under-fluctuation, the Neyman construction (which depends on toy-MC samples) can provide an upper limit on the \(\mathcal{B}(B_d \rightarrow \mu^+ \mu^-)\) measurement.

With the fitted values obtained from the invariant mass fit, eq. 7.17 yields the following values of the branching fraction for both signal decays:
To determine the uncertainty on the measured parameters three different approaches were used. The low available statistics for the analysis imposes a reliable method for the uncertainty calculation. The three approaches that were explored will verify each other and produce the best possible estimation of the branching fraction uncertainty.

The first approach is the so called RCF variance [151] which extracts the uncertainty by approximating the likelihood around the minimum to be Gaussian. The second approach approximated the confidence intervals by using the profile likelihood ratio tool [151]. The last and more accurate approach is the Neyman construction which was used as the main tool for the final result with the first two options used for validation. The list of the combined statistical and systematic uncertainties obtained for both measurements along with their central value can be seen in table 7.3.

Given that the RCF and the profile likelihood assume a Gaussian like distribution of the likelihood the implementation of the Neyman belt is expected to provide a more accurate result for the uncertainty, since no such assumption is needed. Such assumptions in the statistical regime of the analysis are not always fulfilled and can bias the estimation of the uncertainty. In addition, the Neyman construction provides a natural way of extracting the confidence intervals rather than having to use more complicated statistical tools. The resulting likelihood

\[ \mathcal{B}(B^0_s \rightarrow \mu^+ \mu^-) = 3.21 \times 10^{-9} \]

\[ \mathcal{B}(B^0_d \rightarrow \mu^+ \mu^-) = -1.3 \times 10^{-10} \]  \hspace{1cm} (7.19)
7.3 ATLAS result

<table>
<thead>
<tr>
<th>Method</th>
<th>$\mathcal{B}(B^0_s \rightarrow \mu^+ \mu^-)$</th>
<th>$\mathcal{B}(B^0_d \rightarrow \mu^+ \mu^-)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCF</td>
<td>$(3.21 \pm 0.93) \times 10^{-9}$</td>
<td>$(-1.3 \pm 2.1) \times 10^{-10}$</td>
</tr>
<tr>
<td>Profile LL Ratio</td>
<td>$3.21^{+1.02}_{-0.99} \times 10^{-9}$</td>
<td>$-1.3^{+2.3}_{-2.1} \times 10^{-10}$</td>
</tr>
<tr>
<td>Neyman Belt</td>
<td>$3.2^{+1.1}_{-1.0} \times 10^{-9}$</td>
<td>$&lt; 4.3 \times 10^{-10}$</td>
</tr>
</tbody>
</table>

Table 7.3: The ATLAS 15/16 Branching Fraction results with the uncertainties evaluated with three different statistical methods. Only the last row (Neyman Belt) is considered the published quantity with the other two being a cross check [155].

contour from the Neyman belt can be seen in fig. 7.23, which also shows the compatibility of the measurement with the SM expectation. According to the chronology of the available data

![Figure 7.23: Neyman belt contours on the branching fraction plane of the $B^0_s$ and $B^0_d$. The red lines indicate the contours obtained with the statistical uncertainty only whereas the blue lines include both statistical and systematic uncertainty. The SM predicted value can be seen with the black dot. [40].](image)

sets the ATLAS result on the $B^0_{(s)} \rightarrow \mu^+ \mu^-$ branching fraction was published [40] prior to the other two experiment (CMS, LHCb) making the value the most stringent measurement at the time (CMS: [144], LHCb: [146]). Unfortunately, only a 4.7σ significance was reached for the $\mathcal{B}(B^0)$ and no evidence can be claimed. However, the result shows significant improvement on the evolution of the branching fraction measurement for the $B^0_s$ and $B^0_d$ di-muons decays which can be seen in fig. 7.24.

7.3.1 Preparation for full Run-2

At the date of writing, the ATLAS collaboration is preparing for the branching fraction analysis for the $B_{(s)} \rightarrow \mu^+ \mu^-$ with the integration of the full Run-2 data set. While a substantial effort was placed on optimizing the BDT discriminator to gain as much as possible sensitivity to the signal, a significant effort was applied to identify the best trigger strategy and exploiting the
The trigger strategy is an important aspect of the whole analysis, since the B-physics trigger menu is largely pre-scaled in ATLAS and the main limitation of the analysis arises from the low available statistics. Identifying the lowest possible prescaled triggers and maintaining the analysis complexity within nominal reach is an important task for the full Run-2 version.

Additionally, due to the di-muon invariant mass resolution of ATLAS, the $B^0_s$ and $B^0_d$ peaks seen in fig. 7.21 heavily overlap, making them practically indistinguishable, in comparison to the CMS (slightly distinguishing) and LHCb resolutions. Identifying a method for separating the two mesons is expected to provide a significant impact on the measurement of the $B_d$ meson, which as shown above in the 15/16 version of the analysis underfluctuated to a non-physical number of observed events. Furthermore, developing the procedure provides a deeper understanding of the invariant mass models, and their behavior when they are applied on the analysis mass window.

In the following sections of this chapter, a more detailed discussion will be provided on the choices of the triggers for the full Run-2 analysis in addition to the preliminary studies of the peak separation between $B^0_s$ and $B^0_d$.

**Trigger studies**

The partial Run-2 analysis as discussed above (sec. 7.3), tried to simplify the trigger strategy used with a minimal loss in the number of expected $B^{0}_{(s)}$ mesons. For this reason, the decision was made to use only the listed HLT triggers below with the $p_T$ muon cut at 4 and 6 GeV, which are also seeded from the same L1 trigger (L1_MU6_2MU4) and exclude the equivalent trigger with the 6 GeV $p_T$ cut on both muons (HLT_2mu6_bBmumu), since the gain was minimal in
7.3 ATLAS result

comparison to the added complexity.

- HLT_mu4_mu6_bBmumu_Lxy0 (2015 and 2016).

For the full Run-2, unfortunately the $L_{xy} > 0$ version of the trigger was prescaled providing a significant loss of collected luminosity for the B-Physics analysis over the remaining 2017 and 2018 periods. In addition, no other trigger during the already analyzed partial Run-2 data seems to provide a significant improvement. For these reasons, the main triggers contributing the lion’s share in 2017 and 2018 are trigger chains with a $p_T$ cut larger than 4 GeV and 6 GeV, that can be seen in table 7.4. As a topo cut the various L1 topological cuts are referred to (sec. 2.2.6).

<table>
<thead>
<tr>
<th>$p_T &gt; 4 GeV$</th>
<th>no cut</th>
<th>$L_{xy} &gt; 0$ cut</th>
<th>topo cut</th>
<th>topo and $L_{xy}$ cuts</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_T, p_T &gt; 4 GeV$</td>
<td>3.17</td>
<td>0.00</td>
<td>1.03</td>
<td>20.68</td>
</tr>
<tr>
<td>$p_T &gt; 6 GeV, p_T &gt; 4 GeV$</td>
<td>24.29</td>
<td>26.03</td>
<td>4.00</td>
<td>80.00</td>
</tr>
<tr>
<td>$p_T &gt; 6 GeV$</td>
<td>45.53</td>
<td>43.07</td>
<td>21.62</td>
<td>111.81</td>
</tr>
</tbody>
</table>

Table 7.4: Integrated luminosity for trigger with $p_T > 4 GeV$ and various selection cuts (e.g. topological, $L_{xy}$) for 2017 and 2018 data taking periods. The luminosity units correspond to $fb^{-1}$ for each trigger. The values indicated are summations of the luminosities for the various data periods for the different trigger objects, hence include pre-scaling.

As previously mentioned, relying only on the partial Run-2 strategy would yield to a massive reduction on the number of collected events, severely impacting the analysis sensitivity.

To calculate the number of expected events, the efficiency with respect to the lowest $p_T$ trigger was used and can be seen in fig. 7.25. Since the lowest $p_T$ threshold for both muons is 4 GeV, all the other triggers are sub-sets of the $2\mu 4$ trigger. Therefore, the collected luminosity for each trigger needs to be re-weighted with the efficiency factor in order to avoid double counting

Figure 7.25: 17/18 trigger efficiency with respect to $p_T > 4 GeV$ and $p_T > 4 GeV$ muon trigger. The x-axis error bars do not contain any information since the values on x are categorical. For the y-axis the calculation of the uncertainty of correlated numbers (Clopper-Pearson [163]) was used since the numerator and denominator number of events with the different triggers are based on the same MC sample.
events. The expected yields after the correction can be seen in table 7.5.
From the last table the baseline strategy for the full Run-2 analysis was identified, which is

<table>
<thead>
<tr>
<th>no cut</th>
<th>Lxy &gt; 0 cut</th>
<th>topo cut</th>
<th>topo and Lxy cuts</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_T &gt; 4 GeV$</td>
<td>11.59</td>
<td>0.00</td>
<td>3.43</td>
</tr>
<tr>
<td>$p_T &gt; 6 GeV$</td>
<td>73.19</td>
<td>78.16</td>
<td>10.87</td>
</tr>
<tr>
<td>$p_T &gt; 6 GeV$</td>
<td>57.00</td>
<td>56.21</td>
<td>24.41</td>
</tr>
</tbody>
</table>

Table 7.5: Number of signal events after efficiency correction with respect to 2mu4. Each cell contains the different event count for different trigger objects, equivalently done as in table 7.4 for the luminosities.

planning to use the following list of triggers,

- HLT_mu4_mu6_bBmumu_Lxy0 (2015+2016).
- HLT_mu4_mu6_bBmumu_Lxy0_L1BPH-2M9-MU6MU4_BPH-0DR15-MU6MU4 (17/18).
- HLT_2mu4_Lxy0_L1BPH-2M8-2MU4 (2018).

exploiting 94% of the available collected luminosity (good-for-physics) for the full Run-2 data collection period. The various letters in the names indicate the different cuts contained in each trigger object. For example the BPH hints that this trigger is dedicated to B-physics and 0DR15 that the DR distance of the two muons is between 0 and 15.

The baseline strategy for the triggers foresees the addition of 3 topological triggers. Further studies were performed with the use of the $B_s^0$ signal MC used in the 15/16 analysis to cross check whether the inclusion of additional triggers during the 15/16 periods would increase significantly the efficiency with respect to the baseline strategy outlined above. Such studies aim to increase the statistics for the signal and ensure that no $B_{(s)}$ events were excluded from trigger selections, which are the main limitation factors of the analysis.

The study as mentioned above, was based on the signal MC available from the 15/16 version of the analysis. In addition, the information provided by the ATLAS trigger group on the collected luminosity per data collection period for each trigger chain was used. In this way effective prescales were calculated for each trigger with respect to the lowest available unprescaled trigger. Additionally, extra caution was taken to identify the L1 trigger for each trigger chain. The main concern arises from the fact that overlapping L1 triggers (e.g. triggers with $p_T > 6 GeV$ are a subset of $p_T > 4 GeV$) might contain the same events due to their selection. For this reason, only the events passing one of the L1 trigger selections are used for each trigger chain to avoid double counting. The triggers exploited with their additional signal efficiency on the number of expected events can be seen in table 7.6, with the naming convention following the one showed in fig. 7.25. The percentage value quoted in the right most column corresponds to additional percentage with respect to the value calculated for the baseline strategy (shown in sec. 7.3).
Table 7.6: Additional triggers studied for the 2015/2016 period to be included in the full Run-2 analysis.

<table>
<thead>
<tr>
<th>Trigger name</th>
<th>Year</th>
<th>Additional efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>2mu4</td>
<td>2015</td>
<td>0.16%</td>
</tr>
<tr>
<td>2mu6_Lxy0_topo</td>
<td>2016</td>
<td>1.72%</td>
</tr>
<tr>
<td>2mu6_Lxy0</td>
<td>2016</td>
<td>4.67%</td>
</tr>
<tr>
<td>mu4_mu6_topo (2M8)</td>
<td>2016</td>
<td>1.85%</td>
</tr>
</tbody>
</table>

The 2M8 parenthesis shows that the topological item used for this specific trigger is different than the items used in the baseline strategy and that a different mass cut was applied.

As expected the extra trigger chains provide a minimal addition to the collectable signal yield provided by the baseline strategy and are thus not included at first instance in the procedure, since that would inflate the complexity of the analysis. The chain with the largest impact is the third one which will be included in all the analysis ntuples and employed in case further sensitivity is needed (indicated by the fit procedure optimization step).

The final step for the study was to calculate the number of expected events with the baseline strategy of triggers for both the $B_s^0$ and the $B_d^0$, with all the analysis pre-selections applied and the minimal BDT cut. To evaluate the expected yield the following formula was used seen in eq. 7.20.

$$N_{\text{exp}} = N_{\text{Run}} \frac{\sigma_{\text{Full Run2}}}{\sigma_{\text{Run1}}} \times \frac{\epsilon_{\text{Full Run2}}}{\epsilon_{\text{Run1}}} \times \frac{L_{\text{Full Run2}}}{L_{\text{Run1}}}$$

(7.20)

Where the changes due to the increased centre-of-mass energy to the production cross section for the B-mesons were included, the different trigger efficiencies calculated above and the luminosity. Extrapolating the number of expected events leads to the values seen in eq. 7.21,

$$N_{\text{exp}} \approx 400 \quad \text{and} \quad N_{\text{exp}} \approx 44$$

(7.21)

where the cross section ratio 1.7 was used, $L_{\text{Run1}} = 25 fb^{-1}$, $L_{\text{Full Run2}} = 139 fb^{-1}$ and finally for the efficiency of Run1 $\epsilon_{\text{Run1}} = 1$ and the corresponding for the full Run-2 $\epsilon_{\text{Full Run2}} = 0.94$.

Current status

There are many tasks currently developing in parallel for the full Run-2 version of the analysis. One of the most important ones is the inclusion of the mass resolution into the invariant mass fit.

Performing the mass separation study provides multiple benefits for the analysis and a potential decorrelation of the $B_d^0$ and $B_s^0$ peaks. Introducing the mass resolution into the likelihood fit can happen via various routes. The most straightforward one is the approach taken by CMS where the resolution was explicitly included in the likelihood increasing the number of observables by one. In order to minimize complication on the side of the likelihood, the approach taken in ATLAS was to identify a strongly correlated kinematic variable and bin the data-set, not only in BDT but also in that variable. This approach is expected to be less complicated, compared to an unbinned fit in both the kinematic variable and the invariant mass, since no analytical modeling is required for the kinematic variable. Testing the available options in the
15/16 ntuple, the usage of the maximum $\eta$ from the two muons was identified to have a strong correlation with the resolution, as it can be seen in fig. 7.26. From now on the variable will be called $\eta_{\mu^+ \mu^-}^{\text{max}}$ for simplicity.

To exploit the benefit of binning in $\eta$ as well as to the BDT a toy generation model is needed to be set in place that accurately reproduces the data. Identifying the exact number of $\eta$ bins that yield the maximum separation of the two signal peaks requires that the MC generation model for the signal needs to include a dependence of $\eta$. To avoid parametrizing the dependence analytically the same approach as for the background models in the BDT was taken. That means that the model parameters were constrained across different $\eta$ bins with linear slopes. The most straightforward solution would be to include the analytical PDF, however that would require some accurate parameterization of the $\eta$ distribution. Avoiding the complicated full analytical model can be achieved by introducing dependencies of the invariant mass PDFs shape parameters as a function of the $\eta$ variable. The methodology is equivalent to the approach used for the BDT bins in the 15/16 model, where the slope of the background components were constrained with linear (constant) models across the various analysis BDT bins.

To acquire the correct models for the generation a framework was developed that iteratively fits the available background and signal MC samples from the 15/16 analysis to avoid any assumptions on the fit starting values. All the model parameters are left free and determined by the result of the last iteration step. The steps taken to identify the model shapes can be seen in the list below.

- **Signal:**
  1. Binned single Gaussian fit with all parameters free.
  2. Single Gaussian fit with all parameters free starting from previous step fit result.
  3. Double Gaussian fit with all parameters free starting from previous step fit result.
4. Triple Gaussian fit with common mean of the core two Gaussian. Parameters are all free and starting from previous step fit result.

5. Triple Gaussian fit with independent means and all parameters free, starting from previous step fit result.

• **Background:**

1. Binned fit on lower data Side-Bands (SB) with exponential only, all parameters free.

2. Binned fit on upper data SB with Chebychev O(1) only, all parameters free.

3. Unbinned fit on lower data SB with exponential only, all parameters free starting from binned fit result.

4. Unbinned fit on upper data SB with Chebychev O(1) only, all parameters free starting from binned fit result.

5. Unbinned fit on blinded data with combined models, all parameters free starting from two previous unbinned fit result.

A projection of the invariant mass fit result on the signal MC can be seen in fig. 7.27, showing all the different components. Studying the evolution of the parameters after the individual fits in the four BDT bins and a number of different $\eta$ binning schemes, allowed the determination of the dependencies for all the shape parameters across the bins. In this way a simultaneous fit model was constructed to generate toy samples as faithfully as possible in both signal and background.

Studying the dependence of the mass resolution thoroughly is a work in progress for the full Run-2 analysis. However, currently both the generation model and the fit model reached the level that allows testing different generation binning schemes in $\eta$ and fitting them to check their performance.

![Figure 7.27: Result of three Gaussian fit on signal MC for a specific $\eta$ bin.](image)
7.4 Conclusions on $B_{(s)} \rightarrow \mu^+ \mu^-$ analyses

All the three experiments at the LHC attempting to measure the branching fraction of the $B^0 \rightarrow \mu^+ \mu^-$ have provided a result with the 15/16 data set. In all three cases certain limitations appeared mainly due to the low statistics available for such a rare process. Hence, the integration of the full Run-2 luminosity is essential to fully explore the potential of this analysis, as indicated already by the preliminary LHCb result.

As a first step, once all the 15/16 analysis results were available, a combination was attempted in order to fully stretch the statistical reach of the luminosity provided by the LHC [164]. The combination result on the branching fraction can be seen in eq. 7.22.

\[
\mathcal{B}(B^0_s \rightarrow \mu^+ \mu^-) = (2.69^{+0.37}_{-0.35}) \times 10^{-9}
\]

\[
\mathcal{B}(B^0_d \rightarrow \mu^+ \mu^-) = (0.6 \pm 0.7) \times 10^{-10}
\]

(7.22)

Up to today, this is the best available limit for the values of the branching fraction for the $B^0 \rightarrow \mu^+ \mu^-$ decays. CMS and LHCb provided in addition to the branching fraction measurement the $B^0_s$ effective lifetime result, with their 15/16 data sets. Moreover, the combination of the two results was explored yielding to the value seen in eq. 7.23 [164].

\[
\tau_{B^0_s \rightarrow \mu^+ \mu^-} = 1.91^{+0.37}_{-0.35} \text{ ps}
\]

(7.23)

Due to the importance of the effective lifetime result (sensitive to NP), having the equivalent measurement from ATLAS is needed. The main physics work of this thesis focused on the analysis of the $B^0 \rightarrow \mu^+ \mu^-$ effective lifetime with the well known 15/16 data set, based on the existing work from the branching fraction version of the analysis. In the following chapters, the analysis procedure along with the studies for the various systematic sources will be discussed.
In the preceding chapters all the necessary ingredients were provided in order to introduce the measurement on the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime. In this chapter the strategy along with the tools developed to measure the lifetime value with the ATLAS detector on the partial Run-2 data will be described. The main body will focus on the developed tools and optimizations performed on pseudoexperiments generated to mimic the real data, followed by the application of the procedure on the analysis control channel $B^+ \rightarrow J/\psi(\rightarrow \mu^+\mu^-)K^+$ used to estimate the several systematic sources. Finally, the expected uncertainty returned by the fitter will be shown, in preparation for the fit to the data set.

8.1 Analysis strategy

To measure the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime the distribution of the $B_s^0 \rightarrow \mu^+\mu^-$ decay time needs to be extracted from the data sample that passes the selections used in the 2015-2016 branching fraction analysis and are described in app. A. Those selections do not provide a separation between the signal decay time and the decay time of the background, hence the signal lifetime cannot be measured only based on those. To extract the lifetime two possible choices are available for the analysis. The first choice is to define signal and background PDF models and fit the combined decay time distribution, which is effectively the choice CMS made in their main approach (sec. 7.1.2). The second possible approach is to subtract the background decay time distribution from the data and obtain the signal only distribution, which is more aligned to the second approach of CMS (sec. 7.1.2) and the main method of LHCb (sec. 7.2.2).

For ATLAS the fit on the background subtracted signal distribution was decided as explored as the main procedure. To obtain the background subtracted distribution the use of the sPlot method [150] was deployed, for which a detailed description is provided in app. B. The sPlot performs the subtraction of the background in a two step procedure, with the only important requirement to be that the variables used are uncorrelated. In the case of the effective lifetime measurement the two variables were identified to be the invariant mass and the decay time, where the first one is the discriminating variable and the latter one control variable. Their correlations for all the different event sources were measured in MC and can be seen in fig. 8.1.
The first step for *sPlot* is to perform an unbinned maximum likelihood fit on the discriminating variable, where all the individual event sources are modeled with analytical PDFs. The fit determines the various yields for each model and uses them to derive weights for each event, called *sWeights*. Each event gets assigned a number of *sWeights* equal to the number of event sources, with the highest weight value on one of the categories indicating that the event is more likely to belong in this category. The next step for *sPlot* is to apply the specific *sWeights* on the events for the control variable, to create the distribution of that variable for the given category in a form of a weighted histogram. The histogram obtained with the application of the signal *sWeights* represents the background subtracted signal distribution of the control variable (e.g. decay time). If all the weighted histograms, for each category of *sWeights* are summed, the original full distribution is obtained, effectively allowing *sPlot* to be a statistical tool that performs background subtraction without any prior knowledge of the control variable distribution.

The accuracy of the subtraction depends significantly on the extraction of the *sWeights*, derived from the fit on the discriminating variable. In the case of ATLAS the developed models from the branching fraction analysis were employed (sec. 7.3). The final background subtracted distribution for the $B_0$ decay time was chosen to be a binned, rather than an unbinned fit. Since binning the available data reduces the statistical strength, it’s crucial to maximize the available signal events. For this reason a single BDT bin was decided to be used in contrary to the 4 BDT bins used in the branching fraction analysis, to avoid reducing the statistical power of the data sample.

Nominally the decay distribution of a particle is described with an exponential function [3]. Introducing detector and selection effects smears the original distribution. Such effects are the BDT discriminator, the triggers, and the analysis pre-selection, which use lifetime dependent cuts in order to reduce the prompt background in the analysis window. At low decay times for example those effects can easily be seen with a turn-on effect appearing. An accurate analytical modelization is required if a full PDF is needed to be derived. To avoid parameterizing the signal analysis efficiency curve with an analytical model, a fitting approach with the use of lifetime dependent MC histogram templates was used. Acceptance and efficiency effects will be reproduced by the simulation in these templates. The signal lifetime templates from MC are parameterized as a function of the signal lifetime; hence a fit procedure needs to be defined to extract the lifetime and its uncertainty as a function of this parameter. The derived fitting procedure, discussed in more detail in sec. 8.3, was characterized through pseudoexperiments.

![Figure 8.1: Correlation graphs for three source components from full MC samples. The linear correlation factor was calculated for each source between invariant mass and decay time.](image)
The analysis uncertainty is expected to be dominated by its statistical contribution due to the low number of events. For this reason the determination of systematic uncertainties - although thorough - can in many cases afford to take a rather conservative approach. The main systematic sources expected are arising from the fit procedure and the data-MC inconsistencies for the signal. The former can be measured with toy experiments and simulation, while the latter is measured through a data-driven approach relying on the same control channel used for the branching ratio analysis ($B^+ \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^+$). The control channel presents a similar final state, with the same selections and triggers as the signal channel. The fitting procedure defined in $B^0_s \rightarrow \mu^+ \mu^-$ was applied without any changes on the control channel to measure the well known lifetime of the $B^+$ meson. Any discrepancy observed from the accurately known experimental value [41] was included as a systematic uncertainty in the final measurement.

The next sections will first describe the data-set selection and the selections applied. The discussion will then move on to the definition of the invariant mass models used in the sPlot subtraction, followed by the details of the signal fitting procedure the analysis will employ. The chapter will conclude with the prediction of the fit’s statistical performance on the signal, and the results of its application on the control channel.

### 8.2 Samples and triggers

This analysis is performed as a follow up on the partial Run-2 branching fraction analysis. For this reason the data samples used are the $\sqrt{s} = 13$ TeV data taken by the LHC during 2015 and 2016. The integrated luminosity collected during those periods corresponds to 36.2 fb$^{-1}$. However, as shown in the previous ch. 7 not all of the luminosity delivered by the accelerator is detected by ATLAS, and in addition not all of the recorded luminosity is considered good for physics.

On the MC side the branching fraction samples produced at $\sqrt{s} = 13$ TeV describing both the signal and the background in the mass window range of the analysis, for both the analysis channel ($B^0_s \rightarrow \mu^+ \mu^-$) and the control channel ($B^+ \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^+$), were used. The MC samples proved sufficient for the effective lifetime measurement thereafter no further generations were needed. The available samples can be seen in the following list, with a more detailed discussion about the selections applied on them found in app. A.

- $B^0_s \rightarrow \mu^+ \mu^-$ signal MC: Contains in total 166 thousand $B^0_s \rightarrow \mu^+ \mu^-$ signal decays, after the preselection applied at the level of ntuple generation.
- $b\bar{b} \rightarrow \mu\mu X$ background MC: Combined background sample containing both the continuum background and the SSSV partially reconstructed background. Contains almost 6 million background decays after the preselection applied at the ntupling level.
- $B^+ \rightarrow J/\psi K^+$ signal MC: Contains in total 200 thousand $B^+$ signal decays, after the preselection applied at the level of ntuple generation.
8.2 Samples and triggers

- $b\bar{b} \rightarrow J/\psi X$ **background MC**: Contains both background sources for the control channel, as well as a leaked signal distribution which has to be veto-ed to obtain the correct background distribution. The number of events included are roughly 4 million with all the cuts from the preselection at the level of ntuple generation.

The trigger selection applied is the same as for the branching fraction analysis (sec. 7.3) and can be summarized in the following list:

- **Analysis channel**
  - `HLT_mu6_mu4_bBmumu`: Signal channel object for the 2015 period.
  - `HLT_mu6_mu4_bBmumu_Lxy0`: Signal channel object used in both 2015 and 2016 periods.

- **Control channel**
  - `HLT_mu6_mu4_bJpsimumu`: Control channel trigger for 2015.
  - `HLT_mu6_mu4_bJpsimumu_Lxy0 (main)`: Control channel trigger for 2015 and up to a certain point of 2016.
  - `HLT_mu6_mu4_bJpsimumu_Lxy0 (delayed)`: Control channel trigger, equivalent of previous object in another data stream, used mostly during 2016.

As it can be seen the trigger selections between the two channels are almost identical (different inv. mass cuts - $4.9 < m_{\mu\mu} < 5.6$ GeV ($B_s^0$), $2.5 < m_{\mu\mu} < 4.3$ GeV ($J/\psi$)) and thereafter dominated by the same effects. The main trigger selection criteria affecting the decay time distribution for both channels originates from the $L_{xy} > 0$ cut at trigger level imposed in the chains used during 2016.

Having common samples and trigger strategy with the branching fraction analysis allows for a better understanding of the used samples and their statistical properties. A difference in the sample use appears in the application of the BDT, to suppress the background in the mass window. As discussed in sec. 7.3 the branching fraction analysis defines 4 BDT bins with certain signal efficiency in order to perform an accurate measurement on the number of signal events. In the lifetime analysis though partitioning the sample into multiple bins in BDT would result into a complication for the background subtraction with the invariant mass fit. Therefore a single BDT bin was used in the most signal enriched region of the BDT variable. To identify the best possible BDT region an optimization was performed based on the maximization of the figure of merit $S/\sqrt{S+B}$ (S: Expected signal events, B: Expected background events), which will be discussed in sec. 8.2.2 below.

8.2.1 Signal estimation

Before tuning and fully optimizing the analysis, the expected performance on the sample was very roughly estimated to understand the relevance of the result with respect to existing publications. Without any optimization the highest BDT bin ([0.415,1.0]) from the branching fraction
8.2 Samples and triggers

analysis was taken, since it’s the cleanest in terms of background entering the mass window. Extrapolating from the mass sidebands allows the calculation of the number of background events entering the signal region. In this way a rough background subtraction can be performed seen in fig. 8.2, on the decay time distribution.

The background subtracted decay time distribution can be seen with the red line in comparison to the full distribution seen with the blue. To extract the expected background in the signal region an extrapolation taken from the data side-bands in mass was taken. The number of estimated signal and background events can be seen in eq. 8.1.

\[
N_{\text{estim}}^{\text{sig}} \approx 34 \\
N_{\text{estim}}^{\text{bkg}} \approx 8
\] (8.1)

With this rough estimation, it can be seen that the data set collected by ATLAS can provide a similar number of signal events to the two other LHC experiments. Performing a more elaborate background subtraction and identifying a more optimal BDT cut is expected to increase the analysis number of signal events. Therefore the following study shown in the next section was developed to identify the cut with the performance according to the traditional figure of merit nominally used in such optimizations (e.g. \( S/\sqrt{S+B} \)).

### 8.2.2 BDT cut optimization

To identify the optimal BDT cut a detailed study with the background and signal MC sample was performed. The study optimized the figure of merit seen in eq. 8.2 [151] in the signal region ([5166, 5526] MeV) of the invariant mass distribution:

\[
\text{Merit} = \frac{S}{\sqrt{S+B}},
\] (8.2)

where S are the number of signal events in the mass window and B the number of background events. The MC samples used are generated with an arbitrary number of events, therefore to
identify the number of signal events the SM expectation value on the branching fraction was used. For the background the normalization was taken from the number of observed background events only in the side-bands of the invariant mass window when the lowest BDT cut (0.146) taken from the branching fraction analysis was applied.

The BDT variable has as allowed values [-1, 1], where -1 are the events which are likely to be background and +1 the more signal-like events. Since the events with a value close to +1 are signal like, the upper boundary of the expected range was set to the maximum possible value. Thus only the lower boundary of the BDT bin was scanned to identify the value that maximizes eq. 8.2. A set of various scan values can be seen in fig. 8.3 where the optimal lower BDT bound, which maximizes the figure of merit, can be seen with the orange circle.

With this optimization procedure the value of the BDT range that maximizes the merit function was found to be [0.365, 1.00], with the corresponding number of signal events and the expected background in the signal region seen in the table 8.1. Normalizing the background and signal

<table>
<thead>
<tr>
<th>Condition</th>
<th>Merit value</th>
<th>$N_{\text{sig}}$</th>
<th>$N_{\text{SRbkg}}$</th>
<th>$N_{\text{FRbkg}}$</th>
<th>BDT cut</th>
</tr>
</thead>
<tbody>
<tr>
<td>S+B&gt;0</td>
<td>5.604 ± 0.119</td>
<td>48.68 ± 0.28</td>
<td>26.77 ± 3.16</td>
<td>151.7 ± 7.68</td>
<td>0.365</td>
</tr>
</tbody>
</table>

Table 8.1: BDT range optimization results obtained from maximizing the figure of merit in eq. 8.2. The resulting value of all the parameters can be seen along with the optimal BDT lower cut.

MC to the values obtained from the optimization procedure yields the distribution of invariant mass (fig. 8.4a) and decay time (fig. 8.4b) seen below.

The uncertainties seen in fig. 8.4 are the statistical uncertainties only, which for the signal distribution are very small since it’s the product of scaling the full statistics MC to the expectation value assuming the SM prediction. Optimizing the BDT cut increases almost by a factor of 2 the number of expected events in the signal region, in comparison with the rough estimate performed in sec. 8.2.1. The number of background events was found to be small in that given BDT region with a signal to background ratio of $S/B \sim 2$. Applying an elaborate background subtraction such as the sPlot will allow to access almost all of the available signal events and thus increase the statistical power of the analysis.
8.3 Mass model

The sample selection criteria between the branching fraction and effective lifetime analysis are the same, except for the use of the BDT variable. Therefore the invariant mass models from the branching fraction are used also for the invariant mass fit for the extraction of $s$Weights, although the yields are different. The general form of the PDF used for the mass fit can be seen in eq. 8.3,

$$P_{\text{tot}}(m_{\mu^+\mu^-}) = N_{\text{sig}}P_{\text{sig}}(m_{\mu^+\mu^-}) + \sum_i N_{i\text{bkg}}^i P_{i\text{bkg}}^i(m_{\mu^+\mu^-})$$  \hspace{1cm} (8.3)

where $i$ represents the different background sources, $N_{\text{sig}}(\text{bkg})$ the number of observed decays and $P_{\text{sig(bkg)}}$ the analytical PDF models for each source. The functional models along with the source they correspond to can be seen in the following list.

- **$B^0_s$ signal model**: Double Gaussian with independent means and sigmas. Gaussian constraints are applied on the mean and sigma of the signal model to accommodate for systematic effects discussed in the next chapter.

- **SSSV background model**: Part of the partially reconstructed background (same-side/same-vertex) described by an exponential function, with exponential slope shape parameter and the yield free. The missing components are expected to have small contributions and hence not modelled analytically. However they were studied as part of the systematic uncertainties.

- **Combinatorial background model**: Describes the continuum background source, modeled with $O(1)$ Chebychev polynomial, with the slope and yield parameters allowed to vary freely by the fitter.

As it can be seen not all the background models are included, as for the branching fraction analysis. Background sources e.g. the $B_c$ and the $B^0_d$ were studied as sources of systematic un-
certainties, due to their small number of expected events.

The biggest difference between the branching fraction model and the effective lifetime relies on the parameters of the background shapes. In the branching fraction analysis those parameters were constrained by a certain functional model across the BDT bins (e.g. linear for Chebychev slope, constant for exponential slope). In the lifetime invariant mass fit those constraints are irrelevant since only one BDT bin is used. The second difference between the two mass models, concerns the shape parameters of the signal. In the branching fraction case the shape parameters were frozen to the values obtained from the MC fit, and allowed to vary only as part of the systematic uncertainties. In the case of the lifetime measurement the signal shape parameters were constrained with a Gaussian with the central value taken from the MC and the width according to the systematic uncertainties studied in the branching fraction. The values of the parameters and their resolution in the Gaussian constraint can be seen in table 8.2.

From the branching fraction analysis the systematic uncertainties related to the shape of the signal model concern the shift observed between data and MC in the $J/\psi$ and $Y$ mass. The mass scale systematic uncertainty was studied by shifting the position of the $B_{sJ}$ peak by $\pm 5$ MeV. For the lifetime analysis the 5 MeV shifts are included as the width of the Gaussian constraint, effectively including its effect in the statistical uncertainty.

The second shape related systematic uncertainty arises from the signal resolution on the dimuon invariant mass window. In the branching fraction analysis this source was studied with the use of the $J/\psi \rightarrow \mu^+\mu^-$ and $Z \rightarrow \mu^+\mu^-$ decays and a $\pm 5\%$ variation was observed. For the lifetime analysis the strategy is equivalent; the $\pm 5\%$ variation was included in the Gaussian constraint for the Gaussian width. With this approach both shape related constraints are accounted for and treated as statistical uncertainties by the invariant mass unbinned extended maximum likelihood fitter.

The result of the invariant mass fit determines the signal yield and its correlations with the calculation of the covariance matrix. Subsequently, the covariance matrix is inverted in order to extract the per event $s$Weights for each of the event sources. Because of this procedure, the signal $s$Weights will depend - among others - on the backgrounds invariant mass models, therefore an additional systematic effect on the choice of the background models was performed, and discussed in the systematic sources section below. Testing the fit procedure and estimating the performance requires an accurate generation of pseudoexperiments. The methods used for generating toy samples is described in the following section below.
8.3.1 Toy sample generation

The mass fitter validation and the lifetime determination closure tests are performed with pseudoexperiments. Two approaches were taken in the lifetime analysis, where both of them were tested and compared to determine the most accurate representation of the data sample. The first approach is based on the statistical technique of sampling sub-sets of fluctuating expected number of events from the large full MC samples of the analysis. This technique is called bootstrapping [151] and imposes a strict requirement that the number of events extracted from the full sample is significantly smaller than the total sample, to assure that all toy MC are independent, and no oversampling is performed.

From the BDT cut optimization (sec. 8.2.2) it was concluded that in the analysis invariant mass window, ~ 49 signal events are expected and ~ 152 background events. A dedicated method was developed in the analysis framework to extract with a Poisson fluctuation (mimicking the statistical behaviour of particle physics processes) from the full signal MC the expected number of events and the equivalent number of expected background events from the background MC. Both procedures perform the extraction of the events from the corresponding MC samples following the optimal BDT cut and are thus limited in statistics (with quite a severe limitation in the case of the background MC). The details for the bootstrap method can be seen below in table 8.3.

For validating the performance of the invariant mass fit strategy outlined above, it was applied on the bootstrap extractions containing both signal and background. The distribution of the resulting signal and background yields were extracted from the toy fits to estimate the average number of events in each model. Since the background fitted events were expected to be more unstable in comparison to the branching fraction case (due to the missing cross-BDT bin dependence) attempts were performed to stabilize it. The various configurations studied can be seen in the list below.

- **Standard configuration**: Signal shape constrained to accommodate the mass scale, resolution systematic uncertainties and background shape parameters free to be determined from the fitter.

- **Constrained exponential**: Signal shape and exponential slope constrained from the value taken from MC since this is also the value provided by the cross-BDT bin constraint in the branching fraction case. The polynomial shape is free to be determined from the fitter.

- **Constrained polynomial**: Signal shape and polynomial slope constrained according to the value predicted by the cross-BDT constraint from the branching fraction analysis.
8.3 Mass model

The exponential shape was left free to vary from fitter.

- **Frozen signal shape**: Signal shape frozen and background shape parameters free to vary from fitter.

- **Frozen signal shape and constrained exponential**: Signal shape frozen (similar to branching fraction) and exponential slope constrained (from cross-BDT constraint), polynomial slope free to vary from fitter.

The constrained values for the signal shape were taken as described above from the study performed for the branching fraction analysis on MC toys. For the background shapes the constraints were taken from the fit on the full background MC distributions and introduced in the fitter as Gaussian functions penalties in the total likelihood. Each component in the fit has been generated with a Poisson distribution for the number of events, hence the expected distributions for all the components after the fit should reconstruct the generation distribution, if the fitter is working properly. The yield distributions on bootstrap toys can be seen in fig. 8.5, which as expected have a Poisson shape.

The average number of events in all cases are for signal 49 and for the combined background 152. The mean and RMS on the signal and background yields for all the different configurations can be seen in fig. 8.6.

In the combined plots of fig. 8.6 none of the configurations produces a mass fit yield central value that is aligned with the generation. Multiple sources can produce the discrepancies. The most prominent one arises from the oversampling of the background MC due to the low statistics it contains after the optimal BDT cut. Such an effect would yield to the resulted toy samples being affected by statistical fluctuations and thus bias the number of fitted events. To test this approach toys generated from analytical templates were created and tested for their performance.

To generate analytical toys, PDF shapes for both the invariant mass and the proper decay time had to be developed. The models used for the mass were taken from the branching fraction analysis, whereas the proper decay time shapes were developed based on the theoretical expectation for such distributions. As discussed in the beginning of this chapter the decay time distribution of a particle is described with an exponential. In the case of a measurement the exponential is smeared with a Gaussian term. In addition, selection effects enter the distribution (e.g. BDT) and create an acceptance term in the full PDF modelled with an error function. The list of all the analytical models can be seen in table 8.4. The extracted PDF were fitted both in the signal and the background MC, yielding to an acceptable agreement (p-Value > 0.05) between data and models, seen in fig. 8.7. The discrepancy observed in the signal MC peak arises from the fact that the full simulation sample has much higher statistics than the expected value in data. For this reason the disagreement wasn't modelled, since the final data fit will not be sensitive to it. Following the MC fit, the derived templates from the fit result were used to generate toy samples that can be used instead of the already existing bootstrap samples. Since the analytical templates were fitted on the weighted MC distributions the derived toys contain all the required weight effects. To validate their performance the exact same study as described...
8.3 Mass model

(a) Signal yield.

(b) Combinatorial yield.

(c) SSSV yield.

Figure 8.5: Model yields for bootstrap toys generated from the full analysis MC samples after the optimal BDT cut. With black the standard fit configuration (baseline), with red the configuration with the exponential constraint (EC), with green the polynomial constrained configuration (CC), with blue the configuration with the frozen signal shape (FC) and with yellow the frozen signal with the constraint on the exponential slope (FS\+EC). The shape for all different event categories are expected to be Poisson distributions since they are statistical processes around the most probable value.

<table>
<thead>
<tr>
<th>Source</th>
<th>Mass model</th>
<th>Decay time model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal</td>
<td>Double Gaussian with shape constraints</td>
<td>Error func. and exponential smeared with a Gaussian</td>
</tr>
<tr>
<td>Combinatorial</td>
<td>Chebychev $\theta(1)$</td>
<td>Exponential smeared with a Gaussian</td>
</tr>
<tr>
<td>SSSV</td>
<td>Exponential</td>
<td>Exponential smeared with a Gaussian</td>
</tr>
</tbody>
</table>

Table 8.4: Analytical templates used for toy generation in mass and proper decay time.

above with the bootstrap toys was performed with the variations in the background and signal models at the level of the mass fit. The resulting mean and RMS of the distributions can be seen in fig. 8.8 where the difference with 8.6 can observed in predicted average values of the two methods.

In the bootstrap case the number of events injected in each model vary significantly from the values obtained after the fit, whereas in the case of the analytical templates the number of
8.3 Mass model

(a) Signal yield.  
(b) Combinatorial yield.  
(c) SSVV yield.

Figure 8.6: Bootstrap toys summary plots for mean and RMS. The point y-value represents the mean of the distribution for the different configurations of the fitter and the error bar indicates the RMS of the yield distribution.

events are predicted accurately, leading to a robust toy generation. The default fitting configuration for the mass fit reproduces the expected number of events generated with a reasonable accuracy. In addition, it can be seen that the most conservative case, in which the signal shape is frozen to the MC derived fit value, produces similar results with the default configuration, indicating that a frozen shape, as used in the branching fraction analysis, is not required.

The analytical toys allow to overcome the constraint on the number of extractions observed in the bootstrap method and therefore from this point and on, unless specified, the word toys will refer to samples derived from the analytical templates. The last closure test performed to validate that the bootstrap procedure does not have any intrinsic bias was to generate a large toy sample from the analytical templates and bootstrap from those in the same way as performed with the full simulation samples. In this way a cross-check was performed to clarify whether the observed reduced number of fitted events is due to statistical fluctuations or due to other parameters in the bootstrap procedure. The mean value of fitted events for the different configurations when the bootstrap samples were extracted from analytical references can be seen in fig. 8.9. The yellow crosses represent the mean values seen in fig. 8.6 and indicate that indeed the number of fitted events is not lower due to any intrinsic issue on the bootstrap procedure but rather due to the fact that the samples extracted from the full MC are an underfluctuation due to the limited number of background events after the BDT cut.
Figure 8.7: Fit projections on different signal and background components in mass and proper decay time. The fits were performed with all the MC weights applied and hence some of the p-values show a poor agreement between data and fit projection. However, if the Pearson $\chi^2$ is used a better agreement between data and fit models can be achieved.

Having identified and verified the procedure for generating toy samples both in proper decay time and invariant mass, it is possible to produce $sPlot$ signal projections that mimic the observed behavior in data. An example invariant mass fit on a toy sample extracted from the analytical templates with the corresponding $sPlot$ signal projection can be seen in fig. 8.10. The pseudoexperiments generated weren’t only used for the validation of the invariant mass fitter performance. They were also used to generate $sPlot$ signal projections in order to identify the best performing procedure (highest sensitivity on the lifetime) for the extraction of the lifetime value. For this reason the following section will focus on the tests applied on the various
8.4 Lifetime fitting metric

The \textit{sPlot} signal projection contains the background subtracted decay time distribution from which the lifetime can be extracted. The \textit{sWeights} are applied for each event and hence the resulting binned distribution is a histogram filled with weights, which can also contain bin with negative contents. Fitting the binned signal lifetime \textit{sPlot} content to extract the lifetime value is a demanding task and can be performed with various methods.

The first method that was used by both LHCb and CMS [144],[146], uses an analytical template function fitted to the \textit{sPlot} signal projection. Creating such an analytical functional model requires the precise modelization of the signal decay time selection efficiency arising e.g. from detector inefficiencies, selections (e.g. BDT, triggers, $L_{xy} > 0$ cut). In addition, the likelihood used to fit the analytical model needs to be modified to account for the possibility of negative weights. In the case of the LHCb analysis the solution to the likelihood modification was achieved with the use of the \textit{sFit} method [154], and for the CMS case a variable binning was chosen to avoid negative bin contents [151]. In both cases the impact of such a choice affected

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{figure8_8.png}
\caption{Analytical toys summary plots for mean and RMS. The point y-value represents the mean of the distribution for the different configurations of the fitter and the error bar indicates the RMS of the yield distribution.}
\end{figure}
8.4 Lifetime fitting metric

Figure 8.9: Mean number of fitted signal events for the different mass fit configuration when the bootstrap samples are derived from analytical toys. With the yellow crosses the number of fitted signal events in the default configuration where the bootstraps are extracted from the full MC samples can be seen. The different colours indicate the signal yield from different bootstrap toys extracted from analytical reference samples.

(a) Invariant mass fit on toy sample.  
(b) Signal sPlot projection.

Figure 8.10: Example signal sPlot projection obtained from the unbinned maximum likelihood invariant mass fit applied on a toy sample.

mainly the uncertainty returned by the fitter since the treatment of the low statistics bins was not perfect and hence extra corrections had to be applied.

The second approach considered for the lifetime extraction is to perform a fitting based on lifetime dependent templates rather than analytical functional models. Such templates can be obtained from the full signal MC distribution weighted with a per event weight derived from the truth decay time distribution (discussed in detail in sec. 8.4.1). The key advantage of this approach arises from the modelization of the efficiency function. The signal MC provides a faithful modelization of the efficiency derived from the detector simulation along with the selections applied both in data and MC. To extract the $B^0_s$ effective lifetime the signal MC templates need to be fitted to the signal sPlot extracted from data. This is performed with a $\chi^2$ figure of merit described below. Again the limitation factor of treating low statistics bins arises mostly from the assumptions of the $\chi^2$ itself. For this reason a method of treating such bins is
required in order to include them in the fitting metric. Constructing the appropriate form of the $\chi^2$ variable used for the fit will be described in the following sections along with the method used to generate lifetime dependent templates.

### 8.4.1 MC template generation

Fitting based on lifetime dependent full MC histogram templates requires the parameterization of the full simulation sample as a function of the signal effective lifetime. The main limitation of the full simulation sample arises from the fact that it was generated with a fixed value of the lifetime. Changing the generation value would require to re-run the full simulation chain of ATLAS and hence consumes a significant amount of time. To avoid the lengthy process of re-generation a weighting method based on the truth decay time ($\tau_{\text{truth}}$) distribution was employed. Each event in the sample is generated with an exponential distribution, hence taking the ratio of the generation exponential with an exponential of a different truth lifetime value calculates a per event weight. The overall effect of applying those weights to the truth decay time is to re-weight the full signal MC to a new lifetime value. Furthermore the truth based weights can be applied to the reconstructed decay time ($\tau_{\text{reco}}$) value modifying its distribution according to the new generation value. It was known that at truth level that Pythia\,[165] MC generated the decay time distribution of the B meson with an exponential function, for a certain value of the lifetime consistent with the SM prediction ($\tau_{B^0_s} = 1.4727$ ps). Knowing the generation distribution and lifetime values allows the derivation of the weight seen in eq. 8.4,

$$w(\tau_{\text{truth}}) = \frac{e^{-\tau_{\text{truth}}} / \tau_{\text{target}}}{e^{-\tau_{\text{truth}}} / \tau_{\text{generation}}} = \exp \left\{ \frac{\tau_{\text{truth}}}{\tau_{\text{generation}}} - \frac{\tau_{\text{truth}}}{\tau_{\text{truth}}} \right\}$$  (8.4)

where $\tau_{\text{generation}} = 1.4727$ ps is the generation value for the full simulation sample, $\tau_{\text{target}}$ the desired lifetime value for the template and $\tau_{\text{truth}}$ the truth decay time for each event.

The derived weights are subsequently applied to the reconstructed decay time variable, providing effectively a set of templates as a function of the truth lifetime. The re-weighted templates can be seen in fig. 8.11, with the various colours indicating a few example of different lifetime values.

Re-weighting the templates provides a major advantage for the parameterization of the efficiency term of the recorded data. Effects like the detector misalignments are not present since they are not included in the generation file; however, to account for those a detailed study with the use of the control channel was used as part of the systematic uncertainties and discussed in the following chapter.

The re-weighting method has also some limitation factors. The most prominent limitation arises from the shape of the reconstructed decay time distribution itself. Since the shape is an exponential inevitably as the decay time values increase the available statistics will decrease rapidly. This effect causes the re-weighted histogram templates to rely on the same events multiple times especially for larger decay time values; hence, those templates might be affected largely by statistical fluctuations appearing in the tails of the original distribution. For the lifetime case the solution to this limitation comes from the fact that the signal MC sample has a
8.4 Lifetime fitting metric

Figure 8.11: MC histograms templates for various lifetime values. With the black points the default generation value, with red the 1 ps re-weighted template and with green the 2 ps re-weighted version.

large number of events compared to the expected number of events in the signal region. For the lifetime case it was studied and shown that the template generation remains reliable in a region of [0.5, 4.5] ps with the upper limit being significantly further from the world experimental average for the $B^0_a$-meason lifetime [41]. For this reason a set of 450 templates were generated with a step of 0.01 ps per lifetime. The number of templates was identified to be sufficient to produce a fine enough scan for the estimator of choice, which is described in the next sub-section.

8.4.2 Choice of the lifetime goodness of fit estimator

In 1795 Gauss provided a method in order to compare a set of measurements with their uncertainties along with a set of expectation values taken either by an analytical function or theory predictions [163]. In the simplest form when a set of known measurements $y_i$ with their uncertainties $\sigma_i$ was obtained and needs to be compared with the theoretical prediction with a given functional form of $f(x_i; \theta)$, the $\chi^2$ variable takes the form seen in eq. 8.5.

$$\chi^2 = \sum_{i=1}^{N} \left[ \frac{y_i - f(x_i; \theta)}{\sigma_i} \right]^2$$  \hspace{1cm} (8.5)

Each term in the sum raises the dimension of the $\chi^2$ function by one order, making the expectation value $E[f(\chi^2; N)] = N$. The best estimate on the parameter of interest $\theta$ can be found at the position where for different values of $\theta$ the $\chi^2$ has its minimum value.

In the most general case the measurements can be correlated which requires the modification of eq. 8.5 to accommodate for such effects. The resulting general form can be seen in eq. 8.6,

$$\chi^2 = (y - f(\theta))^T V^{-1} [y - f(\theta)]$$  \hspace{1cm} (8.6)

where $y = (y_1, y_2, ..., y_N)$, $f(\theta) = (f(x_1; \theta), f(x_2; \theta), ..., f(x_N; \theta))$ and $V$ the covariance matrix of the measurements $y$ [163]. Again the best estimate $\hat{\theta}$ can be found at the minimum value of eq. 8.6
8.4 Lifetime fitting metric

for various values of $\theta$.

If the measurement uncertainty is found to be Gaussian the $\chi^2$ theory coincides with the maximum likelihood estimator when a Gaussian function is used. The two methods are connected in the case of eq. 8.5 with eq. 8.7,

$$
\chi^2 = -2 \ln L + 2 \ln c
$$

(8.7)

where $c = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_i}$. Using the result of eq. 8.7 allows to determine the uncertainty on the estimated parameter $\hat{\theta}$ from the contour by moving one unit around the $\chi^2$ minimum [163]. An example of such a distribution can be seen in fig. 8.12 with the red line noting the $\pm 1$ location where the uncertainty on the parameter of interest can be found.

In the case of asymmetric errors the distribution of the $\chi^2$ variable as a function of the parameter of interest can be estimated around the minimum with an asymmetric parabola providing a different upper and lower uncertainty [151].

As shown above the $\chi^2$ function is widely used when Gaussian errors that are present in measurements such as the bin content of histograms. For counting histograms the bin content distribution is equal to a Poisson which for a high number of entries is approximated by a Gaussian. For this reason Gaussian errors in each bin can be assumed, where $y_i$ are the contents for each bin. The sPlot distribution however contains weighted events for the signal projection. Each bin content instead of showing the number of events entering the bin has as expectation value and uncertainty as shown in eq. 8.8,

$$
y_i = \sum_{i} n_i w_i \quad \text{and} \quad \sigma_i = \sqrt{\sum_{i} n_i w_i^2}
$$

(8.8)

where $w_i$ are the per event weights entering each of the bins [151]. For the large number of entries in a bin eq. 8.8 approximates a Gaussian distribution for the bin content allowing the use eq. 8.6 for the parameter estimation. However, for the bins with a low number of entries the Gaussian approximation in eq. 8.8 doesn’t hold anymore since the underlying distribution deviates from a Gaussian [151]. Identifying in which sPlot bins the Gaussian approximation

Figure 8.12: Example distribution of $\chi^2$ variable for one parameter of interest. The red line indicates the position in the contour where $\Delta \chi^2 = 1$ and the uncertainty can be calculated.
8.4 Lifetime fitting metric

holds, a large number of toys was used in order to sample the bin content distribution for each bin. The resulting sampled distributions for a high statistics bin and a low statistics bin can be seen in fig. 8.13, where bins with no content are excluded.

In the case of the high number of expected signal events the Gaussian approximation of eq. 8.8 holds, since the sampled distribution has a Gaussian shape (tested with a binned Gaussian fit on the derived distribution). In the case of the low number of signal events bins, seen in fig. 8.13b, the number of entries is lower than in fig. 8.13a, thus the resulting distribution deviates significantly from a Gaussian distribution. Due to the low number of expected signal events in that given sPlot bin the resulting distribution is mostly affected by the PDF shape of the weight distribution rather than the Poisson effect from the event count.

Dealing with low statistics bins needs a careful treatment. The most straightforward solution would be to exclude such bins from the sum seen in eq. 8.5, transforming it for the sPlot case into eq. 8.9.

\[
\chi^2 = \sum_{bin=1}^{N'} \left\{ \frac{\sum_{i=1}^{N} w_i - f(x_i;\theta)}{\sqrt{\sum_{i=1}^{N} n_i^2}} \right\}^2 \quad N > N_{cut} \\
0 \quad N < N_{cut} 
\]

(8.9)

Where \(N'\) is the number of total bins, \(f(x_i;\theta)\) the expectation value taken from the scaled MC template, \(i\) the events entering each bin and \(N_{cut}\) an arbitrary cutoff for including or excluding a bin from the sum based on the number of events entering a given bin. Nominally, in particle physics the cutoff for low statistics bins is set to at least 10 events entering a given bin.

Excluding low statistics bins was used as a first attempt of the parameterization for the fitting on the sPlot signal projection. However, due to the shape of the fitted histogram (e.g. exponential) such a decision yielded to a major reduction on the sensitivity on the parameter of interest (lifetime) when the fitter was tested with toys. To avoid the decrease of the sensitivity on the parameter of interest a treatment for such low statistics bins is required to be performed.

Various modifications to the definition of the \(\chi^2\) variable that allow the inclusion of the sPlot low statistics bins were explored, and tested with toy samples as outlined in the next sections. The parameterization of each approach and their rationale will be discussed in the following
8.4 Lifetime fitting metric

sub-sections. In addition each of the approaches explored will be tested on toy samples generated with the same lifetime value, to identify the impact on the parameter of interest sensitivity (RMS of residuals) and the error estimation returned by the fitter (RMS of pulls).

Toy-based approach

The most brute-force approach to include low statistics bins makes use of toy samples to sample the expectation value and spread of each sPlot bin to use in the goodness of fit estimator. To achieve this estimation a large number of toys was generated for a set of lifetime values. Each of those toys was fitted with the invariant mass models described above and the sPlot signal projection was obtained. In each decay time bin the expectation value and error (eq. 8.8) were obtained to sample their underlying distribution, excluding the bins in toys who have zero entries. Including toys with zero bin content is the main source of bias in using eq. 8.8 directly. The main effect is that the distributions obtain a huge peak at zero which underestimates the uncertainty significantly. An example of such distributions for a set of sPlot bins can be seen in fig. 8.14.

From those distributions both the mean of the expectation value and the content spread can be taken as best estimates for the expectation value and the uncertainty of the sPlot distribution. An alternative approach is to take the mean and RMS of the bin content distribution, attempting to treat each bins as Gaussian distributions. There are multiple limitations from us-

Figure 8.14: sPlot expectation value and error distributions samples from 10000 toys for 12 sPlot decay time bins for 1.4727 ps.
8.4 Lifetime fitting metric

ing the mean and RMS of each distribution or just the mean of each distribution, arising from the fact that in the low statistics bins have only a limited amount of entries; thus, the estimated values are subject to statistical fluctuation. Additionally the distribution of the low statistics bins deviates from a Gaussian distribution, which is also a factor expected to bias the overall performance of the fitter, since an imperfect modelization of those is derived. Measuring such effects is essential to check whether the performance of the fitter improves in comparison to the default case of excluding low statistics bins. The derived templates are called from here on as sPlot templates and their shape can be seen in fig. 8.15.

Comparing the RMS of the bin content distribution with the mean value of the spread distri-

\[ \chi^2 = \sum_{bin=1}^{N} \left( \frac{y_i - mean_i}{RMS_i} \right)^2 \]  

(8.10)

where \( y_i \) is the expectation value from the fitted sPlot distribution, \( mean_i \) and \( RMS_i \) the mean and RMS value of the bin content toy-sampled distribution for each bin. The overall performance of the fitter was tested on toys shown in the next section.

Figure 8.15: sPlot template generate from a certain number of toys. With the red points the mean of the samples distribution in decay time bins and the uncertainty indicates the RMS of such distribution.

Comparing the RMS of the bin content distribution with the mean value of the spread distri-
8.4 Lifetime fitting metric

Poisson error

An alternative approach for parameterizing the $\chi^2$ variable is to use the Poisson uncertainty derived from the expectation value even for the low statistics bins [163]. For the high statistics bins such parameterization holds as indicated by the sampled distribution of the bin content. Treating the low statistics bins (as per the high statistics) yields an overestimation of the uncertainty originating from the fact that the weight PDF is ignored. Such an effect impacts the overall uncertainty returned by the fitter.

Calculating the Poisson uncertainty could were based either on the fitted $sPlot$ or the MC template expectation values. The choice was made towards the MC template since the $sPlot$ expectation value allows the existence of negative bin contents and hence Poisson uncertainties cannot be calculated directly. To account for potential negative fluctuations due to the $sWeights$ in certain bins, the symmetrizing of the Poisson uncertainty was decided; in this way the bound of positive number of events from the Poisson error does not affect the fit result. Effectively, only the upper Poisson uncertainty calculated in each bin is included in the $\chi^2$ evaluation. Using only the upper Poisson uncertainty does not affect the bins which contain a high number of entries since both the upper and lower uncertainties approximate the symmetric Gaussian uncertainty. The resulting parameterization can be seen in eq. 8.11,

$$\chi^2 = \sum_{bin=1}^{N} \left( \frac{y_i - x_i}{\sigma_{UpperPoisson}^i} \right)^2$$

(8.11)

where $y_i$ the fitted $sPlot$ expectation value, $x_i$ the expectation value from the lifetime dependent MC histogram template and $\sigma_{UpperPoisson}^i$ the upper Poisson uncertainty calculated from $x_i$. Nominally Poisson uncertainties assume integer bin contents (counting histograms), however in the case of the MC templates (weighted with lifetime) the bin content can have any non-integer value. For this reason the Poisson uncertainty is calculated based on a linear extrapolation from the closest rounded up and down integer values [151]. Such a parameterization is based on the definition of $\chi^2$ by Pearson, shown in app. B.

Mixed definitions of goodness of fit estimator

The goodness of fit estimator definition can be modified to take advantage of the three approaches above depending on the bin content. Mixing the three mentioned approaches can provide a significant gain on the parameter of interest sensitivity, since it takes advantage of the different statistical properties of each approach. A list of such parameterizations can be seen in table 8.5.

The expectation value can be replaced from the MC template with the $sPlot$ template mean bin content. To avoid complicating the goodness of fit estimator decision the MC templates were compared in terms of expectation values with the toy-based estimation, as seen in fig. 8.16. Both templates show consistent results within the uncertainty and hence can be used interchangeably. Since the $sPlot$ template is computationally demanding and relies more on the sampling of the distribution with toys, the MC templates are going to be used from now on as
8.5 Fitter selection and validation with toy MC

<table>
<thead>
<tr>
<th>Expectation value</th>
<th>Error for high stat. bins</th>
<th>Error for low stat. bins</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC template</td>
<td>Error returned by sPlot</td>
<td>sPlot template estim. error</td>
</tr>
<tr>
<td>MC template</td>
<td>Error returned by sPlot</td>
<td>Symmeterized Poisson error</td>
</tr>
<tr>
<td>MC template</td>
<td>Symmeterized Poisson error</td>
<td>sPlot template estim. error</td>
</tr>
</tbody>
</table>

Table 8.5: Mixed parameterizations of the $\chi^2$ metric for sPlot signal projection fits. Referred as low statistics bins are the sPlot bins that contain less than 10 entries before the sWeights are applied. High statistics are the bins with more than 10 entries.

Figure 8.16: sPlot and MC template comparison for three set of lifetime values. With black the MC template and with the sPlot template as defined above.

the expectation value for each histogram bin.

All the different configurations shown in table 8.5 were tested with toy samples for their performance. The results of all the different tests for the configurations are discussed in the section below, along with the rational for the final decision.

8.5 Fitter selection and validation with toy MC

Selecting the optimal fit metric for the extraction of the lifetime value is one of the important contributions to the analysis performance. Since the sPlot procedure provides a set of weights that are allowed to under-fluctuate into negative bin contents when filled in a histogram, using standard parameterizations for a metric is not possible; hence alternative parameterizations had to be evaluated. In addition a straightforward approach of binning with a variable binning would mean that the analysis sensitivity would be reduced. Identifying the best performing approach was evaluated on toy samples, where the residuals and pull distribution for the parameter of interest was obtained [163].

In the following sections the performance of the $\chi^2$ parameterizations introduced in the previous sub-section will be shown, along with the reasoning for each approach that led to the fitter used on data. For each approach 10 thousand toy samples were generated in invariant mass and proper decay time, on which the sPlot method was applied to extract the background subtracted decay time distribution. Since the bin width and the fit range on the decay time distribution are important parameters for this study (binned fit) as reference values 1 ps bin width in a 0-12 ps range were used, which later were optimized based on their residual and pull performance on the parameter of interest. The residual distribution with respect to the generation value as well as the pull distribution [166] are estimated to evaluate the performance. The
residual distribution indicates if there is any bias on the estimation of the lifetime along with 
the expected sensitivity. The pull distribution provides the order of the bias along with the es-
timation of the uncertainty returned by the fitter. In the following sub-section the optimization 
of the bin width and range will be discussed in detail.

8.5.1 First $\chi^2$ on the high statistics bins

The first approach that was used is the $\chi^2$ definition seen in eq. 8.5. For this approach (as noted 
above and discussed further in app. B) the uncertainty estimated on bins with a small number 
of entries is biased and hence, if not carefully treated, affects the performance of the fitter on 
the parameter of interest by shifting the lifetime distribution towards lower lifetime values. To 
avoid such an issue, a cut-off of a minimum 10 entries on the histogram bin was applied to 
ensure that the calculation of the error approximates a Poisson distribution. The number of 
entries are not evaluated based on the bin content, since this is a weighted histogram, but on 
the actual un-weighted entries of each bin. 

With the parameterization seen in eq. 8.9 the performance can be seen in fig. 8.17. The bias 

![Image](image.png)

(a) Residuals distribution.  
(b) Pulls distribution.

Figure 8.17: Residuals and pulls distribution of the lifetime parameter for the $\chi^2$ configuration 
that excludes the bins with a low number of entries, tested on toy samples.

of the fitter can be evaluated by the mean of the residual distribution. The performance of the 
fitter for its returned error arises from the RMS of the pull distribution. As seen from mean 
value of the residuals, the fitter is unbiased, but a bias is present on the returned error for the 
parameter of interest. This effect arises from the exclusion of the low statistics bins which are 
essential for the estimation of the lifetime, since the information on the lifetime appears in the 
shape of the distribution. The compromise applied on the shape due to the exclusion of the 
low statistics bins yields an average number of fitted bins to be 4 out of 12. 

Additionally, an inconsistency is observed between the upper and the lower uncertainty dis-
tribution seen in the pulls. Such a discrepancy indicates that the correction to the estimated 
uncertainty has to be different between the upper and the lower error. This feature is prominent 
in all the pull distributions and originates mostly from the imperfect modelization of the sPlot 
signal projected distribution with the MC templates. The parameterization of the fitter which 
excludes the low statistics bins still remains a viable fall-back solution, but performs sub-par in
terms of the determination of the effective $B^0_s \rightarrow \mu^+ \mu^-$ lifetime. The public results available on the same lifetime determination are obtained in signal and background conditions that are not hugely different: it is reasonable to assume that the degradation in performance we see with respect to these results can be improved with a better definition of the lifetime goodness of fit estimator.

### 8.5.2 Toy-based $\chi^2$ performance

The first attempt to include the low statistics bins was made using a brute force method where a large amount of toys was required. As discussed above a large set of signal plus background toys was generated, in a large range of lifetime values. For each of the toys, the corresponding signal lifetime $sPlot$ distribution was derived, thus allowing to parameterize the expectation value and RMS of the signal $sPlot$ as a function of the signal lifetime. These two quantities are then used as estimations for the uncertainty and expected value of each $sPlot$ bin, yielding a toy-driven estimation of the goodness of fit estimator.

Following the generation of the $sPlot$ templates the parameterization of the $\chi^2$ seen in eq. 8.10 was used to fit a set of 10 thousand toys generated with the lifetime value found in the full signal MC file ($\tau_{\text{gen}} = 1.4727$ ps). The performance of the toy-based fitter can be seen in fig. 8.18.

The first remark seen in the residual distribution is that the fitter improves the average sensitivity (from 0.511 ps to 0.36 ps) on the effective lifetime as expected from the inclusion of the low statistics bins. However the most prominent effects seen are the multiple peaks appearing on the residual distribution. The spacing of those peaks suggest that the effect originates from the mis-estimation of the error on the low statistic bins. Using the RMS of the sampled bin content distribution yields heavily suppressed error estimation in the template, especially in the longer decay time bins, thus doesn't account properly for the negative fluctuations on the bin content. Another effect is the drop seen around 0 in the pulls distribution (as seen also above in fig. 8.17). Nominally the left and right hand side (upper/lower uncertainty returned by the fitter) would were expected to have a similar distribution. Therefore such behavior is not expected by the approach used, however can be explained from the fact that the treatment of the low statistics

![Figure 8.18: Residuals and pulls distribution on the measured lifetime for the toy-based $\chi^2$ configuration, tested on toy samples.](image)
bins is yet not perfect.

8.5.3 Poisson error in $\chi^2$ performance

Another alternative considered to include the low statistics proper decay time bins in the calculation of the $\chi^2$ is the estimation the uncertainty on the bin content by using the Poisson uncertainty estimation based on the expected average signal count in that bin. As the error estimated is asymmetric, the choice to adopt in each case the largest between the positive and negative uncertainties was made. The performance of such a parameterization in the fitting metric yields the result seen in fig. 8.19.

Comparing the retrieved residual distribution with the approaches above indicates a clear improvement in the performance of the fitter, with a more Gaussian like shape as expected. The sensitivity on the parameter of interest gained on average a factor two compared to the parameterization that excludes the low statistics bins and the multi-peak structure from the toy-based approach disappeared.

The pulls distribution shows the same asymmetry already found. As mentioned this structure is explained from the fact that the error estimation in none of the configurations is perfect, but by no means it is expected. A further step was taken to check whether correcting the upper and lower uncertainty returned by the fitter provides a standard distribution; thus a Gaussian tail fit was applied yielding to a correction factor noted as $c$ and seen in eq. 8.12.

$$
\begin{align*}
c_\sigma^+ &= 1.46 \\
c_\sigma^- &= 0.81
\end{align*}
$$

(8.12)

Applying the correction factor on the corresponding side of the pull distribution, corrects the distribution to standard Gaussian distributions in each side and does not affect the average RMS of the full distribution. This indicates that indeed the effect has statistical roots rather than something much more complicated to grasp. The distribution of the pulls with the Gaussian fits applied on them can be seen in fig. 8.20.
The Poisson $\chi^2$ parameterization provides the best result from the three approaches defined above, however a cross check needs to be performed by introducing the mixed approaches to see if any further improvement can be achieved.

### 8.5.4 Mixed approaches performance

The mixed approach configurations (table 8.5) indicate the different error estimations used depending on the region of the histogram (high/low statistics bins). Those approaches were mainly explored to identify if they produce any significant gain from the three more conventional parameterization discussed previously. In addition with those approaches cross checks can be made on certain assumptions that were taken while developing the fit procedure. Such an assumption is for example the fact that for large statistics the symmeterized Poisson distribution and the $sPlot$ error will provide equivalent results. This assumption originates from eq. 8.8 where for large statistics it is expected to follow a Gaussian distribution. Equivalently the Poisson distribution is approximated for large number of entries with a Gaussian distribution as well. For this reason the result for row 1 and row 3 are expected to be almost identical. The comparison between the two can be seen in fig. 8.21.

As seen from the mean and RMS parameters of the residuals and pull distribution of the effective lifetime the two configurations provide similar results with any potential discrepancies arising from statistical fluctuations from the calculation of the $sPlot$ templates. The last remaining configuration (row 2) uses the $sPlot$ error and the modified Poisson error configuration for the low statistics bins. There only reason to prefer this configuration over the modified Poisson only case, is the performance on the bias and the sensitivity to the lifetime
8.5 Fitter selection and validation with toy MC

Figure 8.21: Residuals and pulls distribution for the mixed configurations in the $\chi^2$ variable, tested on toy samples, for comparison checks.

parameter. The RMS of the residual distribution (average sensitivity) on the $B^0_s \rightarrow \mu^+ \mu^-$ effective lifetime between the two approaches can be seen in the following table 8.6. In both cases the bias observed is almost identical hence the decision is made based only on the expected sensitivity of the fitter configuration. For this reason the modified Poisson only case was chosen as the baseline configuration for our lifetime extraction metric.
8.6 Fitter application on control channel

From the studies shown above the best performing configuration with the most consistent result is the one that relies on the Poisson $\chi^2$. As discussed in the beginning of the chapter one of the main systematic uncertainties explored for this analysis will be evaluated with the use of the control channel. For this reason the fit procedure derived from the $B_s^0$ channel will be applied almost identical in the $B^+ \to J/\psi K^+$. The only difference arises from the analytical templates that were used for the $sWeights$ extraction.

The evaluation of the systematic uncertainty will have to be evaluated in similar statistics as for the $B_s^0$ sample. The control channel has much higher statistics than the analysis channel, thus toy samples can be extracted directly from the data sample. With this method the estimation of the systematic will be performed on a set of data driven toys which will contain all potential effects (e.g. detector misalignments) arising during data taking.

The final systematic extraction will be performed based on the lifetime fitter discussed above. For this reason any assumption on the bin content and the range will be affecting the evaluation of the systematic uncertainty; hence the systematic from the control channel will be studied for the different bin width and fit ranges. The procedure used for the systematic evaluation with the control channel is described below.

Initially as for the $B_s^0$ case an invariant mass fit is required for the background subtraction. The analytical templates used for the unbinned maximum likelihood fit with their constraints can be seen in table 8.7. The applied Gaussian constraints on the signal model are derived based on the $B_s^0$ model to ensure a similar configuration of the fitter. For instance the mean and sigma of the Gaussians were constrained to accommodate the 5MeV variation on the peak position (mass scale) and the 5% variation in the resolution (di-muon resolution). For the background models the shape constraints were applied to mostly keep the fitter under control, except the constraint applied on the switching point of the inverse error function which is correlated to the position of the signal peak; hence evaluated similarly based on the mass scale constraint.

All the yield parameters on the contrary of the shape parameters were left free to vary by the extended unbinned maximum likelihood fitter. Configuring the invariant mass fitter in such

<table>
<thead>
<tr>
<th>Source</th>
<th>Mass model</th>
<th>Constrained parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal</td>
<td>Double Gaussian</td>
<td>Constraint on shape parameters</td>
</tr>
<tr>
<td>Combinatorial</td>
<td>Chebychev $\theta(1)$</td>
<td>Constraint on shape parameters</td>
</tr>
<tr>
<td>Partially Reconstructed</td>
<td>Inv. error function</td>
<td>Constraints on shape parameters</td>
</tr>
</tbody>
</table>

Table 8.7: Invariant mass model for control channel unbinned maximum likelihood fit.
way allowed to use the same likelihood parameterization as for the analysis channel and thus achieve a similar performance of 95% converging fits when the fitter was tested on toys. As can be seen, the mass models used for the control channel differ significantly from the mass templates used in the branching fraction analysis. Such a difference is not expected to be affecting the measurement of the systematic uncertainty since the studied samples have much less statistics than the full data sample; thus the full and more complicated model isn’t required. The full data set allows for the extraction of three different sets of toys for testing the fit metric derived from the analysis channel. The main reason for having derived three different categories arises from the different S/B between the analysis and the control channel. The different background contributions between the two channels mean that isn’t directly possible to derive toys with exactly the same properties as for the $B^0_s$ channel. For this reason the three types of data driven toys listed in table 8.8 were derived to explore all the possible configurations where either the background or the signal events are the same as for the $B^0_s$ or have the same significance between the two channels.

The configuration of the fitter and its performance was validated on MC to ensure that the

<table>
<thead>
<tr>
<th>Configuration</th>
<th>Number of total events</th>
<th>Number of toys</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same signal</td>
<td>80</td>
<td>1000</td>
</tr>
<tr>
<td>Same significance</td>
<td>30</td>
<td>1000</td>
</tr>
<tr>
<td>Same background</td>
<td>405</td>
<td>160</td>
</tr>
</tbody>
</table>

Table 8.8: The three configurations are calculated based on the expected number of events seen from the $B^0_s$ sample. Hence, data driven toys with the same number of signal events, background events and significance (S/Sqrt(S+B)) were created.

biases due to the fitter are equivalent between the $B^+$ and the $B^0_s$. An example fit result on a toy with the same number of events as the $B^0_s$ can be seen in fig. 8.22, with the corresponding sPlot derived projections.

Following the sPlot application step the lifetime fitter is applied to the different toy categories.

Figure 8.22: Example invariant mass fit on a toy sample with the same signal statistics as for the $B^0_s$ and the corresponding signal sPlot projection.

The extracted value in each case is compared with the SM prediction, which is a well known property. The derived residual distribution can be seen in fig. 8.23, estimated with 1 ps bin
8.7 Fitter expected performance

width in a range of 0-12 ps on the data driven toys with the same number of signal events. The fitter performance is similar to the performance seen on the \( B^0_s \) toys, with any difference arising from the discrepancies between the two decays. Validation of the fit procedure with the \( B^+ \) control sample allows the use of it in the upcoming described study for the expected performance of the fitter for different bin widths and fit ranges.

8.7 Fitter expected performance

To identify the best performing fit range and bin width for the lifetime fit metric three components are required. The first component concerns the expected statistical uncertainty returned by the fitter for all the different ranges and bin widths. The second component concerns the evolution of the observed fit bias originating when testing the fit metric on pseudoexperiments (systematic uncertainty). The last component concerns the main systematic uncertainty, for which the control channel discrepancies are expected to be the major contribution and hence the fitter performance applied on the control channel is included.

To estimate the expected upper and lower uncertainty returned by the fitter when tested on toys the upper and lower error distributions were obtained for a different set of ranges and bin widths. Examples of such distributions can be seen in fig. 8.24, which are obtained for the default 1 ps bin width and 0-12 ps fit range configuration. The expected uncertainty in each case is corrected by the asymmetric Gaussian fit on the pull distribution. The fit is repeated for the different bin width values since it cannot be guaranteed that it’s constant as a function of it. The bias of the fitter and the \( B^+ \) systematic are derived for every configuration by obtaining the mean value of the residual distribution with the SM prediction. All the three different com-
8.7 Fitter expected performance

Figure 8.24: Distribution of the upper and lower error returned from the fitter when applied on toys.

Components are therefore included in the fig. 8.25, seen below. The plot shows the evolution of the combined uncertainty (combination in quadrature of each uncertainty) with the bin width. Equivalent plots were created for the fit range which indicated that the overall uncertainty is constant for any fit range that includes the whole number of events. Hence, the 0-12 ps range was chosen, since it's aligned also with the choices of the other experiments. The red belt around the statistical uncertainty shows the RMS of the error distribution to indicate that the variation expected on the statistical uncertainty is much larger than any variation of the bin width and therefore secondary.
Including the RMS in the overall picture shows that the bin width does not affect the fitter performance (constant behavior), with the combined uncertainty central value indicating an inclination towards smaller bin widths. However, the choice for the bin width is 1 ps due to the overall more stable performance in terms of the systematic uncertainties (fit bias/control channel). This decision is also aligned with the decisions made by the other two experiments (CMS: [144], LHCb: [146]).

8.8 Summary for fitter configuration

To conclude the studies performed on the toy samples with the fit metric that uses a symmeterized Poisson uncertainty the expected uncertainties on the parameter of interest can be seen in eq. 8.13,

\[
\begin{align*}
\sigma^+ &= 0.615 \\
\sigma^- &= 0.216
\end{align*}
\]  

(8.13)

after applying the correction factor derived from the pull distribution. In addition, due to the imperfect fitter parameterization (specially for the low statistics sPlot bins) a fit bias is taken as a systematic uncertainty of \( \sigma_{\text{fit bias}} = 0.011 \) ps. Another source of systematic uncertainties was estimated with the usage of the control channel which is discussed in more detail in the following section. The effect from the control channel is \( \sigma_{\text{control}} = 0.041 \) ps. The overall configuration of the fitter producing the most stable performance is with a 0-12 ps fit range and 1 ps bin width. In addition to the statistical effects and the two systematic sources explored in this section, there is a further set of systematic effects appearing in the measurement which are explored in the next section. Identifying all the systematic effects is important prior to running the fitter on the 15/16 data sample, to ensure that all the effects affecting the measurement are under control. The estimation of the systematic uncertainties was made mostly with the use of pseudoexperiments, and when not possible the \( B^+ \) was used.
The fitter configuration allows for a measurement on the lifetime accounting only for statistical uncertainties. Systematic uncertainties arise in the analysis mostly from assumptions in the procedure followed. These assumptions will be challenged in this chapter, with uncertainties assessed where appropriate. The main systematic sources will be discussed, with their rationale and the method used for their estimation. The chapter will conclude presenting the final result on the lifetime measurement, including statistical and systematic uncertainties, as well as the published CMS and LHCb results.

9.1 Fit accuracy

What is defined as fit accuracy are all the effects that are intrinsic to the fitter configuration. Closure tests with toy simulations generated in the null hypothesis show a bias in the extraction of the lifetime at the level of 11fs. The source of this relatively modest bias has not been investigated in detail, given its marginal impact in comparison to the expected statistical uncertainty (expected on average to be 30x larger). The full bias observed is included as systematic uncertainty in the final result, seen in eq. 9.1.

\[ \sigma_{\text{fit accuracy}} = -0.011 \pm 0.003 \text{ ps} \]  

(9.1)

The bias could come about as a shift to a fairly Gaussian distribution of the residuals, or due to a substantial non-Gaussianity of the toy residuals (e.g. due to the presence of toy categories with substantially different residual distributions). For safety, it was verified that the residual bias is not the consequence of a significant number of toy simulations falling at a large distance from the core of the distribution. The test is performed by comparing the fitter bias estimated with and without a cut-off on the residual distribution tails. Fig. 9.1b compares these two bias estimations for several sPlot bin widths and fit ranges: the behavior of the residuals is substantially unchanged between the cases with and without the cut-off. Fig. 9.1a corresponds to a cross-check that the source is not only due to a shift of the residual distribution core. For this reason the residual distribution for the different bin widths were fitted to the residual distribution that indicated the least bias (e.g. 0.125 ps/bin). The fit result was used as a correction on the various bin width distributions. The observed trend proved to be reduced and however not
9.2 Invariant mass fit and decay time distribution assumptions

The effective lifetime measurement depends on the subtraction of the background from the decay time distribution. The subtraction as discussed throughout the thesis was performed with the sPlot statistical technique which derives event weights based on an extended unbinned maximum likelihood fit performed on the invariant mass.

The assumptions applied on the invariant mass models are crucial for the extraction of the
lifetime signal projection and hence testing their effect is required. Such assumptions concern the analytical PDF models that were used to describe the various event sources; moreover, the background interference with the signal could be mis-estimated due to an imperfect background mass model but also because of neglected backgrounds and the assumption that the background proper time distribution in the signal region is well reproduced by background events further away from the signal. Each of the effects below was tested with appropriate methods as discussed below depending on the nature of the effect under consideration toy MC studies were performed generating events from either analytical models or bootstrapping from ATLAS MC samples.

The overall number of events in all the available event categories studied in this section, for the invariant mass regions after the analysis BDT cut, can be seen in the following table 9.1. In this way a rough estimate of the effect of each source can be obtained.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Data 15/16</td>
<td>146</td>
<td>65</td>
<td>24</td>
</tr>
<tr>
<td>Signal MC</td>
<td>3</td>
<td>45</td>
<td>1</td>
</tr>
<tr>
<td>Combinatorial MC</td>
<td>59</td>
<td>31</td>
<td>32</td>
</tr>
<tr>
<td>SSSV MC</td>
<td>30</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(B^0_d \rightarrow \mu^+ \mu^-)</td>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>(B^0_d \rightarrow \pi^- \mu^+ \nu_{\mu})</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(B^0_s \rightarrow K \mu^+ \nu_{\mu})</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(\Lambda_b \rightarrow p \mu^- \bar{\nu}_{\mu})</td>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>(B_c \rightarrow )</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 9.1: Expected number of events for the various event sources entering the invariant mass region. The values were calculated based on the SM expectations and derivations performed for the branching fraction analysis.

9.2.1 Signal mass model

The first event source tested for it’s mass model is the signal. From the branching fraction analysis it was known that the position of the mass peak between data and MC has a discrepancy of about 2 MeV [155]. In addition the resolution of ATLAS on dimuon masses produces on average a 5% systematic effect [155].

To accommodate for both such effects a different approach to the branching fraction analysis was taken. Instead of varying the generation model by ±5 MeV and the resolution by ±5%, the means and RMSs of the signal Gaussian models were constrained with Gaussian terms in the likelihood, rather than frozen. Applying such constraints includes the systematic effect of the resolution and peak position in the statistical uncertainty, instead of having to estimate it separately.
9.2.2 Background models

The background events can affect with various ways the measurement on the effective lifetime, although the fit on lifetime is performed in the background subtracted distribution. There are three ways the background can affect the fit result. The first way is via the assumed background modelling. The analytical form of the background models was validated on simulated background samples, but is not motivated by specific physics considerations (such as in the case of a resonance). The reliance on MC for these models is somewhat mitigated by the fact that the model parameters are not tied in any way to the MC fits, however the choice of alternate analytical models stills owes to be explored as part of the possible sources of systematic uncertainties affecting proper decay time distributions; thus changing analytical PDFs might affect the fit result. When using analytical models for the generation of toy simulations, the background parameterization was derived in shape and parameter values from simulated background samples. Although reasonable for the generation of toys, those models need to be tested for adequacy and consistency with what is obtained from real data.

The second way that the background can affect the lifetime measurement is via its normalization. If the number of background decays entering the signal region is mis-estimated, then the sPlot distribution might identify wrongly some background decays as signal and vice-versa.

Finally, the last method concerns mostly event categories with small contributions. Such, decays have a small number of expected events in the analysis range and hence introducing analytical PDF models to describe them is just over-complicating the total PDF. However, their effects need to be tested to cross check their impact in the measurement of the lifetime, since some of them might have decay time distribution which in principle could interfere substantially with the signal lifetime distribution.

Assumptions on background inv. mass models

As mentioned above the invariant mass models describing the background distributions are choices that provide a reasonable performance when applied on toys. A different selection of PDF can be used, hence in theory all available options need to be tested. A thorough exploration of functional models is however impossible; hence a reasonable compromise is usually made by testing a set of the most likely alternatives to the original choices.

The alternative models used for the SSSV and combinatorial models can be seen in table 9.2. To estimate the effect of changing the analytical templates the following procedure was applied. The toys generated with the nominal models were used and fitted in mass for the background subtraction by changing the fit model component for the corresponding background to the al-

<table>
<thead>
<tr>
<th>Source</th>
<th>Ref. Model</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combinatorial</td>
<td>Chebychev O(1)</td>
<td>Exponential</td>
</tr>
<tr>
<td>SSSV</td>
<td>Exponential</td>
<td>Gaussian tail</td>
</tr>
</tbody>
</table>

Table 9.2: Alternative PDF modelization for the different background sources along with the nominal invariant mass models.
ternative from table 9.2. Models were swapped one at a time. Example fit results for a specific toy can be seen in fig. 9.3.

For each of the changes the residual distribution in lifetime can be obtained. However a more accurate result if compared to fit result obtained with the reference fit models is to take the distribution of the residual differences for each toy. This approach excludes effects due to toys fluctuations in the estimation of the lifetime residual shift. The resulting distributions for each PDF can be seen in fig. 9.4.

The bias due to the assumed background models can be seen for both cases in eq. 9.2.

\[
\begin{align*}
\sigma_{SSSV Model} &= 0.033 \pm 0.002 \text{ ps} \\
\sigma_{CombModel} &= -0.0014 \pm 0.0007 \text{ ps}
\end{align*}
\] (9.2)

Decay time distribution and background contamination

The second way on the background can affect the lifetime measurement on the signal is due to the assumptions on the normalization and decay time models. The two different background
components have a different behavior in this systematic category and hence a different approach is employed for each component.

The first component discussed is the SSSV. The difference between the SSSV and the combinatorial is that the SSSV is rapidly falling in the lower invariant mass side-band and hence entering very little in the signal region. To measure the effect of the SSSV component affecting the signal decay time distribution the following method was used. The model was removed completely from the toy generation and maintained in the fitting of the toys. In addition bootstrap toys were generated and included on the top of the original distribution, effectively doubling the normalization of the component.

Both methods yielded similar results when the distribution of the difference between the residuals was taken with the nominal approach. To ensure that no statistical fluctuation affected our result the largest of the two values were taken as the effect of the SSSV contamination, seen in eq. 9.3.

$$\sigma_{SSSVCont} = 0.046 \pm 0.006 \text{ ps}$$

(9.3)

For the combinatorial component the procedure is slightly different. Mainly, because the combinatorial background extends across the whole invariant mass region. To ensure that the combinatorial distribution is the same between the two regions and hence the effect of the contamination is constant across the full invariant mass range, the difference between the sPlot of the upper and lower side-band of the combinatorial background distribution was taken and can be seen in fig. 9.5. As it can be seen the sPlot between the upper and lower distribution are averaging to 0, showing no indication of a difference in proper decay time for the upper and lower side-band regions of the combinatorial background. If a potential mis-modeling due to a possible proper decay time and mass correlation for this background is excluded, the only other potential source of systematic uncertainty in the signal lifetime measurement due to this background could arise from a mis-estimation of the number of combinatorial background events.
9.2 Invariant mass fit and decay time distribution assumptions

in the signal region. For this the following procedure was used. The combinatorial background model shape (linear slope) was frozen to the value obtained from the reference fit (full range). The same shape was used - with a free normalization parameter - to repeat the invariant mass fit restricted to a range where either the upper or the lower mass side-bands were excluded. The residual difference between the nominal approach and the upper and lower side-band fits was taken. To ensure again that no effect is misestimated the larger of the two values was taken as a systematic uncertainty, seen in eq. 9.4.

\[ \sigma_{\text{CombContModel}} = 0.034 \pm 0.001 \text{ ps} \]  \hspace{1cm} (9.4)

Neglected background sources

The last way of affecting the lifetime measurement due to the background contributions is with not modeling decays with small number of expected events in the signal region (table 9.1). Such, decays are any semi-leptonic decay with a mis-identified meson or hadron as a muon, the decay of the $B_c$ meson, and the $B^0_d \rightarrow \mu^+\mu^-$ decay. All three categories have a small contribution in the total number of events, thus the same method was used to measure their effect on the signal lifetime. One by one the different categories were introduced in the generation of the toys. Then fitted with the nominal invariant mass model to obtain the signal projected distribution, and fit in lifetime for the residual difference distribution.

The first category discussed is the $B^0_d \rightarrow \mu^+\mu^-$ decay which is expected to have 4 events after the BDT cut of the analysis. The $B^0_d$ inclusive MC sample was used to introduce those decays as bootstrap samples in the toy generation. Again the difference between the residuals with and without the $B^0_d$ events was obtained to measure the effect of not including the $B^0_d$ PDF in the invariant mass fit. The scale of the effect can be seen in eq. 9.5.

\[ \sigma_{B^0_d} = 0.016 \pm 0.001 \text{ ps} \]  \hspace{1cm} (9.5)

The second category of events that was not been modelled analytically in mass is the $B_c \rightarrow J/\psi(\rightarrow \mu^+\mu^-)\mu^+\nu\mu$. Since the mass of the $B_c$ is beyond the $B^0_s$ mass such decays enter the signal region. The total number of $B_c$ decays is expected to be 4 in the full mass window of the analysis. Following the equivalent approach as for the $B^0_d$, bootstrap samples were included for the $B_c$ component, with the scale of the effect being:

\[ \sigma_{B_c} = 0.003 \pm 0.0002 \text{ ps} \]  \hspace{1cm} (9.6)

The last missing background component concerns three decays which are categorized as semi-leptonic decays ($\Lambda \rightarrow p\mu\nu, B^0_d \rightarrow \pi\mu\nu, B^0_s \rightarrow K\mu\nu$). Those decays although having a different final state than the signal can appear within the fit range when one of the hadrons in the final state is mis-identified as a muon. From the branching fraction analysis it was found that the purity of muons is 97% [155], hence the expected number of the combined semi-leptonic decays is 5 events in the full range.

To estimate their contributions in the lifetime measurement a more conservative approach was
taken. Instead of including each component with its contribution corrected by the fake rate of
the corresponding final state hadron, the sum of all the three components assuming they sat-
urate the 5 events contributions was taken. Overall, 15 semi-leptonic events were added from
bootstrap samples with the resulting effect being:

$$\sigma_{semi} = 0.008 \pm 0.001 \text{ ps}$$  \hspace{1cm} (9.7)

9.3 MC modeling

In addition to the systematic effect arising from the event categories modeling, both in invari-
ant mass and decay time, there are also effects arising from the discrepancies between the data
sample and the MC samples that were used in the analysis. Those effect were studied either
with the use of the control channel or directly with the $B^0_s$ samples when possible.

9.3.1 Control channel discrepancies

The lifetime fitter uses signal MC templates parameterized with the lifetime. The main as-
sumption arising from their use is that the efficiency term in addition to the any other effect
(e.g. BDT selection) is reproduced faithfully by the simulation. Since the MC simulation is not
perfect, testing this assumption is crucial for the performance of the analysis.
The simulation between the control and the analysis channel is equivalent, from the point of
view of the software release that was used, and most of the selections applied in the two chan-
nels (different invariant mass cut to the muons). Furthermore the control and analysis data
samples were collected at the same period with triggers applying if not identical, very similar
selections. To first order, it is expected that the MC mis-modeling effects will enter similarly the
signal and the control channel.
As discussed in the previous chapter (sec. 8.6) the control channel has much higher statistics
available than the analysis channel, therefore the data-set can be randomly partitioned and
each sub-sample used to repeat the same measurement multiple times. For this reason three
different categories of toys were made seen in table 8.8. The whole optimization of the invariant
mass fitter was performed initially on the MC sample to avoid any biases. The final systematic
effect, concerning discrepancies between data and signal MC and detector inefficiencies, was
measured on data directly. The samples were partitioned to mimic the statistics of the anal-
ysis channel to ensure that the statistical regime between the two channels is the same. Since
the S/B between the two channels is different, three different types of sub-samples was cre-
ated, with the systematic uncertainty arising conservatively from the sample category with the
highest discrepancy (same signal). The measured systematic effect can be seen in eq. 9.8.

$$\sigma_{B^+} = 0.041 \pm 0.008 \text{ ps}$$  \hspace{1cm} (9.8)
9.3 MC modeling

η dependence

Another effect arising from the discrepancies between the data and MC is the potential difference in performance of the ATLAS detector when the two muons in the final state are crossing each other (cowboys) or are bending away (sailors) from each other. Studying this effect for the lifetime analysis was performed with the use of the control channel, where both the MC samples and the data have much higher statistics. To simplify the study a kinematic variable was identified which is sensitive to the trajectory of the muons. Such variable was identified to be the Δη between the two muons always calculating it by subtracting the positively charged muon η from the negative. Following this calculation allows to partition both the MC and the data samples into the two muon categories. The calculation of the Δη can be seen in eq. 9.9,

$$
\Delta \eta = \eta_{\mu^+} - \eta_{\mu^-}
$$

where with \( \eta_{\mu^+} \) the η of the positive charged muons is noted and with \( \eta_{\mu^-} \) the corresponding negatively charged muon. The separation of the two categories was made based on the sign of the Δη variable.

To identify any potential effect the signal MC Δη distribution was obtained in addition to the data distribution from the control channel. In order to compare the two distributions a background subtraction had to be performed for the data sample. This was achieved with simply extracting the number of background events found with an invariant mass fit. The two distributions then were binned and plotted together, with a variable binning which allows to have a finer binning close to \( \Delta \eta = 0 \), where the two muons are more likely to cross paths, seen in fig. 9.6.

Thereafter the ratio plot of the two distributions was used to generate a per event weight for each candidate in the analysis channel. The next step was to apply the Δη re-weighting on the
toy events extracted from bootstraps and use the lifetime fitter. The difference between the residuals with and without the $\Delta \eta$ weight was obtained, indicating to an effect seen in eq. 9.10.

$$\sigma_{\Delta \eta} = -0.01 \pm 0.004 \text{ ps}$$ (9.10)

### 9.3.2 Weight effects

The last systematic uncertainty arising from discrepancies between data and MC concerns the application of the MC weights. As discussed in the app. A, in order to obtain a faithful representation of the data distributions with the MC a set of weights had to be applied to the MC distribution. Three categories of weights were identified to have a potential effect on the lifetime measurement.

The method to measure the effect of these weights is identical. The weights were excluded in the toy generation but maintained in the signal templates employed for the lifetime extraction, finally the difference of the residual distribution was taken with the case where the weights are present both in generation and fitting. The three categories studied concern the kinematic weights called (Data-Driven Weights (DDW),Quark Level Corrections (QLC)), the trigger weights and finally the isolation weight. All three where studied on the analysis channel with the effect seen in eq. 9.11.

$$\sigma_{DDW/QLC} = -0.009 \pm 0.001 \text{ ps}$$

$$\sigma_{Trigger} = 0.000 \pm 0.000 \text{ ps}$$

$$\sigma_{Isolation} = -0.008 \pm 0.002 \text{ ps}$$ (9.11)

The effect of the trigger weights was found to be insignificant and hence not contributing in the final combined systematic uncertainty, for this reason it was removed from table 9.3.

### 9.4 Heavy-Light state asymmetry

The last source of systematic uncertainty explored for the analysis concerns the potential effect arising from the possible selection bias between heavy and light mass eigenstates. Event selections sensitive to proper decay time are performed throughout the analysis and could therefore change the natural relative abundance of these states. Therefore there is always the concern that the light state decays might be present but discarded due to the selection of the analysis samples.

Such an effect is part of the efficiency as a function of proper time shape. If the functional shape was flat then the admixture of the two states wouldn’t be affected, however having an asymmetric shape indicates that certain decay time values are more favorable than others. The efficiency function seen in fig. 9.7, reports the signal selection efficiency as a function of proper decay time at different stages of the analysis (at the level of sample derivation, signal preselection and final analysis cuts). The final BDT cut clearly shows a bias in the signal proper decay...
9.5 Combined uncertainty

The complete summary of the systematic uncertainties tested for this measurement can be seen in table 9.3. The total uncertainty shown in the last row of the table is the combination of the uncertainties, treating them as uncorrelated. Although during the calculation the uncertainties have a sign indicating the direction of the effect, for the combination a symmeterizing approach was used. This approach is acceptably conservative, since the total systematic uncertainty is still well below the statistical uncertainty of the measurement.

The major sources of systematic uncertainty for the ATLAS experiment arise from the SSSV component leaking into the signal region and the discrepancies between data and MC, measured with the $B^+$ control channel. Having completed the study on the systematic uncertainty, in addition to the expected statistical uncertainty, allows to advance in performing the fit on the data sample to extract the lifetime value observed in ATLAS with the 15/16 sample.
Table 9.3: Summary table of systematic effects and their value, all converted into uncertainties.

<table>
<thead>
<tr>
<th>Uncertainty source</th>
<th>Uncertainty value (ps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fit accuracy</td>
<td>0.011</td>
</tr>
<tr>
<td>SSSV contamination</td>
<td>0.046</td>
</tr>
<tr>
<td>Combinatorial lifetime model</td>
<td>0.034</td>
</tr>
<tr>
<td>SSSV mass PDF</td>
<td>0.033</td>
</tr>
<tr>
<td>Combinatorial mass PDF</td>
<td>0.001</td>
</tr>
<tr>
<td>MC modeling</td>
<td>0.041</td>
</tr>
<tr>
<td>$\Delta \eta$ correction</td>
<td>0.01</td>
</tr>
<tr>
<td>$B_s^0$ kinematics</td>
<td>0.009</td>
</tr>
<tr>
<td>Isolation re-weighting</td>
<td>0.008</td>
</tr>
<tr>
<td>Light-Heavy state mixing</td>
<td>0.012</td>
</tr>
<tr>
<td>$B_d^0$ mis-modeling</td>
<td>0.016</td>
</tr>
<tr>
<td>Semi-leptonic background mis-modeling</td>
<td>0.008</td>
</tr>
<tr>
<td>$B_c$ background mis-modeling</td>
<td>0.003</td>
</tr>
<tr>
<td>Total systematic uncertainty</td>
<td>0.083</td>
</tr>
</tbody>
</table>

9.6 Final result

The first step to extract the lifetime is to perform the unbinned maximum likelihood fit on the invariant mass required by the sPlot statistical tool. The fit has converged nominally without any issue producing the result seen in fig. 9.8, after the analysis BDT cut.

The fit result is in good agreement with the data and hence can be trusted. The resulted number of events from the fit is $51 \pm 10 \ B_s^0$, $36 \pm 23 \ \text{SSSV}$ and $146 \pm 22 \ \text{combinatorial decays}$.

A final validation of the whole procedure was performed by comparing the background generation lifetime models with the distributions obtained in data to avoid any mis-modeling. Any big discrepancy between the generation and the distribution observed in data might put into question the procedure used to estimate many of the systematic uncertainties. For this reason the toy generation model was compared against the data derived background distribution seen in figure {\ref{fig:9.8}}.

Figure 9.8: Invariant mass fit with full model on the 15/16 data set.
9.6 Final result

fig. 9.9. No significant deviation was observed between the two distribution, which indicates that the toy generation is faithfully reproducing the background models. The last remaining step following this validation is applying the lifetime fitter to the data signal projection. Applying the $sWeights$ to the decay time distribution produces the background subtracted signal projection. Applying the fit procedure defined for the analysis produces the following result seen in eq. 9.13.

$$\tau_{B_d^0} = 1.185^{+0.478}_{-0.19} \pm 0.083 \text{ ps}$$  \hspace{1cm} (9.13)

Where the statistical uncertainty was corrected according the pull distribution (sec. 8.7 and finally the fit result can be seen in fig. 9.10.

The found value is compatible with the SM expectation value for the heavy mass eigenstates.

Figure 9.10: Signal sPlot projection from 15/16 data sample. With the red line the best fit signal MC template is shown. The uncertainties on the data points are obtained from the parameterization of the fitter discussed in the previous section.
by \( \sim 1\sigma \) and with the light state lifetime by \( \sim 0.5\sigma \). In addition the returned uncertainty is compatible with the most probable value found in the uncertainty distributions (fig. 8.24) extracted from toys. Given that the measurement is largely statistically limited it's highly important to repeat the measurement once the full Run-2 data set is available.

### 9.6.1 Comparison with other experiments

The result obtained with the use of the 15/16 data set from ATLAS is compatible with the SM. Equivalently the results obtained from the two other experiments are compatible with the SM expectation value. A comparison between the three results can be seen in fig. 9.11.

Between the two general purpose detectors (e.g. ATLAS,CMS) it can be seen that ATLAS has a more stringent uncertainty than CMS, although the two have roughly the same number of events in the signal region. This discrepancy is likely to arise from the fact the S/B ratio for CMS is 1, whereas for ATLAS due to the optimization for the BDT cut the S/B was close to 2.

Since all three results are now available a combination in the assumption of uncorrelated uncertainties, can be performed. The combination of the three experiments uses the combined statistical and systematic uncertainty in quadrature. To combine the three results an iterative method was used that assumes a linear extrapolation of the uncertainty around the minimum of the likelihood of each experiment [167]. The resulted lifetime with the combined uncertainty can be seen in eq. 9.14.

\[
\tau_{B^0_s}^{comb} = 1.922^{+0.274}_{-0.265} \text{ ps} \tag{9.14}
\]

The combined result does not exclude significantly neither the heavy nor the light mass eigenstate lifetime value. The main limitation factor arises from the low number of \( B^0_s \) decays ob-
served. For this reason the statistical uncertainty is the highest factor which can be improved only with the addition of the full Run-2 data from all three experiments. As shown all three experiments perform an extensive study on the systematics indicating a good understanding of them and their limited impact on the final result. The dominant value in the combination is arising from the LHCb result due to the smaller uncertainty and the relatively large spread of the results. For this reason no prediction on NP can be directly made, however a clear indication of more statistics from all major experiments is shown.
CONCLUSIONS

This thesis has two core components both explored as part of the ATLAS collaboration. The first part concerns mostly a technical task in the context of the ATLAS Phase-II upgrade, whereas the second part is dedicated to the first effective lifetime measurement of the $B_s$ di-muon decays in ATLAS with the 15/16 data-set.

In regard of the first part the HTT system has been discussed with its expected layout and functionality for ATLAS during the HL-LHC era. Within this project the test facility at Sussex has been started to test and validate the host boards called TP. To ensure the seamless operation of the test facility the use of the HTT legacy hardware has been made, for which an extensive firmware block has been developed to allow the full functionality of the boards. In addition a set of hardware validation tests has been performed providing useful insight of issues that might appear during the final tests.

The firmware block developed for the legacy hardware, has as a main goal to enable the operation of the AM ASIC's mounted on with the use of the IPbus protocol. During those studies a characterization of the IPbus protocol has been performed providing useful information to the HTT community. The outcome of this measurement set the IPbus as the main monitoring protocol for the whole HTT system.

Furthermore the test-bed setup has been created which provided all the required ingredients to perform the measurement of the AM chip temperature during operation. Such measurement would have been key for an HTT system based on AM ASICs. Finally the legacy hardware has been used to evaluate the performance of a typical ATCA back-plane. The signal integrity across all the accessible high-speed communication lines has been measured.

Regarding the second core part of the thesis the measurement of the $B_s$ effective lifetime has been performed on the 15/16 data sample collected at 13TeV in ATLAS. The measurement uses the sPlot statistical tool to perform a measurement of the background subtracted signal distribution. The fit is performed with the use of lifetime dependent MC histogram templates with a use of the $\chi^2$ based fitting metric.

The analysis is using the $B^+ \rightarrow J/\psi(\rightarrow \mu^+ \mu^-) K^+$ control channel to estimate the main systematic uncertainty. In addition all remaining significant systematic uncertainties have been assessed, obtaining a final result with both the statistical and systematic uncertainty:
\[ \tau_{B_s \to \mu^+ \mu^-} = 1.185^{0.478}_{-0.19} \pm 0.083 \text{ps} \] (9.15)

When compared to the SM prediction for the heavy and light states it can be seen that the value is \( \sim 0.5\sigma \) away from the light state lifetime and \( \sim 1\sigma \) away from the heavy state lifetime. None of the two values can be excluded due to statistical limitation of the analysis, hence proving the importance of increasing the statistics in this kind of flavour physics measurements in ATLAS.
ANALYSIS SAMPLES AND SELECTIONS

The effective lifetime analysis presented in this thesis as well as the branching fraction analysis on which it was based on, were performed on the 15/16 datasets collected by ATLAS. For the development of the analysis procedures, the control channel and the systematic studies a set of MC samples has been required. In the current appendix the list of the MC samples will be presented in addition to the selections applied to both data and MC, for the derivation of the final ntuples. Finally, the weights applied on the MC samples to mimic the data distribution faithfully will be shown, with a brief discussion for their rationale.

A.1 MC samples

To perform the effective lifetime analysis a limited amount of MC samples are being required, since most of the studies concerning the invariant mass shapes and exploring the properties of the data-set happened for the branching fraction analysis. For this reason in this chapter the full list of MC samples is being presented in table A.1. The samples are categorized depending

<table>
<thead>
<tr>
<th>Channel</th>
<th>Type</th>
<th>Events</th>
<th>Generator</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B^0 \rightarrow \mu^+ \mu^-$</td>
<td>exclusive signal</td>
<td>1,000,000</td>
<td>PYTHIA+EvtGen</td>
</tr>
<tr>
<td>$B^0_d \rightarrow \mu^+ \mu^-$</td>
<td>exclusive signal</td>
<td>1,000,000</td>
<td>PYTHIA+EvtGen</td>
</tr>
<tr>
<td>$B^+ \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^+$</td>
<td>exclusive reference</td>
<td>1,997,000</td>
<td>PYTHIA+EvtGen</td>
</tr>
<tr>
<td>$B^- \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^-$</td>
<td>exclusive reference</td>
<td>1,999,500</td>
<td>PYTHIA+EvtGen</td>
</tr>
<tr>
<td>$B^+ \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)\pi^+$</td>
<td>exclusive bkg</td>
<td>498,000</td>
<td>PYTHIA+EvtGen</td>
</tr>
<tr>
<td>$B^- \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)\pi^-$</td>
<td>exclusive bkg</td>
<td>500,000</td>
<td>PYTHIA+EvtGen</td>
</tr>
<tr>
<td>$B^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)\phi(\rightarrow K^+ K^-)$</td>
<td>exclusive control</td>
<td>5,000,000</td>
<td>PYTHIA+Photos</td>
</tr>
<tr>
<td>$B \rightarrow h h'$</td>
<td>exclusive bkg</td>
<td>5,000,000</td>
<td>PYTHIA</td>
</tr>
<tr>
<td>$B^0 \rightarrow K^- \mu^+ \nu_\mu$</td>
<td>exclusive bkg</td>
<td>250,000</td>
<td>PYTHIA+EvtGen</td>
</tr>
<tr>
<td>$B^0 \rightarrow \pi^- \mu^+ \nu_\mu$</td>
<td>exclusive bkg</td>
<td>500,000</td>
<td>PYTHIA+EvtGen</td>
</tr>
<tr>
<td>$A_{h}^b \rightarrow p \mu^- \nu_\mu$</td>
<td>exclusive bkg</td>
<td>250,000</td>
<td>PYTHIA+EvtGen</td>
</tr>
<tr>
<td>$b \bar{b} \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)X$</td>
<td>inclusive bkg</td>
<td>10,000,000</td>
<td>PYTHIA+EvtGen</td>
</tr>
<tr>
<td>$b \bar{b} \rightarrow \mu^+ \mu^- X$</td>
<td>inclusive bkg</td>
<td>650,000,000</td>
<td>PYTHIA+Photos</td>
</tr>
</tbody>
</table>

Table A.1: Simulation samples for the branching fraction signal, reference, control and bkg samples. All the samples have been generated with the use of EvtGen except the $B_s^0 \rightarrow J/\psi \phi$ sample, since it’s shared with other analysis and therefore generated with a flat angular distribution. After generation the SM distributions are recovered with a hit and miss approach.
on the process they are simulating in the branching fraction analysis; thus in the "signal" category belong all the samples simulating the signal process, in the "reference" category all the MC for the reference channel, in the "control" category for the control channel and finally in the "bkg" category the background MC samples. In the last two columns of the table the number of generated decays for each sample is being showed along with the used generators. For all the samples listed in table A.1 the ATLFAST-II detector simulation has been used for the physics object reconstruction, except in the samples used to identify the muon mis-identification probability and the peaking background studies. For those special samples the full detector simulation has been employed.

In the case of the $b\bar{b} \rightarrow \mu^+ \mu^- X$ inclusive background it is known from Run-1 [158] that most of the candidates originate from decays of hadrons produced by $b\bar{b}$ pairs. For this reason the generation of this sample includes only muons originating from the $b\bar{b}$ decay chain. Prompt di-muons from $c\bar{c}$ pairs were not included in the simulation since their contribution would have been removed by the BDT used for the analysis.

In all the generated samples a set of cuts is being applied to select decays that are as closely as possible to the characteristics from the process under analysis. Those cuts are applied at various stages of the ntuple generation and are being discussed in the following section.

### A.2 Event preselection

Initially in all the samples a first set of loose selections are being applied to all the MC samples generated as well as on the data in order to maintain data size under control and a uniform reference selection across all analysis steps. This first set of cuts are normally referred as analysis preselection. In table A.2 the preselection cuts for all the samples listed above are being listed along with an explanation of what each cut concerns.

For the cuts using the B decay vertex (SV) there are two different mass variables that can be used to determine the vertex position. The first available invariant mass variable uses only the ID information to reconstruct the mass of the B meson, whereas the second one adds also the MS information. The first approach has been identified to provide a better resolution than the later in the signal MC and hence is being used to evaluate the properties of the B candidates. For the reference and control channel since the Kaon contains more tracks in the ID the second approach is being used instead.

From the invariant mass cuts applied the regions shown in table A.3 have been used in both the branching fraction and the effective lifetime analysis. Finally, an additional set of cuts is being applied in all channels and samples at the so called ntupling level which are shown in the list below.

- $\Delta R_{\text{flight}} < 1.5$
- $|\alpha_{2D}| < 1.0$
- $L_{xy} > 0$
### A.2 Event preselection

<table>
<thead>
<tr>
<th>Cut</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>combined muons [157]</td>
<td>all</td>
<td>Requirement on muon candidates to have compatible tracks in the ID and MS.</td>
</tr>
<tr>
<td>$p_T(\mu_1) &gt; 6 \text{GeV}, p_T(\mu_2) &gt; 4 \text{GeV}$</td>
<td>all</td>
<td>Transverse momentum of the muon candidates.</td>
</tr>
<tr>
<td>$</td>
<td>\eta(\mu)</td>
<td>&lt; 2.5$</td>
</tr>
<tr>
<td>$p_T^B &gt; 8.0 \text{GeV}$</td>
<td>all</td>
<td>Transverse momentum of the B candidate.</td>
</tr>
<tr>
<td>$</td>
<td>\eta^B</td>
<td>&lt; 2.5$</td>
</tr>
<tr>
<td>$\chi^2_B/NDF &lt; 6$</td>
<td>all</td>
<td>Reduced $\chi^2$ obtained from the vertexing procedure employed to reconstruct the SV.</td>
</tr>
<tr>
<td>$4766 \text{MeV} &lt; m_B &lt; 5966 \text{MeV}$</td>
<td>signal</td>
<td>Invariant mass of the B candidates.</td>
</tr>
<tr>
<td>$4930 \text{MeV} &lt; m_B &lt; 5630 \text{MeV}$</td>
<td>control</td>
<td>Invariant mass of the B candidates.</td>
</tr>
<tr>
<td>loose track quality selection [168]</td>
<td>reference</td>
<td>Quality requirement on the tracks.</td>
</tr>
<tr>
<td>$p_T(K) &gt; 1.0 \text{GeV}$</td>
<td>reference</td>
<td>Transverse momentum of the kaon candidates.</td>
</tr>
<tr>
<td>$</td>
<td>\eta_K</td>
<td>&lt; 2.5$</td>
</tr>
<tr>
<td>$2915 \text{MeV} &lt; m_{J/\psi} &lt; 3275 \text{MeV}$</td>
<td>reference</td>
<td>Invariant mass of the $J/\psi$ candidates.</td>
</tr>
<tr>
<td>$\chi^2_{J/\psi}/NDF &lt; 6$</td>
<td>reference</td>
<td>Reduced $\chi^2$ obtained from the vertexing procedure employed to reconstruct the $J/\psi$ candidate.</td>
</tr>
<tr>
<td>$1005 \text{MeV} &lt; m_\phi &lt; 1035 \text{MeV}$</td>
<td>control</td>
<td>Invariant mass for the $\phi$ candidates.</td>
</tr>
<tr>
<td>$\chi^2_\phi &lt; 10$</td>
<td>control</td>
<td>Reduced $\chi^2$ obtained from the vertexing procedure employed to reconstruct the $\phi$ candidate.</td>
</tr>
</tbody>
</table>

Table A.2: Preselection cuts applied on the signal, reference and control channels. The first column indicates the cut applied on the various objects, the second column the samples that the cut is applied. Finally, the last column provides a short description for the applied cut.

Where $\Delta R_{flight}$ is the three-dimensional opening, defined as $\sqrt{\Delta \phi^2 + \Delta \eta^2}$, between the B candidate reconstructed momentum and the vector between the PV and SV. The $|\alpha_{2D}|$ is the absolute value of the transverse-plane projection, and finally the $L_{xy}$ the projection of the difference between the PV and the SV along the transverse momentum plane of the B candidate.

Although, the selections applied in data and MC are attempted always to be equivalent in analyses, there are still discrepancies between the two samples that might affect the analysis outcomes. To avoid such biases weights can be derived that are correcting for effects that are present on data and not simulated in MC, which can be applied on the MC samples.
A.3 Simulation samples re-weighting

The main sources causing discrepancies between data and MC samples are and imperfect detector and the response of the trigger in simulation. An additional factor arises from the different pile-up profile between data and MC and finally the kinematic differences between the B mesons in data and MC due to the imperfect b quark phase space and hadronization models. Accounting and correcting for such differences is achieved with the use of weights that are mostly extracted with data-driven approaches. The weights are subsequently applied on the MC samples. A list of the weights required for the branching fraction and effective lifetime analyses of the $B_s \rightarrow \mu^+ \mu^-$ decays can be seen below.

- Pile-up weights
- Muon offline efficiency weights
- Muon trigger weights
- Kinematic corrections
  - DDW
  - QLC

For each of the listed weights a short description is provided in the sections below, for explaining the reasons for requiring the weights.

A.3.1 Pile-up re-weighting

Normally when MC samples are generated data taking has not been concluded; thus the information of the pile-up distribution during data taking is not available for the MC samples. Correcting for this missing information in the MC is being performed with the use of the average pile-up value $\langle \mu \rangle$ which accounts for both the in-time pile-up (number of interactions in the same bunch) and the off-time pile-up (interactions from neighbouring bunch crossings). Re-weighting for pile-up in ATLAS is applied in all physics analyses via the PileupReweighting tool which in practice takes the ratio between the average number of pile-up interaction in data and MC and derives the required weights. Additionally, the tool accounts for effects due
to trigger prescaling that affects specifically the $B_s \rightarrow \mu^+ \mu^-$ analysis. Finally, the uncertainty of the pile-up weights is also provided which can be used as a systematic uncertainty in ATLAS analyses.

### A.3.2 Offline efficiency weights

The offline efficiency weights concern all ATLAS analyses that are dealing with muons. This set of weights is applied in every muon in the final state and describes the discrepancies between the detector response for data and MC. The weights include the different working points for the muons\(^1\) and hence a different set is available for each of these categories.

The weights are calculated based on the tag-and-probe method\(^{[163]}\) to calculate the efficiency ratio of muons in data and MC\(^{[157]}\). Well known SM decays like the $J/\psi \rightarrow \mu^+ \mu^-$ and the $Z \rightarrow \mu^+ \mu^-$ are used for the efficiency calculations in different $p_T$ and $\eta$ regions. The method’s tag leg is required to be a muon triggering a single muon trigger and the probe leg a reconstructed muon by a system independent to the tag\(^{[157]}\).

Since, in ATLAS muons are reconstructed using information from three different parts of the detector (ID, calorimeter and MC) three different probes are available. The efficiency of one of the probes is being calculated based on the track and tagging efficiency of the remaining two options, making the final efficiency being obtained from the combination of those.

As for the pile-up weights the corresponding uncertainty is being provided for every weight in order to be included as a systematic uncertainty in the ATLAS analyses.

### A.3.3 Trigger weights

This category of the weights describes the different response of muon triggers in data and in simulation. As for the former two categories the corrections are necessary for all ATLAS analyses dealing with muons.

The method to evaluate the weights is very similar as for the offline efficiencies. A tag-and-probe method is employed with the use of the $J/\psi$ and $Z$ di-muon decays for low and moderate $p_T$ thresholds. For high-$p_T$ values semi-leptonic $t\bar{t}$ and $W+jets$ events are being used with the same procedure. Since, the $B_s \rightarrow \mu^+ \mu^- \mu^-$ analyses described in this thesis are in the low-$p_T$ region the most important contribution arises from the weights calculated with the $J/\psi \rightarrow \mu^+ \mu^-$ decays.

In contrast to the offline efficiency calculation the trigger weights require the estimate of the correlation between the weights of the two muons, since they are expected to fire di-muon triggers. Such, correlations appear mostly when the opening angle of the two muons is very small and hence the two muons are likely to fire the same muon chamber in the detector. For the effective lifetime and branching fraction analysis the opening angle of the $B_s$ di-muon candidates is beyond the level where correlations become relevant and hence not included.

---

\(^1\) ATLAS has a different set of cuts that create the muon objects, depending on the strictness of those cuts different categories (working points) of muons arise (e.g. loose, tight, etc.)
A.3 Simulation samples re-weighting

Similarly to all the up to know discussed weights the uncertainties are being provided to be included as potential systematic sources.

A.3.4 Kinematic weights

The last set of weights required for the analysis concern the discrepancies between the B meson kinematics in data and MC. This category of discrepancies arises from the various set of cuts applied at the generation of the MC samples and the imperfect modelization of the physics processes in simulation. Correcting for those differences gives rise to two categories of weights outlined in the list below, with a short description.

- **Quark Level Corrections (QLC):** Account for the generation biases due to $p_T(B)$ and $\eta(B)$ selections in the MC $b$ quarks.

- **Data-Driven Weights (DDW):** Account for the residual discrepancies between data and MC ($p_T(B), \eta(B)$) mostly due to the $b$ quark phase space and hadronization models, which are not perfectly modelled in the MC generation.

Both weights are calculated in a grid of different $p_T(B)$ and $\eta(B)$ values only for the exclusive MC samples. Applying the weights in combination produces the most consistent and faithful representation of the data samples. In the following two sub-sections a short description for the methods used to derive the weights will be provided. For a more detail description the documentation of the branching fraction analysis need to be checked, since both categories were introduced and studied there [155].

**Quark Level Corrections**

At the process of generating the exclusive signal MC samples a set of kinematic cuts at truth level are being applied in order to enhance the production speed of the samples. Such cuts introduce a bias on the signal kinematics that is being corrected, with the application of the QLC weights. The list of the generation cuts can be seen in table A.4 for the different exclusive samples of the branching fraction analysis.

The calculation of the QLC weights is being made based on $p_T(B), \eta(B)$ bins relative to the signal distribution in an unbiased reference phase space volume common to all the different signal MC [155]. The samples used for the calculation are the following which are samples only at truth level since the bias treated by the QLC is introduced at the same level; hence no detector simulation is required.

- **unbiased MC:** Samples generated with looser generation level cuts on the quarks with respect to the default MC samples of the analysis.
A.3 Simulation samples re-weighting

| Process                                      | $\hat{p}_T \text{min}$ | $|\eta_{b}|$ | $p_T$ |
|----------------------------------------------|-------------------------|-------------|-------|
| $B^0_{s,d} \rightarrow \mu^+\mu^-, B \rightarrow hh'$ | 5GeV                    | 2.6         | 5GeV  |
| $B^+ \rightarrow J/\psi(K/\pi)^\pm$, $B^0 \rightarrow (K/\pi)^-\mu^+\nu_\mu, \Lambda^0_b \rightarrow p\mu^-\bar{\nu}_\mu$ | 7GeV                    | 2.6         | 7GeV  |
| $B^0_s \rightarrow J/\psi\phi$              | 11GeV                   | 2.5         | 9GeV  |

Table A.4: Quark-level cuts employed in the MC exclusive samples generation. With $\hat{p}_T \text{min}$ the lower cut applied on the parton $p_T$ produced in hard scattering at the reference system of the incoming partons. The last two columns indicate the cuts applied on the $p_T$ and $|\eta|$ of the b quarks from which the B meson of interest originates.

- **quark-biased**: Samples with the same quark-level cuts shown in table A.4 as the default MC samples without a set of cuts called Final State (FS) cuts \(^2\).

The calculation of the weights is based on those two samples and is given by eq. A.1,

$$w_{QLC} = v_{FS \text{c}uts}^{quark \text{Bi}ased} \left( \frac{\sigma_{Pythia}^{quark \text{Bi}ased}}{N_{tot}^{quark \text{Bi}ased}} \right) / v_{FS \text{c}uts}^{unbiased} \left( \frac{\sigma_{Pythia}^{unbiased}}{N_{tot}^{unbiased}} \right)$$

(A.1)

where with $v$ the number of entries in a given $p_T(B), \eta(B)$ bin from the unbiased and the quark biased samples after applying the FS cuts is given, with $\sigma_{Pythia}$ the Pythia-calculated generation cross-section and finally with $N_{tot}$ the number of generated events for the given samples. In practice the QLC are not only correcting for the different $p_T(B)$ and $\eta(B)$ discrepancies but also for the correct normalization of the samples due to the generator-level cuts [155]. For the analysis the weight applied in each event is the inverse of eq. A.1, and their uncertainties are provided for additional systematic studies in the analyses.

**Data-Driven Weights**

The last set of weights applied in the analysis MC samples are the DDW which account of residual differences between data and MC. In practice this means that all the derived weights till this point need to be applied on the MC samples before extracting the DDW which would account for any other possible effect.

Calculating the correction is a complicated task and is performed using the MC samples for the reference and control channels as well as the data. To perform the comparison between data and MC a background subtraction needs to be performed in data. Additionally, to avoid dependencies to the specific MC samples that are used only the odd events have been used for the weight extraction in both data and MC whereas the even events have been used as part of the branching fraction reference channel mass fits.

During the calculation of the weights it has been proven that the weights from the reference channel ($B^+ \rightarrow J/\psi K^+$) and the control channel ($B^0_s \rightarrow J/\psi\phi$) have been found to be interchangeable [155]. The resulting discrepancies between data-MC thus are more likely due to

---

\(^2\) Cuts applied on the final state (muons); $p_T > 3.5$GeV $|\eta| < 2.6$; in case of the hadrons the cuts are applied on the kinematic properties $p_T > 0.8$GeV and $|\eta| < 2.6$
kinematic reasons rather than other factors like the differences in fragmentations between $B^+$ and $B^0$.

The method employed for the calculation of the weights is slightly different than the method used for the QLC. Instead of binning in $p_T(B)$ and $\eta(B)$ an iterative approach has been chosen due to the limited statistics of the two samples, which has been performed for the two variables independently as they were two different variables. The calculation is repeated multiple times until the calculated weight is consistent with 1. Additionally, an add-hoc algorithm has been developed to account for the calculation of the uncertainty on the weights including all the potential correlations between the various steps of the iterative approach.

The formula used for calculating the weights can be seen in eq. A.2,

\[
W^1_{pT} = \frac{\sum_\eta D_{\eta(B),pT(B)}}{\sum_\eta D_{\eta(B),pT(B)}} \cdot \frac{\sum_\eta MC_{\eta(B),pT(B)}}{\sum_\eta MC_{\eta(B),pT(B)}}
\]

\[
W^n_{pT} = \frac{\sum_\eta D_{\eta(B),pT(B)}}{\sum_\eta D_{\eta(B),pT(B)}} \cdot \frac{\sum_\eta MC_{\eta(B),pT(B)}}{\sum_\eta MC_{\eta(B),pT(B)}} \prod_{m=1}^{n-1} W^m_{pT} W^m_{\eta}
\]

where the first component is the equation for the first iteration, which allows the extraction of $W^1_{pT}$ weights, and the second component the corresponding $n_{th}$ iteration, for the extraction of $W^n_{pT}$. In eq. A.2 the $MC_{\eta(B),pT(B)}$ refers to a particular $\eta(B),pT(B)$ bin of the MC sample, weighted with all the weights described above in this section, while $D_{\eta(B),pT(B)}$ refers to an $\eta(B),pT(B)$ bin in the data distribution. Acquiring the formulas for the $\eta$ weights an exchange of the indices of $pT$ with $\eta$ is required in eq. A.2.

To calculate the final weight for the DDW the product of the $pT$ and $\eta$ weights after convergence is being taken with the final form seen in eq. A.3.

\[
W_{DDW} = \prod_{m=1}^{m=N} W^m_{pT} \cdot \prod_{m=1}^{m=n} W^m_{\eta}
\]

For a more detailed discussion about the DDW extraction the analysis documents of the branching fraction analysis need to be considered found in ref. [155].
In this appendix section the statistical tools used for the lifetime analysis are being discussed along with their usage in the different cases. The main tool discussed initially concerns the sPlot technique used for the background subtraction in the decay time distribution. A detailed description of the logic, along with the derivation of the required weights for distribution unfolding with sPlot will be provided. Following to that an overview of the goodness-of-fit tools that have been used to evaluate the performance of weighted unbinned maximum likelihood fitters shown throughout the thesis.

B.1 Distribution unfolding method

A common approach in particle physics to extract parameters of interest from a data sample is to use the maximum likelihood fit estimator. In the case of multiple sources of events in a given data sample, the various species are represented with different components in the maximum likelihood estimator shown in eq. B.1,

\[ \mathcal{L} = \sum_{e=1}^{N} \ln \sum_{i=1}^{N_s} N_i f_i(y_e) - \sum_{i=1}^{N_s} N_i \]

where with \( N \) the total number of events is noted, \( N_s \) the number of event categories (e.g. signal, background, etc.), \( N_i \) the number of expected events on average for the \( i^{th} \) species, \( y \) the set of observables and \( f_i \) the theoretical prediction expressed as a PDF for each of the event categories [150].

The main reason for using the logarithm in eq. B.1 is due to the computational advantage for minimizing a function rather than maximizing; hence instead of maximizing the likelihood for the best estimate, the minimization of the log-likelihood is being performed. Identifying the quality of the resulting maximum likelihood fit can become a difficult task. Various approaches can be used, with the most common one described in the next sub section. However, the binned \( \chi^2 \) goodness-of-fit is an approximation for an unbinned fit. An alternative method would have been to use the fit result and based of that unfold the distribution on another variable \( x \) which from now on is going to be called control variable. Obtaining the unfolded control distributions and comparing them with theoretical predictions can indicate whether the estim-
Unfolding such distributions can be achieved directly with cuts applied on the control variable only if there are prominent discrepancies between the different event sources. The main limitation of this approach arises from the reduction on the statistics appearing from the cut based selections, thus making the obtained distributions sub-samples of the original and prompt to statistical fluctuations. Solution to that issue can be achieved with the use of the sPlot formalism, which provides a reliable method for unfolding the distributions of different event categories on the control variables, with the use of the maximum likelihood fit on the observable with the known PDF models, here after called discriminating variable [150].

sPlot is a statistical tool which provides the capability to obtain the truth distribution for a number of control variables $M_n(x)$, which are uncorrelated to the discriminating variable $y$, without requiring any knowledge on the distribution of the control variable. In the next two subsections the formalism of the sPlot method will be discussed. Initially the first step towards sPlot will be discussed, called inPlot, which assumes that $x - y$ are correlated variables. Following the derivation of the sPlot method and the corresponding sWeights, derived to unfold the distributions of the control variable, will be provided. For both formalisms a pair of variables will be assumed, but the whole derivation can be applied easily for any number of control and discriminating variables.

### B.1.1 inPlot

In the case of the total correlation between the discriminating variable ($y$) and the control variable ($x$), a functional form exist which between $x - y$, effectively including the $x$ variable in eq. B.1, and forcing the a-priori knowledge of its PDF. Ignoring that correlation and attempting to infer the truth distributions $M_n(x)$ from the result of the unbinned maximum likelihood fit performed on $y$ with the known PDF for each of the event categories yields to the definition of the per-event weight seen in eq. B.2 [169].

$$P_n(y_e) = \frac{N_n f_n(y_e)}{\sum_{k=1}^{N_n} N_k f_k(y_e)} \quad (B.2)$$

If the weights are applied on the $x$ variable its distribution can be obtained with the definition seen in eq. B.3,

$$N_n \tilde{M}_n(\tilde{x}) \delta x = \sum_{e \in \delta x} P_n(y_e) \quad (B.3)$$

where the sum runs over all the events that enter a histogram bin centered around $\tilde{x}$ and with a width of $\delta x$. In practice the distribution $\tilde{M}_n(\tilde{x})$ is the distribution of the $x$ variable when filled in a histogram with the weight defined in eq. B.2 [150]. If the sum in eq. B.3 is being replaced with an integral it can be proven that $\tilde{M}_n(\tilde{x})$ reproduces on average the $M_n(x)$ truth distribution. This derivation can be seen in eq. B.4.
\[ \langle N_n \tilde{M}_n(\bar{x}) \rangle = \int d y \sum_{j=1}^{N_s} N_j f_j(y) \delta(x(y) - \bar{x}) \mathcal{P}_n(y) \]
\[ = N_n \int d y \delta(x(y) - \bar{x}) f_n(y) \]
\[ = N_n M_n(\bar{x}) \]  

(B.4)

Therefore, the sum of all the weighted events for each of the event categories provide an estimate of the control variable \(x\) distribution \([150]\). Such distributions reconstruct accurately the \(M_n(x)\) truth distribution only if the \(x\) variable is a subset of the \(y\) variable. The major limitation from this procedure arises from the implicit entering of the control variable PDF for each event category in the expression of the weight eq. B.2. However, if there is no correlation between the control and the discriminating variable the a-priori knowledge of the control variable PDF’s is not needed and hence the sPlot technique described in the next section can be used.

### B.1.2 sPlot

In the sPlot method applying directly the weight defined in eq. B.2 cannot be performed, since the PDF’s of the control variable will appear in the derivation seen in eq. ?? but are not present in the likelihood function anymore. This fact gives a rise to a correlation term, seen in eq. B.5, which in the case of total discrimination between \(x - y\) this is equal to covariance matrix between the different event categories.

\[ V_{n j}^{-1} = \frac{\partial^2 (-\mathcal{L})}{\partial N_n \partial N_j} = N_j \int d y \frac{f_n(y) f_j(y)}{\sum_{k=1}^{N_s} N_k f_k(y)} \]  

(B.5)

Replacing the correlation term with the covariance matrix in the derivation of eq. B.4 for all the different species yields to the following determination of the \(M_n(x)\) truth distributions, seen in eq. B.6.

\[ \langle \tilde{M}_n(\bar{x}) \rangle = \sum_{j=1}^{N_s} M_j(\bar{x}) N_j \langle V_{n j}^{-1} \rangle \]  

(B.6)

Inverting the matrix, would allow the definition of the truth \(M_n(\bar{x})\) distributions, which leads subsequently to the re-definition of the weight in eq. B.2 to the so called sWeights seen in eq. B.7.

\[ \mathcal{P}_n(y_e) = \frac{\sum_{j=1}^{N_s} N_j f_j(y_e)}{\sum_{k=1}^{N_s} N_k f_k(y_e)} \]  

(B.7)

Applying the sWeights derived from the result of the maximum likelihood fit on the y-variable to the x-variable unfolds on average for each species of events the truth distribution \(M_n(\bar{x})\), if histograms are filled with their corresponding event weights. The uncertainty in each bin can easily be proven to be the square root of the sum of the squared weights. Additionally, the overall normalization is being maintained, meaning that adding all the sPlot distribution for the various species the initial distribution can be obtained \([150]\).

sPlot is a reliable statistical tool allowing the unfolding of the control variable distributions without making any assumption on the underlying PDF.
B.2 Goodness-of-fit

For all the unbinned maximum likelihood fits the binned $\chi^2$ test with the use of the uncertainty from the data has been used to identify the quality of the fit. The calculation of the $\chi^2$ variable is based on eq. B.8,

$$
\chi^2 = \sum_{i=1}^{N_{\text{bins}}} \frac{(O_i - E_i)^2}{\sigma_i^2}
$$  (B.8)

where $N_{\text{bins}}$ the number of total bins in the histogram, $O_i$ is the expectation value of the data in a given bin, $E_i$ the integral of the PDF in the given bin and $\sigma_i$ the Poisson uncertainty of the data in a given bin. Due to the properties of the $\chi^2$ variable for bins with low statistics the fit quality might be biased. Although the binned $\chi^2$ does not provide a fully accurate result for unbinned fits, it’s been widely used to evaluate the performance of fitters, without affecting the outcomes of both the effective lifetime and branching fraction analysis since all the assumptions are being tested with systematic uncertainties.

In the case of weighted events the usage of eq. B.8 cannot be used directly, since bins with no entries do have a bias in the calculation of the uncertainty. This effect appears in the calculation of the sum as terms with very large values and hence causing the fit quality to be poor. Such, behavior does not compromise the performance of the unbinned maximum likelihood fitter and for this reason a more accurate method known as Pearson $\chi^2$ [163] is being used. The form of the Pearson $\chi^2$ can be seen in eq. B.9,

$$
\chi^2 = \sum_{i=1}^{N_{\text{bins}}} \frac{(O_i - E_i)^2}{E_i}
$$  (B.9)

where $N_{\text{bins}}$ the number of bins in the histogram, $O_i$ and $E_i$ the observed and expected number of events in a bin. The Pearson definition does not suffer from the same issues as eq. B.8, since it evaluates the consistency between the model and the data based on the expected events rather than the observed (e.g. sum of weights).
SUMMARY OF PERSONAL CONTRIBUTIONS

The thesis attempts to present a coherent story of the hardware tests performed for the HTT project at Sussex, along with the first ever measurement of the ATLAS $B_s \rightarrow \mu^+ \mu^-$ effective lifetime. As the author of this thesis my contributions in all three chapter although significant were not the only ones to acquire the presented result. In this appendix a list of my individual contributions will be presented in all the projects, I participates.

HTT contributions

As a member of the HTT community in Sussex I contributed heavily in the organization and operation of the lab facility, which included the legacy hardware used for the HTT studies, the ATCA shelf and the commercial Xilinx evaluation boards. I was the only firmware developer of the group designing most of the firmware blocks discussed in the corresponding chapter, and responsible for the safe operation of the available hardware boards. In more detail.

- Implementation of the IPbus firmware for the Virtex-7 evaluation card.
- Design of the PRM06 firmware block, with either integrating pre-made blocks or developing new ones to accommodate all potential operations.
- Measuring the latency of the IPbus protocol with the test-bed setup.
- Safe power up of ATCA crate and Pulsar-IIb boards.
- Configuring the Xilinx IBERT firmware block for the ATCA back-plane tests.
- Developing the testing firmware for the HTTONline software tests.

Full Run-2 branching fraction contributions

Shortly after joining the Sussex group the branching fraction publication for the 15/16 analysis was published [40], opening the path for the design of the full Run-2 measurement. As a member of the B-Physics group I had a leading role in the following tasks:
• Selection of triggers for the full Run-2 version of the analysis.

• Development of the framework for the MC generation for the studies including the \( \eta \) dependence in the mass fit.

**Effective lifetime measurement contributions**

Finally, again as a member of the B-Physics ATLAS group I was the main analyzer of the effective lifetime measurement for the \( B_s \) dimuon decays on the 15/16 data-set. My contributions in the analysis are summarized in the list below.

• Designing, implementing and validating the analysis framework for the lifetime extraction in both the analysis and control channel.

• Integrating the sPlot method into the framework for background subtraction.

• Identifying best possible parameterization for the fitting metric to the background subtracted decay time distribution.

• Extracting the expected result with the corresponding expected statistical uncertainty from toys.

• Perform the estimation of the main systematics sources with the use of the \( B_s \) channel.

• Extraction of the final lifetime result.
LIST OF ACRONYMS

ATLAS  A Toroidal LHC ApparatuS
ALICE  A Large Ion Collider Experiment
ASCOT  Apparatus with Super Conductor Toroids
ATCA  Advanced Telecommunications Architecture
AM    Associative Memory
AMTP  Associative Memory Track Processor
ASIC  Application Specific Integrated Circuit
BR    Branching Fraction
BSM   Beyond the Standard Model
BDT   Boosted Decision Tree
CMS   Compact Muon Solenoid
CAM   Content-Addressable Memory
CoM   Centre-of-Mass
CMOS  Complementary Metal-Oxide-Semiconductor
CTP   Central Trigger Processor
CSC   Cathode Strip Chambers
CKM   Cabibbo-Kobayashi-Maskawa
CRC   Cycle Redundancy Check
CERN  European Organization for Nuclear Research
DDR   Dual-Data Rate
DCS   Detector Control System
DRP   Dynamic Reconfiguration Port
DDW   Data-Driven Weights
EAGLE Experiment for Accurate Gamma Lepton and Energy measurements
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EF</td>
<td>Event Filter</td>
</tr>
<tr>
<td>ECAL</td>
<td>Electromagnetic Calorimeter</td>
</tr>
<tr>
<td>FPGA</td>
<td>Field Programmable Gate Array</td>
</tr>
<tr>
<td>FEX</td>
<td>Feature EXtractor</td>
</tr>
<tr>
<td>FELIX</td>
<td>Front End Link eXtractor</td>
</tr>
<tr>
<td>FTK</td>
<td>Fast Tracker</td>
</tr>
<tr>
<td>FMC</td>
<td>Field Mezzanine Connector</td>
</tr>
<tr>
<td>FSM</td>
<td>Finite State Machine</td>
</tr>
<tr>
<td>FS</td>
<td>Final State</td>
</tr>
<tr>
<td>GMII</td>
<td>Gigabit Medium Independent Interface</td>
</tr>
<tr>
<td>GIM</td>
<td>Glashow-Iliopoulos-Maiani</td>
</tr>
<tr>
<td>HTT</td>
<td>Hardware Tracking for the Trigger</td>
</tr>
<tr>
<td>HTTIF</td>
<td>HTT Interface</td>
</tr>
<tr>
<td>HBM</td>
<td>High Bandwidth Memory</td>
</tr>
<tr>
<td>HTTSim</td>
<td>HTT Simulation</td>
</tr>
<tr>
<td>HLT</td>
<td>High Level Trigger</td>
</tr>
<tr>
<td>HCAL</td>
<td>Hadronic Calorimeter</td>
</tr>
<tr>
<td>ID</td>
<td>Inner Detector</td>
</tr>
<tr>
<td>IPMI</td>
<td>Intelligent Platform Management Interface</td>
</tr>
<tr>
<td>IPMC</td>
<td>Intelligent Platform Management Controller</td>
</tr>
<tr>
<td>IP</td>
<td>Intellectual Property</td>
</tr>
<tr>
<td>I2C</td>
<td>Inter-Integrated Circuit</td>
</tr>
<tr>
<td>IBERT</td>
<td>Integrated Bit Error Rate</td>
</tr>
<tr>
<td>ITK</td>
<td>Inner Tracker</td>
</tr>
<tr>
<td>IBL</td>
<td>Insertable B-layer</td>
</tr>
<tr>
<td>JTAG</td>
<td>Joint Test Action Group</td>
</tr>
<tr>
<td>KG</td>
<td>Klein-Gordon</td>
</tr>
<tr>
<td>LHCf</td>
<td>Large Hadron Collider forward</td>
</tr>
<tr>
<td>LHCb</td>
<td>Large Hadron Collider beauty</td>
</tr>
<tr>
<td>LHC</td>
<td>Large Hadron Collider</td>
</tr>
<tr>
<td>LEP</td>
<td>Large Electron Positron</td>
</tr>
<tr>
<td>LS</td>
<td>Long Shutdown</td>
</tr>
<tr>
<td>Linac</td>
<td>Linear Accelerator</td>
</tr>
</tbody>
</table>
**List of Acronyms**

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>LH</td>
<td>Left-Handed</td>
</tr>
<tr>
<td>LVDS</td>
<td>Low Voltage Differential Signaling</td>
</tr>
<tr>
<td>LAr</td>
<td>Liquid-Argon</td>
</tr>
<tr>
<td>MDT</td>
<td>Monitored Drift Tubes</td>
</tr>
<tr>
<td>MS</td>
<td>Muon Spectrometer</td>
</tr>
<tr>
<td>MFV</td>
<td>Minimal Flavour Violation</td>
</tr>
<tr>
<td>MC</td>
<td>Monte Carlo</td>
</tr>
<tr>
<td>NP</td>
<td>New Physics</td>
</tr>
<tr>
<td>NSW</td>
<td>New Small Wheel</td>
</tr>
<tr>
<td>PS</td>
<td>Proton Synchrotron</td>
</tr>
<tr>
<td>PRM</td>
<td>Pattern Recognition Mezzanine</td>
</tr>
<tr>
<td>PCB</td>
<td>Printed Circuit Board</td>
</tr>
<tr>
<td>PCIe</td>
<td>Peripheral Component Interconnect Express</td>
</tr>
<tr>
<td>PDF</td>
<td>Probability Density Function</td>
</tr>
<tr>
<td>PV</td>
<td>Primary Vertex</td>
</tr>
<tr>
<td>PDG</td>
<td>Particle Data Group</td>
</tr>
<tr>
<td>RPC</td>
<td>Resistive-Plate Chambers</td>
</tr>
<tr>
<td>QA</td>
<td>Quality &amp; Assurance</td>
</tr>
<tr>
<td>QSFP</td>
<td>Quad Form-Factor Pluggable Transceiver</td>
</tr>
<tr>
<td>QFT</td>
<td>Quantum Field Theory</td>
</tr>
<tr>
<td>QCD</td>
<td>Quantum Chromodynamics</td>
</tr>
<tr>
<td>QED</td>
<td>Quantum Electrodynamics</td>
</tr>
<tr>
<td>QLC</td>
<td>Quark Level Corrections</td>
</tr>
<tr>
<td>RTM</td>
<td>Rear-Transition Module</td>
</tr>
<tr>
<td>RTL</td>
<td>Register-Transfer Level</td>
</tr>
<tr>
<td>RH</td>
<td>Right-Handed</td>
</tr>
<tr>
<td>RMS</td>
<td>Root Mean Square</td>
</tr>
<tr>
<td>SM</td>
<td>Standard Model</td>
</tr>
<tr>
<td>SUSY</td>
<td>Supersymmetry</td>
</tr>
<tr>
<td>SPS</td>
<td>Super Proton Synchrotron</td>
</tr>
<tr>
<td>SSTP</td>
<td>Second Stage Track Processor</td>
</tr>
<tr>
<td>SI</td>
<td>Standard Units</td>
</tr>
<tr>
<td>ShMM</td>
<td>Shelf Manager</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>SGMII</td>
<td>Serial Gigabit Medium Independent Interface</td>
</tr>
<tr>
<td>SoC</td>
<td>System on Chip</td>
</tr>
<tr>
<td>SMA</td>
<td>SubMiniature version A</td>
</tr>
<tr>
<td>SV</td>
<td>Secondary Vertex</td>
</tr>
<tr>
<td>SCT</td>
<td>SemiConductor Tracker</td>
</tr>
<tr>
<td>SSSV</td>
<td>Same-Side Same Vertex</td>
</tr>
<tr>
<td>SB</td>
<td>Side-Bands</td>
</tr>
<tr>
<td>TOTEM</td>
<td>TOTal cross-section, Elastic scattering and diffractive dissociation Measurement</td>
</tr>
<tr>
<td>TCP</td>
<td>Transmission Control Protocol</td>
</tr>
<tr>
<td>TDAQ</td>
<td>Trigger and Data Acquisition System</td>
</tr>
<tr>
<td>TP</td>
<td>Track Processor</td>
</tr>
<tr>
<td>TFM</td>
<td>Track Fitting Mezzanine</td>
</tr>
<tr>
<td>TRT</td>
<td>Transition Radiation Tracker</td>
</tr>
<tr>
<td>TGC</td>
<td>Thin-Gap Chambers</td>
</tr>
<tr>
<td>UDP</td>
<td>User Datagram Protocol</td>
</tr>
<tr>
<td>VHDL</td>
<td>VHSIC Hardware Descriptive Language</td>
</tr>
<tr>
<td>VEV</td>
<td>Vacuum Expectation Value</td>
</tr>
<tr>
<td>XVC</td>
<td>Xilinx Virtual Cable</td>
</tr>
</tbody>
</table>
BIBLIOGRAPHY


[47] T. Aaltonen and et. al., Search for $B^0 \rightarrow \mu^+\mu^-$ and $B^0_d \rightarrow \mu^+\mu^-$ decays with the full CDF Run II data set, Physical Review D 87 no. 7, (2013). http://dx.doi.org/10.1103/PhysRevD.87.072003.


http://dx.doi.org/10.1140/epjc/s10052-015-3602-7.

http://dx.doi.org/10.1016/j.physletb.2018.10.051.


https://doi.org/10.1088/1748-0221/3/08/s08003.


https://doi.org/10.1088/1748-0221/3/08/s08002.

http://cds.cern.ch/record/926196.

https://doi.org/10.1088/1748-0221/3/08/s08007.

[64] W. Herr and B. Muratori, *Concept of luminosity*,
https://cds.cern.ch/record/941318.

[65] J. Jowett and et. al., *The 2015 Heavy-Ion Run of the LHC*,
https://cds.cern.ch/record/2207379.


   www.comtel-online.com.

[119] CMS Collaboration, S. Ajuha and et. al., A Full Mesh ATCA-based General Purpose Data
   https://cds.cern.ch/record/2621635.


[126] D. Magalotti, L. Alunni, N. Biesuz, G. Bilei, S. Citraro, F. Crescioli, L. Fanò, G. Fedi,
   Recognition Mezzanine based on Associative Memory and FPGA technology for Level 1
   Track Triggers for the HL-LHC upgrade, Journal of Instrumentation 11 no. 02, (2016)


   environment friendly low power ROM design on FPGA,


[130] A. Annovi, M. M. Beretta, G. Calderini, F. Crescioli, L. Frontini, V. Liberali, S. Shojaii, and
   A. Stabile, AM06: the Associative Memory chip for the Fast Tracker in the upgraded ATLAS
https://doi.org/10.1088/1748-0221/12/04/c04013.


http://static6.arrow.com/aropdfconversion/f2ce08b0de12d4b2deef5cbecc6bd6ed52561b63/atcaf125-d51.pdf.


https://www.xilinx.com/support/documentation/ip_documentation/gig_ether
net_pcs_pma/v16_0/pg047-gig-eth-pcs-pma.pdf.


https://www.xilinx.com/support/documentation/sw_manu

https://www.xilinx.com/support/documentation/ip_documentation/axis
infrastructure_ip_suite/v1_1/pg085-axi4stream-infrastructure.pdf.

https://www.xilinx.com/support/documentation/ip_documentation/tri
-mode_ethernet_mac/v9_0/pg051-tri-mode-eth-mac.pdf.


[158] A. Collaboration, *Study of the rare decays of $B_s^0$ and $B^0_s$ into muon pairs from data collected during the LHC Run 1 with the ATLAS detector*, 2016.


[162] A. Flavio, *Observation of $B_s^0 \rightarrow \mu^+\mu^-$ and first measurement of its effective lifetime with LHCb Run-2 data*. Observation of $B_s^0 \rightarrow \mu^+\mu^-$ and first measurement of its effective lifetime with LHCb Run-2 data., https://cds.cern.ch/record/2253469.


This thesis was typeset using the \LaTeX\ typesetting system created by Leslie Lamport. The body text size is set to 11 pt with *Utopia Regular* with *Fourier* font, part of \LaTeX\ Live. The bibliography was typeset using the \texttt{ATLAS}-paper style.