Optimisation of the SHiP experimental design

Oliver James Immanuel Lantwin
Imperial College London
Department of Physics

A dissertation submitted to Imperial College London
for the degree of Doctor of Philosophy
The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-NonCommercial 4.0 International. Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.
Abstract

The SHiP experiment is a proposed experiment aiming to search for new super-weakly interacting particles. The concept is based on using a very intense and high energy proton beam at the CERN Super Proton Synchrotron (SPS) which is delivered to the new Beam Dump Facility (BDF), where the experiment will search for New Physics (NP) in a zero background environment.

This thesis describes several studies for the optimisation of this concept, in order to maximise its physics potential. These include studies of a benchmark signal model to understand acceptance effects, studies of the muon induced background using both simulation and a dedicated experiment at the SPS, and the optimisation of the muon shield—a crucial component of SHiP—using machine learning techniques.
Declaration

I declare that this thesis presents my own original research, and that work by others is properly referenced.

Oliver Lantwin
Acknowledgements

I would like to thank my colleagues at the Yandex School of Data Analysis, Imperial College and CERN for their collaboration on the muon shield optimisation, in particular Fedor Ratnikov, Denis Derkach, Andrey Ustyuzhanin, Hans Dijkstra, Mitesh Patel and Andrey Golutvin.

Furthermore, I would like to thank Federico Leo Redi and our student Alexandre Luc Grandchamp for our successful collaboration on the study of the passive shielding; Plamenna Venkova who I worked with closely for the muon $\pi \pi$ background; Konstantinos Petridis for his advice and collaboration on the muon combinatorial background, and Alex Marshall for continuing my work on it; the whole muon flux and charm teams who did impressive work to make the measurements a success.

As well, I would like to extend my thanks to Richard Jacobsson and Nicolà Serra, for many helpful discussions on the $b\bar{c}$ integration and physics studies respectively.

Finally, I would like to thank my family and friends for their support throughout my research and this thesis, but also for making sure I do not lose my balance; all those brave enough to proof-read one of the many drafts; and last but not least Andrey Golutvin for his advice and supervision.
During my studies I have worked on many different aspects of the SHiP experiment. Initially I studied the relationship between the signal acceptance and the experimental geometry to understand what could be gained by an optimised muon shield, as described in Chapter 4.

Following this, I started studying the muon induced backgrounds presented in Chapter 5. First, the occupancy of muons in the Hidden Sector Detector (hsd) was studied in preparation for the optimisation of the muon shield, which necessarily requires quantifying the muon background, and its associated electromagnetic (em) background. It was shown that they could be tackled independently, which allows the optimisation to focus on muons alone. A proof of concept study here showed that the em background could be dealt with by shielding against it, a study which was later continued by Alexandre Grandchamp under Federico Redi’s and my supervision.

After these preparatory studies, the optimisation itself was undertaken, resulting in a method for optimising the muon shield using machine learning techniques, which can and will be used in future in order to update the design in accordance with evolving engineering constraints, but also in a concrete new muon shield design which robustly reduces the muon flux, even with a reduced field compared to the previous design. This work is presented in Chapter 6.

Subsequently, I returned to the study of muon induced backgrounds: The combinatorial background, which I studied together with Konstantinos Petridis and was then completed by Alexander Marshall, and the Deep Inelastic Scattering (dis) backgrounds, which was studied initially together with Plamenna Venkova, before being completed by myself.

Finally, I worked on the measurement of the muon flux (see Chapter 7) and charm cross-section at dedicated experiments using the cern sps, where I was responsible for the online system and data processing.
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.6</td>
<td>Aside: SHiP software</td>
<td>49</td>
</tr>
<tr>
<td>4</td>
<td>Signal studies</td>
<td>51</td>
</tr>
<tr>
<td>4.1</td>
<td>Method</td>
<td>52</td>
</tr>
<tr>
<td>4.1.1</td>
<td>Experimental geometry</td>
<td>53</td>
</tr>
<tr>
<td>4.1.2</td>
<td>Signal production</td>
<td>53</td>
</tr>
<tr>
<td>4.1.3</td>
<td>Reconstruction &amp; selection</td>
<td>54</td>
</tr>
<tr>
<td>4.2</td>
<td>Results</td>
<td>54</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Behaviour for vessel distance relative to target</td>
<td>56</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Comparison of muon shield configurations</td>
<td>58</td>
</tr>
<tr>
<td>4.2.3</td>
<td>Moving vessel front while keeping tracker position constant</td>
<td>61</td>
</tr>
<tr>
<td>4.2.4</td>
<td>Acceptance envelope</td>
<td>61</td>
</tr>
<tr>
<td>4.3</td>
<td>Summary</td>
<td>65</td>
</tr>
<tr>
<td>5</td>
<td>Background studies</td>
<td>67</td>
</tr>
<tr>
<td>5.1</td>
<td>Signal selection</td>
<td>69</td>
</tr>
<tr>
<td>5.2</td>
<td>Muon flux</td>
<td>70</td>
</tr>
<tr>
<td>5.3</td>
<td>Muon combinatorial background</td>
<td>74</td>
</tr>
<tr>
<td>5.4</td>
<td>Muon Deep Inelastic Scattering (dis) background</td>
<td>76</td>
</tr>
<tr>
<td>5.4.1</td>
<td>dis in the vessel</td>
<td>79</td>
</tr>
<tr>
<td>5.4.2</td>
<td>dis in the experimental hall</td>
<td>83</td>
</tr>
<tr>
<td>5.5</td>
<td>Passive shielding of electromagnetic (em) background</td>
<td>88</td>
</tr>
<tr>
<td>5.6</td>
<td>Optimisation of the SHiP vessel shape</td>
<td>91</td>
</tr>
<tr>
<td>5.7</td>
<td>Summary</td>
<td>94</td>
</tr>
<tr>
<td>6</td>
<td>Optimisation of the muon shield</td>
<td>95</td>
</tr>
<tr>
<td>6.1</td>
<td>Loss function</td>
<td>98</td>
</tr>
<tr>
<td>6.2</td>
<td>Parametrisation of the muon shield</td>
<td>103</td>
</tr>
<tr>
<td>6.3</td>
<td>The muon sample</td>
<td>104</td>
</tr>
<tr>
<td>6.4</td>
<td>Bayesian optimisation</td>
<td>108</td>
</tr>
<tr>
<td>6.4.1</td>
<td>An example problem</td>
<td>110</td>
</tr>
<tr>
<td>6.5</td>
<td>Computing architecture</td>
<td>113</td>
</tr>
<tr>
<td>6.6</td>
<td>Optimisation runs</td>
<td>114</td>
</tr>
<tr>
<td>6.6.1</td>
<td>Continuous optimisation</td>
<td>115</td>
</tr>
<tr>
<td>6.6.2</td>
<td>Low-field optimisation</td>
<td>116</td>
</tr>
<tr>
<td>6.7</td>
<td>Results</td>
<td>117</td>
</tr>
</tbody>
</table>
6.8 Conclusion

6.8.1 Future research

7 Measurement of the muon flux

7.1 Design of experiment

7.2 Online and offline computing

7.3 Data taking

7.4 Analysis

7.5 Summary

8 Muon shield technology and prototyping

8.1 Muon shield technology and engineering challenges

8.2 Realistic field maps

8.3 Assembly

8.4 Prototyping and plans

8.5 Implications for the optimisation

9 Conclusion

Bibliography

List of figures

List of tables

List of abbreviations
Chapter 1

Introduction

With the end of the second run of the Large Hadron Collider (LHC), particle physics is in a state of aporia. Following the great success of the Higgs discovery \[1,2\], the LHC placed impressive constraints on the existence of new particles, which were hoped to solve the big open questions of fundamental physics.

Yet the Standard Model (SM) is withstanding scrutiny of more and more precision measurements \[3\]. While there are hints here and there such as the flavour anomalies \[4\] and hints at lepton flavour violation (LFV) \[5–9\], observed at LHCb and Belle, there is no unambiguous indication that more data or slightly higher energies, as discussed for future circular colliders, will necessarily lead to new discoveries.

Nonetheless, there is firm observational evidence —mainly from neutrino physics, cosmology and astrophysics— that there is New Physics (NP), which will most likely require new particles beyond the Standard Model (BSM) to explain:

**Dark Matter:** There are several astrophysical and cosmological phenomena, such as structure formation, weak lensing, rotation curves of galaxies \[10\], and behaviour of galaxy clusters \[11\] that would require either modifications to general relativity, or new massive particles with very weak, or possibly only gravitational, couplings to the SM particles. Some of these phenomena have been known for decades, but still have no confirmed explanation. So far, searches for different dark matter candidates have not observed them, steadily pushing the exclusion limits for the popular models further. In the meantime models such as light dark matter (LDM) are becoming more and more attractive.

**Baryon Asymmetry:** While several manifestations of CP-asymmetry have been established and measured in the SM, they are currently orders of magnitude too small to
explain the observed dominance of matter over anti-matter in the universe. There are several proposed mechanisms, involving new particles interacting with the known \( \text{sm} \) particles, which could lead to more \( \mathcal{CP} \)-asymmetry.

**Neutrino Masses:** The origin of the neutrino masses is not yet known. As the only uncharged fermion in the standard model, there are several mechanisms available to give the neutrinos their masses, and which could also explain their smallness, e.g. the Seesaw-mechanism. Most of these mechanisms require new and potentially observable particles, such as the Heavy Neutral Lepton (\( \text{HNL} \)) of the Neutrino Minimal Standard Model (\( \nu\text{msm} \)) (see Section 2.3.3). However, generally neutrino masses and oscillations could be explained by the addition of two sterile neutrinos of virtually any mass \[12\].

Complementing the standard model of particle physics there is also a standard model of cosmology, the \( \Lambda \text{cdm} \) or concordance model. It is observationally very well-established and requires a cosmological constant \( \Lambda \) (or a mechanism such as dark energy to generate it), as well as cold dark matter (\( \text{cdm} \)) or something fulfilling its role.

It is clear that we have not yet discovered the fundamental origin of these observed phenomena. But where should we look for this new physics? Generally, new physics could be either at higher energies or at weaker couplings than studied so far (or a combination of both). The highest-energy scale we can study in a controlled way is the electroweak scale (\( \mathcal{O}(100 \text{ GeV}) \), the energy scale of the weak vector bosons and the Higgs boson), which is currently under study at the LHC. Beyond it, the only known scale is the Planck scale (\( \mathcal{O}(10^{19} \text{ GeV}) \), the scale at which gravitational effect will become important), but we do not know which—if any—scales fall in between: With the discovery of the Higgs, it has been determined that the \( \text{sm} \) itself is mathematically consistent up to the Planck scale \[13\], so no further energy scales are needed from the mathematical perspective.

Even if the current hints of LFV are confirmed, this does not immediately allow determining the scale of \( \text{np} \). Similarly, Muon \( g - 2 \) \[14\] and other precision experiments cannot unambiguously pinpoint the energy scale of new physics, as the interpretation of their results usually requires assuming either an energy or a coupling.

Facing this potentially vast desert, without any energy scale to target, the age of no-lose theorems or discovery guarantees, such as for the Higgs or something else at the LHC, is over. Experiments have to lead the way into the unknown.

There are several complementary frontiers here that can be explored experimentally:
Precision experiments can infer the existence of new particles by studying flavour physics and processes mediated by loop diagrams. They can be sensitive to contributions from particles much heavier than producible directly, as the mass of the particles running in the loop is unlimited. However, they cannot measure the mass and coupling of new contributions independently.

At the energy frontier, experiments are searching for the direct production and decay of new particles at unexplored energies. These experiments usually use hadron colliders in order to maximise the centre-of-mass energy achievable. However, the downside of this is a very challenging experimental environment, as plenty of soft quantum chromodynamics (QCD) interactions accompany the hard interactions studied. Due to this complex environment, the irreducible backgrounds limit the sensitivity to weakly coupled $np$.

The background environment is much cleaner at lepton colliders, but electron colliders have reduced energy reach and muon colliders are still in the early R&D phase.

To look for very weakly interacting particles there are several strategies. Direct detection experiments such as LZ [*] look for cosmogenic dark matter, in particular for so-called weakly interacting massive particles (WIMPs), interacting in a large fiducial volume in extremely controlled environments.

Alternatively one can attempt to produce new particles in the lab, at the intensity frontier in which case one needs to choose between collider and fixed-target set-ups, each with their own trade-offs.

Fixed target set-ups offer much higher intensities, as the interaction probability with a target is orders of magnitude larger than that with a colliding beam, but have lower centre-of-mass energy. Also, especially with dense targets, the initial state of the interaction is harder to control and thus to determine than with a lepton collider. Nonetheless, with few exceptions most current and proposed intensity frontier experiments are fixed-target experiments, as the increased intensity is crucial to study e.g. neutrinos and other very weakly interacting particles.

[*] Confusingly, at accelerators usually direct detection is taken to mean the observation of either a particle itself or of its final state, such that it can be identified. Inference of new particles via scattering or missing-mass techniques would normally be called indirect. We will use this stricter definition of direct detection throughout the thesis.
The SHiP experiment [16,17] is a proposed zero-background experiment setting out to look for a wide range of new physics models [12] at the intensity frontier in a fixed-target configuration at the CERN SPS.

The main signature looked for is a decay vertex with reconstructed mass and particle identification (PID) from the decay of beyond standard model particles in an evacuated, hermetic decay volume. This signature is a smoking gun for new physics assuming a zero background environment. This makes achieving, and proving that we indeed achieve, a zero background environment the key design challenge of the SHiP experiment.

The active muon shield, which magnetically deflects the muons produced in the beam dump of the high intensity beam away from the detector, is the crucial part of our solution to this challenge. Its performance determines the physics performance of the experiment, as it determines signal acceptance and possible intensity. This thesis focuses on the optimisation of the muon shield, and related studies on both signal and background, as well as on related test beam measurements and prototyping efforts.

In Chapter 2, the theory background and context are reviewed, in particular the so-called portal formalism, which allows categorising the huge variety of hidden sector models. Moreover, some common benchmark models used in the community† to compare the prospects of future experiments are discussed. The \( \nu_{\text{MSM}} \) will then be discussed in more detail as the go-to example for the remainder of the thesis.

Chapter 3 gives an overview of the current experimental layout of the SHiP experiment, highlighting in particular changes relative to the configuration described in the technical proposal (TP) [17].

Chapter 4 describes the acceptance studies performed using the \( \nu_{\text{MSM}} \) as a benchmark model to determine the effects of the muon shield length and decay vessel dimensions on the signal acceptance and the physics performance.

Several background studies are presented in Chapter 5, in particular the muon induced backgrounds, which are crucial to quantify in order to prove that SHiP can achieve a zero background environment, and directly motivate and inform the optimisation of the muon shield. The optimisation of the vessel shape and of the passive shielding is also presented in this Chapter.

† To prepare for the update of the European strategy for particle physics (ESPPU), CERN created the physics beyond colliders study group (PBC) to collect and study proposals for future non-collider experiments at CERN. As part of this study, several benchmark models for BSM physics were defined.
Chapter 6 discusses the muon shield optimisation itself, outlining the challenges, and mathematical and computational tools used and developed, and presents the resulting configuration alongside future plans for further improvement.

In Chapter 7 the measurement of the muon flux is presented, which was performed at the sps in summer 2018. Both the design of the experiment and the analysis, which is ongoing at the time of writing, will be described.

Chapter 8 outlines prototyping plans for the muon shield, which will test manufacturing and assembly techniques available for grain-oriented steel, with a focus on the implications for the future changes to the design of the muon shield and for future optimisation runs.

Finally, in Chapter 9 I will summarise the work done and conclude.
Chapter 2

Theoretical background

“The experiments and theory of the 1960s and 1970s gave us today’s Standard Model [...] a beautiful manuscript with some unfortunate Post-it notes stuck here and there with unanswered questions written on them. The last 40 years of effort has not removed even one of those Post-it notes.”

— Burton Richter 1931–2018 in reference [18]

This chapter reviews firstly, what we know about the Standard Model (sm), and secondly, what we currently know and what we know we do not know about New Physics (np).

As SHiP is searching for bsm physics in a model independent way, it is crucial to understand the sm precisely, as this knowledge is used to suppress all sm backgrounds. Due to the wide range of bsm models studied, this chapter focuses on breadth rather than depth in their discussion. For more detailed discussions of the many models, see reference [12], reference [19], reference [20] and references therein.

2.1 Frontiers of our current knowledge

2.1.1 The Standard Model

The Standard Model (sm) of particle physics was established in the 1960s and 1970s in an interplay of experimental observations and theory development, and unifies
our understanding of the strong and electroweak interactions. It does not attempt to explain gravity.

The SM is formulated as a quantum field theory (QFT) with two important characteristics: it is a gauge theory, i.e. it is assumed that it obeys a specific gauge symmetry group (specifically $SU(3) \times SU(2) \times U(1)$), which fully determines the possible interactions of the fields and it uses spontaneous symmetry breaking via the Brout-Englert-Higgs (BEH)-mechanism [21–23]† to allow for massive vector gauge bosons without violating gauge invariance.

Due to this mathematical structure, the field content alongside its Lagrangian density fully specify this model. The Lagrangian density before spontaneous symmetry breaking is

$$\mathcal{L}_{\text{SM}} = \frac{1}{2} \text{Tr} G_{\mu \nu} G^{\mu \nu} - \frac{1}{2} \text{Tr} W_{\mu \nu} W^{\mu \nu} - \frac{1}{2} \text{Tr} B_{\mu \nu} B^{\mu \nu} + \text{Fermion kinematic terms} + \text{Fermion mass terms} + \text{Higgs terms}$$

where $G$, $W$, and $B$, are the gauge fields; $\Phi$ is the Higgs field with its potential $V(\Phi) = \lambda \left( |\Phi|^2 - \frac{v^2}{2} \right)^2$; $D_\mu$ are the covariant derivatives; $Y$ are the Yukawa couplings; and $f$, $g$ run over the flavours and generations respectively. The fermion fields are $\chi_L^f$ for the left-handed leptons and $e_R^g$ for the right-handed charged leptons; and $q_L^f$ for the left-handed quarks, and $u_R^g$ and $d_R^g$ for the right-handed up- and down-type quarks respectively. This separation into left- and right-handed fields is due to the chiral nature of the weak force.

After spontaneous symmetry breaking, the $B$ and $W$ fields mix and result in the massive $Z$ and the massless $F$ fields via the BEH-mechanism.

*Remarkably, except for $CP$ violation in QCD, all allowed interactions are observed.
† Unfortunately, due to the arbitrary rules of the Nobel prize, the contribution of Brout, Hagen, Guralnik and Kibble are often forgotten.
Considering the simplicity and apparently arbitrary assumption of gauge invariance, it might seem miraculous, but the \text{sm} accurately describes nature, wherever we have been able to test it.

In addition to the apparent arbitrariness, the \text{sm} has 26 free parameters that have to be measured. Many physicists would like to see both of these features explained by a more fundamental theory.

However, while it is mathematically consistent up to the Planck scale, the \text{sm} fails to account for several observed phenomena, which are outlined in Section 2.2.

\subsection*{2.1.2 Lambda cold dark matter (\Lambda\text{CDM})}

Complementing the \text{sm} of particle physics, there also is a cosmological standard model, the so called \Lambda\text{CDM} or \textit{concordance} model. As the name already implies, its main features are a cosmological constant, \(\Lambda\), and cold dark matter (\text{CDM}).

The cosmological constant started as an Einsteinian \textit{fudge}, to generate a steady-state solution from his equations of general relativity, which otherwise would have predicted that the universe contract. However, when it was discovered that the universe is not only not contracting, but expanding [24], and subsequently, that the expansion was accelerating [25], it allowed adjusting general relativity to reflect this.

While it has not yet been identified with a particular (or several) particle(s), dark matter is already a pillar of \text{CDM}, and is key to one of its greatest successes: \text{CDM} explains the observed power-law spectrum of the cosmic micro-wave background with only a few parameters. Additionally, this power-law spectrum gives us information on the relative abundance of dark matter to matter, which is measured to be \(\sim 5.4\), most recently by \textsc{planck} [26].
2.2 Observed beyond the Standard Model (bsm) phenomena

2.2.1 Dark matter

While dark matter is well-established cosmologically, the particle physics implications are less clear. There are many models that could explain dark matter, and several experiments are exploring, so far unsuccessfully, the available parameter space.

When categorising different types of dark matter there is usually the distinction between cold, warm, hot, and non-thermal dark matter. The former three categories all assume that dark matter starts out in thermal equilibrium with the baryonic matter in the universe, before decoupling—the so called freeze-out—from visible matter. The timing of this freeze-out is determined by the coupling and interactions between dark and visible matter. Together with the mass of the dark matter particle(s), they allow simulating the effects at the cosmological scale, and predict the current abundance and energy.

Cold dark matter is non-relativistic today, allowing for the observed formation of structures in the universe, while hot dark matter is still relativistic (e.g. the standard model neutrino), and can only be a small part of the total dark matter abundance in order to be consistent with observations. Warm dark matter finds itself in the grey-zone between the two where it is not yet hot enough to conflict with cosmological requirements. Non-thermal dark matter, such as axionic dark matter, is not produced via the freeze-out mechanism and thus is less stringently constrained by cosmological observations.

The wimp “miracle”, the suggestive coincidence of the predicted abundance of weak-scale cold dark matter and the observed abundance, led to much speculation about its mass and cross-section, and to several experiments to search for these wimps. Alas, the favoured parameter space has now been almost fully ruled out [27].

More recently, astrophysical observations, such as a positron excess at pamela [28] and ams-02, a possible feature at 511 keV in the γ-ray spectrum observed by integral [29], and a possible line at 3.5 keV in the X-ray spectrum observed by the xmm-Newton tele-
scope \[^{30}\] point towards lighter dark matter candidates in the MeV to GeV range. Candidates in this mass range would not have been observable at the current generation of direct detection experiments.

It is not straightforward to fit all cosmological phenomena indicating the presence of dark matter with a single candidate, and no model is consistently favoured by the available evidence as a result. While minimal solutions are popular in physics, considering the complexity of the visible matter of the $\text{sm}$, it seems likely, that the more abundant dark matter could consist of a rich range of particles itself—constituting but a part of a larger Hidden Sector (HS).

### 2.2.2 Neutrino masses

The observation of neutrino oscillations, showed that neutrinos, originally believed to be massless, have a tiny but non-zero mass.

At the inception of the $\text{sm}$ neutrino mass terms were not included, as they were believed to be massless. It is trivial to add right-handed neutrinos, which would be sterile and unobservable, to generate Dirac or Yukawa mass terms for neutrinos. Adding these terms would allow for the observed masses, but they would not justify the vastly different magnitude of the neutrino masses and those of the other $\text{sm}$ fermions. So while possible, this is not a particularly satisfying solution to the problem of neutrino masses.

As neutrinos however are the only uncharged fermions of the $\text{sm}$, other mass terms are possible, as discussed in Section 2.3.3.

From the oscillation experiments and cosmology, there are several constraints on the neutrino masses. Cosmology gives an upper limit on the sum of active neutrino masses in the sub-eV range, and also shows that there are only three active neutrino generations.

From oscillation experiments, two mass splittings are known \[^{3}\]: \(\Delta m^2_\odot \equiv m_2^2 - m_1^2 = (7.53 \pm 0.18) \times 10^{-5} \text{eV}^2\) and \(\Delta m^2_{\text{atm}} \equiv m_3^2 - m_2^2 = (2.51 \pm 0.05) \times 10^{-3} \text{eV}^2\) with \[^{3}\]Unfortunately, the Hitomi satellite, which could have confirmed or ruled out this excess was lost before it could complete its mission.

\[^{3}\]In the inverted hierarchy case, \(\Delta m^2_{\text{atm}} \equiv m_3^2 - m_2^2 = (-2.56 \pm 0.04) \times 10^{-3} \text{eV}^2\)
\( |\Delta m^2_\odot| \ll |\Delta m^2_{\text{atm}}| \). This leads to a degeneracy in the ordering of the masses: if \( m_3 > m_2 \), the neutrino mass hierarchy is said to be normal, and if \( m_3 < m_2 \), it is said to be inverted.

### 2.2.3 Baryon asymmetry of the universe (BAU)

It is generally believed that matter and anti-matter were produced in equal proportions at the Big Bang. However, in the observable universe we only observe baryonic matter. This indicates that an asymmetry in the evolution of matter and anti-matter leads to the observed quantity of matter after all anti-matter has annihilated with matter. We talk about the baryon asymmetry of the universe (BAU), as the matter density is dominated by protons and neutrons.

To explain the observed asymmetry, the Sakharov conditions [31] are required:

1. Baryon number violation,
2. \( C \)- and \( CP \)-violation, and
3. departure from thermal equilibrium.

While \( CP \)-violation has been observed in the weak force in many hadronic processes, it is orders of magnitude too small to account for the observed abundance of matter.

There could be \( CP \)-violation due to neutrino mixing, and the next generation of neutrino experiments may be able to observe it.

Several extensions of the \( sm \) can produce large \( CP \) violation in the neutrino sector, and there are several mechanisms that can generate a baryon asymmetry from an asymmetry in the lepton sector, e.g. via leptogenesis [32].

### 2.3 Categorising the Hidden Sector (HS)

Assuming that \( bsm \) is describable by \( qft \)—while not strictly necessary, outside of the realm of quantum gravity, there are no indications that this is not the case—, any extension of the standard model would entail new Lagrangian terms.\(^*_1\)

---

\(^*_1\)Precision measurements have ruled out the possibility of additional generations, \( i.e. \) the simple addition of new fields without modifying the Lagrangian density, which would also not help to solve any of the problems with the \( sm \).
Table 2.1: Possible portal interactions

<table>
<thead>
<tr>
<th>Portal</th>
<th>Interaction term</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scalar (e.g. dark scalar, dark Higgs)</td>
<td>((H^\dagger H)\phi)</td>
</tr>
<tr>
<td>Vector (e.g. dark photon)</td>
<td>(\epsilon F_{\mu\nu}F'_{\mu\nu})</td>
</tr>
<tr>
<td>Fermion (e.g. Heavy Neutral Lepton (HNL))</td>
<td>(H^\dagger \bar{N}L)</td>
</tr>
<tr>
<td>Axion-like particle (ALP)</td>
<td>(aF_{\mu\nu}F'_{\mu\nu})</td>
</tr>
</tbody>
</table>

These generally can be split into those that involve SM particles (\(\mathcal{L}_{\text{portal}}\)), and those that do not (\(\mathcal{L}_{\text{hs}}\)):

\[
\mathcal{L} = \mathcal{L}_{\text{sm}} + \mathcal{L}_{\text{portal}} + \mathcal{L}_{\text{hs}} \tag{2.2}
\]

The resulting Hidden Sector (HS) of particles that do not directly interact with the SM could be arbitrarily complicated, and may be inaccessible due to the lack of direct interactions.

The portal terms however, which by definition involve the SM fields, have to result in gauge invariant Lagrangian terms, which allows us to limit the possible types of terms using our knowledge of the SM.

While there could well be hidden sector models without any portal terms, these could not explain any of the observed BSM phenomena.

An overview of the possible interaction terms is given in Table 2.1. In addition to the renormalisable vector, scalar and fermion portals, there is also the pseudo-scalar Axion-like particle (ALP) portal. In following, each of these portals will be briefly introduced.

### 2.3.1 Vector Portal

One of the simplest ways to add a hidden sector which interacts with the SM is by introducing an additional \(U(1)\) symmetry, under which hidden sector particles are charged, analogous to quantum electrodynamics (QED). Due to the close analogy to QED, the introduced gauge boson is commonly called dark photon, written \(A'\) with the field strength tensor \(F'\).
To connect the hidden sector to the SM, one can introduce a kinetic mixing term, essentially a cross-term of the $F$ and $F'$ kinetic terms:

$$\epsilon F_{\mu\nu} F'_{\mu\nu},$$  \hspace{1cm} (2.3)

where $\epsilon$ is the coupling coefficient to the SM. Through this term the dark photon inherits the photon couplings to the SM particles, suppressed by $\epsilon$.

By introducing the kinetic mixing term, it is unnecessary to add any new charges to the SM fields.

Dark photons can be produced through the production of high-energy (virtual) photons, e.g. in meson decay, proton bremsstrahlung and many QCD processes. They decay via pair-production; if kinematically allowed, invisibly to hidden sector particles $\chi$, or visibly to SM charged particles (e.g. $\ell^\pm \ell'^\mp$).

Dark photons can play the role of dark matter mediators, particularly for light dark matter, for which $\chi$ would be an ideal candidate. They were also a popular explanation of the muon $g - 2$ anomaly, but the parameter space which could have explained it has since been nearly completely ruled out [33].

### 2.3.2 Scalar Portal

With the discovery of the Higgs boson, it has been shown that scalar bosons exist in nature. It is therefore natural to study the possibility of additional scalar bosons.

The simplest way to extend the SM by addition of light scalar bosons is by adding an interaction term

$$ (H^\dagger H) \phi, $$  \hspace{1cm} (2.4)

where $\phi$ is a new singlet scalar coupling to the square of the Higgs field $H$. In the simplest case, the mixing with the SM Higgs boson is the only interaction, and the coupling coefficient in terms of the mixing angle is called $\sin \theta^2$. Of course, more complicated cases exist.
2.3.3 Fermion Portal

The fermion portal\[\text{[34]}\] naturally arises, when one considers the neutrino mass mechanism.

While it is straightforward to simply add sterile right-handed neutrinos to give the neutrinos Dirac mass terms, there is a second possibility of adding Majorana mass terms \[\text{[34]}\], as the neutrinos are uncharged. As Majorana neutrinos would imply lepton-number violation, they could be part of the solution to the \text{BAU}.

However, one could also add both mass-terms to the neutrinos, which would result in a non-diagonal mass matrix. Diagonalising this matrix would allow suppressing the active neutrino masses, via the \text{see-saw} mechanism \[\text{[35]}\], resulting in light active neutrinos, and at least one, heavy Heavy Neutral Lepton (HNL), which is uncharged under the \text{SM} forces and thus sterile.

Their resulting interaction term with the SM would be

\[H^\dagger NL,\] \hspace{1cm} (2.5)

where \(L\) is the left-handed lepton doublet and \(H\) is the Higgs field.

Due to the mass mixing with the active neutrinos, there is an indirect effective coupling coefficient of \(U_{e,\mu,\tau}^2\) to the charged leptons.

To explain the observed active neutrino masses and the \text{BAU}, at least two HNL are required.

A minimal model that also offers a warm dark matter candidate is the Neutrino Minimal Standard Model (\(\nu\text{MSM}\) \[\text{[36]}\]). It extends the SM by adding three right-handed neutrinos:

- a light \(N_1\) with a mass of \(O(10 \text{ keV})\) and essentially decoupled from the other \(N_i\) is a warm dark matter candidate, while

- two heavy \(N_{2,3}\) (HNLs) with degenerate masses \(O(1 \text{ GeV})\) and weakly coupled to the Standard Model, set the active neutrino masses and create the baryon asymmetry of the universe via leptogenesis.

\[\text{[34]}\] Often called neutrino portal, as a super-weakly interacting portal can only be constructed with a neutral fermion, of which the only option in the SM are the neutrinos.

\[\text{[35]}\] The degeneracy in masses mirrors the pattern of active neutrino masses in the inverted hierarchy case.
The ability to explain dark matter, baryon asymmetry and neutrino masses make this model attractive cosmologically. In turn, if one requires that these problems be solved, an upper and lower limit on the coupling, and a lower limit on the mass can be set, which results in a well-motivated region of parameter space to target experimentally. For a summary of the cosmological constraints, see e.g. reference [37].

In the following chapters, the $\nu$MSM will be used as a benchmark model.

### 2.3.4 Axion-like particle (ALP)

The original QCD axion was first theorised by Peccei and Quinn [38] to explain the absence of $CP$-violation in QCD. This requirement tightly constrains its mass and coupling.

Inspired by this however, one can extend the SM by the addition of an Axion-like particle (ALP). The ALP is the pseudo-Nambu-Goldstone boson (PNGB) of a new spontaneously broken symmetry. If the symmetry is not exact, the ALP is massive, with a mass suppressed by the symmetry breaking scale. This allows straightforwardly accommodating small ALP masses, even if the scale of NP is large.

The ALP, $a$, can couple to the SM via a coupling to two gauge bosons,

$$aF^{\mu\nu}{\tilde{F}^{\mu\nu}} , \quad (2.6)$$

here two photons, but could also couple to two gluons. The ALP portal is a special case of a portal, as this term is non-renormalisable.

Alternatively, ALPs could also directly couple to fermions.

### 2.4 Implications for experimental searches

By construction, portal particles couple very weakly to the SM. In some cases, a large cosmogenic abundance of stable hidden sector particles can be expected, e.g. for dark matter candidates. In this case, zero-background experiments with a large fiducial target, such as current neutrino observatories and direct dark matter detection experiments can be effective. However, the latter are limited by the kinematics of nuclear
recoil: if the dark matter particles are too light ($O(\text{GeV})$ and lighter), scattering of nuclei becomes suppressed.

For all other cases, the new particles have to be produced before they can be detected. For this a large number of interactions needs to be achieved, which is easiest at fixed target experiments.

In addition to scattering, for unstable particles their visible decays can be searched for directly.

One should note, that detection of the particle, either indirectly via its scattering, or directly via its decay, entail an additional suppression factor of the coupling constant compared to indirect searches for missing-mass, which infer the production of undetected particles by using kinematic conservation laws.

In the case of decays, the dominant final states for the decay of hidden sector particles discussed here are tabulated in Table 2.2. Differentiating between these final states using PID allows distinguishing between different hidden sector models, and to reject some SM background candidates. However, as most of these final states are common SM final states, it is crucial to create a zero background environment, to allow studying portal particle decays.

For HNL, the most dominant production and decay modes at SHiP are shown in Figure 2.1. For both beauty and charm mesons the annihilation diagrams dominate the production. The contribution from $B_c$ hadrons has a large theoretical uncertainty, as the fragmentation fraction is not known in the SHiP energy range.

When studying the decay of hidden sector particles we differentiate between partially and fully reconstructed final states. Partially reconstructed final states (e.g.

---

One could argue that reconstructible would be clearer, but reconstructed is the term commonly used.
Theoretical background

\[ W^+ \nu_\mu \mu N_{2,3} \]

\[ B_{(c)}^+ \]

\[ W^+ \nu_\mu \mu N_{2,3} \]

\[ c \rightarrow s \]

\[ D_s^+ \]

\[ u \rightarrow b \]

\[ B_{(c)}^+ \]

\[ \ell^\pm \ell^\mp \nu_\ell \] include particles that cannot be detected, such as neutrinos, whereas for fully reconstructed final states (e.g. $\ell^\pm \pi^\mp$) all final state particles are detectable. As a result, if one wants to study partially reconstructed final states, kinematic cuts need to be looser, as e.g. pointing cannot be required. Additionally, $\text{PID}$ becomes more important to distinguish between background and possible signal, as e.g. lepton number violation can be used to infer the presence of an undetected neutrino.

The topology and kinematics of these final states are very similar between different hidden sector models, as they mostly feature two- and three-body decays of relatively massive long-lived neutral particles, allowing us to concentrate on the general behaviour for detector optimisation.

To understand some of the features of the sensitivity of experiments to these processes, consider the $hnl$ parameter space, shown in Figure 2.3 in the mass-coupling plane.

Current experimental limits are shown in dark-grey, while the big-bang nucleosynthesis (BBN) and Seesaw bounds are theoretical:

The BBN bound is due to the effect $hnl$ could have on the production of light elements in the early universe. Particularly, if the lifetime of the $hnl$ is larger than $O(1 \text{ s})$, they would affect the observed abundance of different light nuclei, resulting in the mass-
dependent lower bound on the coupling. Note, that the curves of constant lifetime run approximately parallel to this bound.

The Seesaw bound is due to the so-called Seesaw relation (see e.g. reference [12])

$$|U_{\mu}| \sim m_\nu / M_N,$$

(2.7)

where $m_\nu$ is the active neutrino mass, which is approximated as $m_\nu \approx \sqrt{\Delta m_{\text{atm}}^2}$, and $M_N$ is the hnl mass.

SHiP is designed to get as close as possible to these bounds for hnl, while also being sensitive to a wide variety of other models.

SHiP searches for hnl produced in a dense target in heavy flavour decays, which then decay to sm particles in the decay vessel, so that these decay products can be detected and measured in the spectrometer. Figure 2.2 shows this conceptually.

The lower boundary of the SHiP sensitivity is mainly determined by the hnl yield, acceptance. To improve it, yield and acceptance need to be maximised.

The main factors responsible for the shape are the different production mechanisms:

Up to about 2 GeV, hnl can be produced in decay of D mesons, which are copiously produced in the beam-dump, giving SHiP its best sensitivity in this range.

Above this threshold, hnl can only be produced in the decay of B mesons, which have a lower yield at SHiP, as the production threshold is comparable to the SHiP centre-of-mass energy.

For hnl lighter than about 500 MeV production in kaon decays would dominate, however the SHiP target is designed to minimise kaon production and reabsorb them before decay. Nonetheless, SHiP nearly reaches the bbn bound for these low masses.
Above the $B$ meson production threshold the next energy of interest for $h\nu\ell$ production would be the $Z$ pole. To target it, one would use an entirely different type of experiment, ideally at a lepton collider.

Because of this, the SHiP beam energy is chosen to be well above the charm and beauty thresholds. Higher energies would not significantly improve the heavy flavour yield while resulting in higher energy muons, which would require a longer muon shield, reducing the acceptance. For a comparison of hypothetical yields if SHiP were to be at other accelerators, see reference [39].

The upper boundary of the SHiP sensitivity region is determined by the $h\nu\ell$ lifetime: above it the $h\nu\ell$ tend to decay before they reach the SHiP detectors, as the expected flight distance is similar to the separation of the detector from the target. In this region, searches at colliders compliment the SHiP sensitivity, as they are sensitive to $h\nu\ell$ that decay much closer to the production point.

By reducing the distance from the target to the decay vessel, SHiP can become sensitive to shorter lifetimes. Moving closer to the target also improves the overall acceptance to $h\nu\ell$ of any lifetime as shown in Chapter 4, which also affects the lower boundary of the SHiP sensitivity curve, as the probability to decay within the decay vessel increases.

While the experimental and theoretical bounds differ for other models, the overall shape, and in particular its lifetime dependence, of the sensitivities is similar.
Figure 2.3: SHiP sensitivity to $hnl$ coupled preferentially only to muon neutrinos. The solid and dot-dashed line indicate the uncertainty over the $B_c$ contribution at SHiP energies. The kaon contribution is not included, so below the kaon threshold the sensitivity is underestimated (dashed line). Experimental limits are shaded in grey, while theoretical limits are indicated in the lighter shaded regions and described in the text. Figure taken from reference [40].
Chapter 3

The SHiP experiment

In 2015 the SHiP collaboration presented its technical proposal (TP). Since then, during the comprehensive design study (CDS), the entire experiment has undergone several rounds of re-optimisation, resulting in a new layout, which we will call CDS configuration here.

This Chapter will briefly review the overall concept for SHiP before describing the new configuration, and differences with respect to the TP configuration, when of interest for this thesis. While the overall layout and concept remains the same, the design of the geometry and the sub-systems have been significantly improved and have matured as result of simulation studies, test beam measurements and prototyping. For a more in-depth discussion of the TP configuration, see reference [17]. The complete current status of the SHiP experimental design is laid out in reference [41], with more detail on the experimental facility available in reference [42].

The changes with respect to the TP discussed here were in part motivated by the signal and background studies presented in Chapter 4 and Chapter 5.

The optimisation of most of the sub-systems depended on the optimisation of the muon shield, as discussed in Chapter 6, as it defines the ultimate envelope available to all systems downstream of it.

3.1 Concept

SHiP is unique among experiments aiming at studying similar physics, in that it can reconstruct the decay vertex, measure the invariant mass of the decayed particle, and
can identify its final state using particle identification (pid), without having to assume pointing or timing, which can instead be used to control the background. Key to this concept are a very high intensity beam dump, and a combination of active and passive shielding, as well as background taggers and pid to reduce the background to zero.

The overall layout of the SHiP experiment can be seen in Figure 3.1. The discussion will proceed along the beamline, starting with the accelerator and target complex, and ending with the spectrometer of the Hidden Sector Detector (hsd).

### 3.2 The SHiP coordinate system

The SHiP coordinate system defines $y$ to be vertical, $z$ to be along the beam-axis and $x$ in the bending plane of the muon shield. Different definitions of the origin are common for $z$: For the signal studies presented in Chapter 4, $z = 0$ is defined as the production point, while usually in the SHiP software framework it is defined as the centre of the decay vessel. $x$ and $y$ are always measured from the beam axis.

### 3.3 Beam Dump Facility (BDF)

The Beam Dump Facility (BDF) encompasses the SHiP beamline and target complex, including the first part of the muon shield. For more detail, please see its cds report [43]. An overview of the BDF is shown in Figure 3.2.

The key challenges it has to address are the maximisation of delivered protons on target, maximising the production of new particles via the most well-motivated production mechanisms, while minimising the production and escape of light hadrons and neutrinos, as well as keeping radiation under control.

From the sps the beam is extracted at 400 GeV using slow extraction to ensure long, uniform spills. This minimises combinatorial background due to pile-up. Over the planned run-time of five years, the beam will deliver at least $2 \times 10^{20}$ protons on target, with $4 \times 10^{13}$ protons on target per 1-second long spill. A new, dedicated 380 m transfer-line transports the beam to the target complex. The intensity of the beam is comparable to that used for the cngs facility, however unprecedented when using slow extraction.
The SHiP experiment

Target/Magnetised hadron absorber
Hidden Sector Detector ($\text{hsd}$)
Active muon shield
Scattering and Neutrino Detector ($\text{snd}$)
Decay volume

$400 \text{ GeV}$ p
$\pi$
$\mu$
hnl

$115 \text{ m}$
$10 \text{ m}$

Figure 3.1: Diagram of the SHiP layout as resulting from the comprehensive design study (cds).
The SHiP target is a water-cooled sandwich of tzm—an alloy of molybdenum and tungsten, optimised to allow for sufficient heat dissipation, as even with the beam spiralling and smearing, the beam power threatens to break the target. The $12X_0$ interaction lengths ($2X_0$ more than in the tp) of the target ensure that the entire beam is stopped, that hadronic showers are contained within, and that light mesons are forced to interact before they can decay leptonically.

As a result the expected yield is about $1.6 \times 10^{18} D$-mesons and $1 \times 10^{14} B$-mesons, which includes an enhancement factor of $\times 2.3$ and $\times 1.7$ respectively to account for the cascade of interactions in the long target.

Finally, the hadron absorber stops all $\nu$ particles apart from muons and neutrinos. Since the tp it has been partially magnetised, so that it already starts to separate the muon polarities, which improves the efficiency of the downstream muon shield significantly, and thus reduces the distance between the target and detectors, which, as shown in Chapter 4, improves the acceptance for hidden sector decays.

While it is not possible to magnetise the entire target region due to radiation protection concerns and its complexity, the hadron absorber after the proximity shielding is magnetised. The exact coil layouts and fields are still being optimised, but for the rest of the design, a uniform field of $1.6 \text{T}$ over $\sim 4 \text{m}$ is conservatively assumed. As the
iron in the hadron absorber is close to saturation, higher fields require a non-linear increase in magnet currents and cooling power.

The current design of the target complex is shown in Figure 3.3.

3.4 SHiP muon shield

For SHiP to succeed, the muon shield is required to reduce the flux of muons in the decay vessel by a factor of at least $O(10^6)$, if not more. However, passive shielding of the muons would require a long distance between the target and the detectors, which would drastically decrease acceptance, and large amounts of dense material, such as tungsten or lead. After early studies for the letter of intent and the rp, it was clear, that an alternative to passive shielding of the muons had to be found.

The resulting concept for the muon shield using electromagnets is shown in Figure 3.4: A first set of magnets tries to sweep out the muons as far as possible to create empty space for a second set of magnets with a flipped magnet polarity. This allows the second set to catch any muons swept back in by the return field of the first magnet set.
Figure 3.4: Illustration of the muon shield concept. The muon shield is shown in a $z$–$x$ section along the bending plane, and the two colours indicate the opposite field directions. Note, that the fields are fully contained in material.

Figure 3.5: A single magnet of the SHiP muon shield. The magnet yoke is shown in blue, green and grey to differentiate the regions with approximately uniform field directions and strengths. The coil is shown in copper and field lines of the magnetic circuit are indicated in red. The stray field outside of the magnet is negligible. Figure 3.4, Figure 3.6a, Figure 3.6b and Figure 3.7 show the muon shield configurations as sections through the magnet yokes in $z$–$x$ and $z$–$y$. 
Each magnet is a dipole electromagnet with a yoke and return yoke, fully containing the field. For the first set of magnets, the coil is wrapped around the central yoke, while for the second set the coils are wrapped around the external return yoke to maximise the field in the parts of the magnet with the highest flux of muons. At the height of the beam, the field is approximately uniformly up or down. In Figure 3.4, Figure 3.6a, Figure 3.6b and Figure 3.7 these two polarities are indicated in green and blue. To illustrate the function of the muon shield, one of these magnets is shown in Figure 3.5. This concept is simple yet effective, and has been used for all further muon shield designs.

3.4.1 Technical proposal (tp) configuration

The first implementation of this concept is the muon shield of the tp (Figure 3.6a). The magnets are warm, i.e. not super-conducting, at 1.8 T. This version of the muon shield was designed by hand to clear space for a cylindrical vessel. Even though the underlying concept is simple and elegant, the tp muon shield design is very long and heavy: The decay vessel cannot start before about 60 m from the target, and the magnets use several thousand tons of iron.

Technical feasibility of constructing such large magnets and space for the coils was not yet considered for this design; only the achievable field was considered.
3.4.2 Configuration resulting from previous optimisation

To improve on the manual design, a second iteration was optimised with Minuit\textsuperscript{44} using toy simulation, as described in reference\textsuperscript{45}, and subsequently implemented in the full SHiP simulation framework. The resulting new baseline configuration (see Figure\textsuperscript{3.6b}) had several interesting features that allow it to use the space more effectively, thus reducing its overall size: The hadron absorber was magnetised with a field of 1.8 T\textsuperscript{∗} equal to that of the rest of the magnets, and the shield was designed with a conical vessel in mind, which reduces the angle that needs to be swept out while retaining most of the acceptance of the cylindrical vessel. Combined these improvements allowed the decay volume to start at about 35 m from the target. Nevertheless, it fulfilled stricter requirements than the \( t_p \) muon shield: In the toy simulation, no muons with \( P > 1 \text{ GeV}/c \) reach the fourth tracking station (\( t_4 \)) of the straw tracker. The weight of the shield was also reduced by \(~40\%\) relative to the \( t_p \) configuration.

The result of this optimisation was further tuned by hand, by splitting a magnet into two to reduce its length for engineering reasons, by clipping the corners of the last magnet to avoid having them scatter muons, and by introducing some steel supports to simulate support structures, as well as raising the floor.

However, some of the assumptions, such as a 1.8 T field in the magnets, and particularly in the hadron absorber were not very conservative, and checks with the full simulation framework (see Section\textsuperscript{3.6}) have since shown, that processes not taken into account in the toy simulation, such as catastrophic energy loss and resulting large-angle scattering, allow muons to slip through the muon shield.

This prompted the re-optimisation with the full simulation framework presented in Chapter\textsuperscript{6}.

3.4.3 Comprehensive design study (\textsc{cDs}) configuration resulting from full optimisation

The configuration resulting from the full optimisation is shown in Figure\textsuperscript{3.7}.

It offers at least the same performance as the result of the previous optimisation, while being \(~25\%\) lighter at the same length. However, most importantly, it is optimised

\textsuperscript{∗} Further studies have shown that a minimum field of 1.6 T is a safer assumption for the hadron absorber.
to only require a field of 1.7 T in the free standing magnets, and only 1.6 T in the hadron absorber. Thus, even with the most conservative estimates of the achievable field, it can achieve the required performance.

For details on the optimisation and its performance, see Chapter 6.

3.5 SHiP detectors

The SHiP experiment comprises two detectors optimised to search for two complementary signatures: the Scattering and Neutrino Detector (snd) is a hybrid detector utilising emulsion cloud chamber technology and electronic detectors looking for scattering of very weakly interacting particles and studying neutrino interactions, while the Hidden Sector Detector (hsd) searches for decay vertices of new particles and is capable of studying their properties.

3.5.1 Scattering and Neutrino Detector (snd)

Conceptually, the Scattering and Neutrino Detector (snd) is very similar to the opera experiment, albeit much smaller, due to its proximity to the target.

Figure 3.7: Diagram of cds muon shield design. As in Figure 3.4, the colours indicate the idealised field direction. Sections in $z-x$ and $z-y$ are shown.
Figure 3.8: Schematic of the Scattering and Neutrino Detector (snd). Figure taken from reference [41]. Note, that the magnet yoke is cut away to show the detector within in the side view.

The snd, shown in Figure 3.8, combines several technologies: Layers of absorber and nuclear emulsion provides precision tracking over a scattering target of $\sim 10$ t, interspersed by fast electronic trackers—the Target Trackers—to allow later matching of the emulsion tracks to events in the muon identification system. For these trackers, scintillating fibres, as used for the LHCb upgrade, and other technologies are being evaluated. The same tracking technology is also employed for the Downstream Tracker of the snd, whose purpose is the momentum measurement of particles leaving the scattering target. RPCs are used for the muon identification system.

While the Tp design still resembled a scaled-down version of opera, the completely re-optimised version for the CDS is diverging from its inspiration. Due to the constraints in length and transverse envelope, as studied in Chapter 5, using the available space efficiently is of utmost importance. Moreover, the addition of light dark matter to the physics case motivated some optimisations to improve the sensitivity without sacrificing performance for neutrino physics.

To solve this design-challenge, the roles of the two magnets of the original design were combined, moving the tracking spectrometers into a new specially designed dipole magnet with a minimum field of 1.2 T, with only the muon identification system
remaining outside of the magnet. To further optimise the design, the last layers of
the muon identification system now act as the upstream veto tagger (uvt) of the hsd,
eliminating the need for a dedicated detector at the entrance of the decay vessel.

Advances in fast automatic scanning, allow regular (biannual) replacement and
scanning of all emulsion films. This enables the use of emulsions in the SHiP envi-
ronment, and search for processes which might not create a signal in the electronic
detectors. Nonetheless, to avoid overexposure of the nuclear films, the rate of low-
momentum muons and associated particles needs to be under control, which motivates
the shielding study presented in Section 5.5.

3.5.2 Hidden Sector Detector (hsd)

While the snd has a very specialised mix of technologies, the hsd is designed around
well-established technologies. It consists of the decay vessel and the hidden sector
spectrometer.

A schematic diagram of the hsd can be seen in Figure 3.9.

The decay vessel is a 50 m long conical frustum, hermetically surrounded by the
surrounding background tagger (sbt) and uvt to tag ionising particles entering the
vessel and the associated interactions. The sbt is provided by either plastic or liquid
scintillators, which are being studied in parallel, while the role of the uvt is now fulfilled
by the muon identification system of the snd. To reduce neutrino interactions to the
required level, the vessel is evacuated to $O$(mbar).

The acceptance of the hsd is defined by the aperture of the spectrometer magnet as
5 m $\times$ 10 m in $x$–$y$. As described in Chapter 6, the muon shield is optimised to sweep
muons out of this acceptance. The residual rate and envelope of the residual muons,
as studied in Section 5.2 in Chapter 5 define the volume available to the decay vessel
and the hsd detectors.

Since the tp, where it was still an elliptical cylinder, the vessel shape has been
optimised in response to the studies presented in Chapter 4 and Chapter 5. The
optimisation of the geometric parameters is described in Section 5.6.

The spectrometer can measure the decay vertex, mass and impact parameter relative
to the target of decaying hs particles accurately. It consists of four straw-tracking
stations ($t_1$–$4$) of very low material budget ($\sim$0.5 % of a radiation length per station)
The SHiP experiment surrounding background tagger (sbt) decay vessel magnet spectrometer /h.scnl Scattering and Neutrino Detector (snd) vacuum upstream veto tagger (uvt) tracking station t/one.taboldstyle tracking station t/two.taboldstyle tracking station t/three.taboldstyle tracking station t/four.taboldstyle timing detector SplitCal ∆x = 5 m(∆y = 10 m) Figure 3.9: Schematic diagram of the Hidden Sector Detector (hsd) (not to scale). For illustration of a signal candidate, an hnl decaying in the vessel is shown in red.
to minimise multiple scattering, two each up- and down-stream of the spectrometer magnet. The magnet has a field integral of 0.5 T m. The baseline design is a warm magnet, but a superconducting option is being considered.

To reduce the number of channels and the material budget, the straw diameter was changed from 10 mm to 20 mm since the TP. The tracking performance is not significantly changed.

A timing detector with $O(100 \text{ ps})$ resolution helps suppressing combinatorial backgrounds.

The electromagnetic calorimeter (ECAL) role has been extended to include reconstruction of the diphoton final state, which would be a smoking gun for ALPS, differentiating them from dark photons, which share the other final states. This improved ECAL, called SplitCal, is segmented longitudinally and includes high-spatial resolution layers, which can determine the photon pointing to $\sim 5 \text{ mrad}$. The segmentation also improves electron-hadron separation, which allows removal of the dedicated TP hadronic calorimeter (HCal) without sacrificing PID performance. Only the absorber of the HCal remains to act as a muon filter.

The experiment is situated in a 120 m long experimental hall, underground for radiation protection. To minimise the scattering of muons or particles produced in the hall from entering the decay volume and mimicking decays of Hidden Sector particles, the hall is 20 m wide throughout, and all infrastructure and support structures are above or below the experiment, to avoid causing interactions with the flux of swept-out muons.

### 3.6 Aside: SHiP software

The SHiP software framework, FairShip, is based on the FairRoot framework, that integrates the ROOT software framework, the PYTHIA6 and the PYTHIA8 event generators, and the GEANT4 toolkit for the simulation of the passage of particles through matter (and other packages not directly relevant for the analyses discussed here).

The simulation has been tuned to available data (see e.g. reference [52]), and a dedicated experiment [53] was run to cross-check the simulated muon spectrum,
whose results are being currently analysed. For more details, see Chapter 7. Cascade production of signal and background in the SHiP target is taken into account and has been implemented in the simulation [54].

The SHiP software framework is still young and constantly changing. For all studies presented here, it had to be modified and extended significantly, which fed back into the framework. Large parts of the current implementation of the SHiP experimental geometry in simulation were updated for or in response to the presented studies.
Chapter 4

Signal studies

At the time of writing SHiP has nearly completed its comprehensive design study (cds), which included a global re-optimisation of the experimental layout. The most essential parts of the experiment to re-optimise were the muon shield and the vessel: The muon shield is crucial for the experimental feasibility and determines the volume that is available to the vessel and detectors, and the minimum possible distance to the target, while the shape and dimensions of the vessel determine the overall acceptance. Thus, before they were optimised, no other sub-systems could be optimised.

In order to optimise the vessel and muon shield to maximise SHiP’s physics performance, it is important to quantify the performance and study the effects of changes to the geometry on it. For SHiP, as a zero background experiment, the sensitivity is proportional to the expected signal yield, so once backgrounds are under control, optimising the signal yield i.e. acceptance becomes important.

As the muon shield defines how close the vessel can be to the target, studying the acceptance as a function of the distance to the target shows how much can be gained by the optimisation of the muon shield.

As a benchmark channel, the hnl of the $\nu_{\text{MSM}}$ are studied here, but the topology and kinematics of other hidden sector decays are very similar. Particularly, if the hidden sector particles are massive relative to the $s_{\text{M}}$ particles they are produced from, they typically have large $p_T$ and as a result a large angular spread around the beam direction. To confirm this, the presented studies can be easily repeated with other signal models implemented in FairShip using the same framework as developed for this study. Previously, studies like this were performed for different channels using a variety of toy simulations, preventing direct comparison.
Here we study the effect of the fiducial volume geometry: the solid angle covered, as well as the distance from the target, and its length. The fiducial volume selection requires that reconstructed vertices are at least 5 cm away from the decay vessel walls, and 20 cm upstream of the first tracking station \( t_1 \). As the kinematic cuts have a very high efficiency on signal by design, and for the Monte Carlo (mc)-truth level studies presented here the kinematics are exactly signal-like, only the fiducial volume cuts are considered for these studies. For a detailed discussion of the selection, please see Section 5.1.

### 4.1 Method

This study considers the relationship of the acceptance to the length of the decay vessel, its distance from the target, and the distance of the tracking stations from the target. To also capture the interdependencies, one would naively expect having to perform a simulation for each combination of these parameters.

In practice however, one can simplify the problem, such that only for the variation of the distance of the tracking stations from the target simulation is needed: As only the reconstruction of signal events needs to be simulated, and their interactions with material are negligible, only the tracking stations are needed in simulation. All other sub-systems, and crucially the vessel itself, are not required, and are not simulated to allow the re-use of simulations to study different vessel geometries. Now, for each distance between trackers and the target a simulation is performed, and the other parameters can be studied during the analysis by defining the fiducial volume, and applying the appropriate fiducial volume cut.

Where initially \( N^3 \) simulations would have been required to perform this study for a grid of \( N \) choices per parameter, only \( N^1 \) simulations need to be performed. Furthermore, the length of the decay vessel and its distance from the target can both be studied with arbitrary granularity, as long as the corresponding distance of the tracking stations to the target was simulated.
4.1.1 Experimental geometry

For this study, which lead to the current CDS vessel design, the TP configuration of the HSD was used. Compared to the configuration shown in Figure 3.9, the only relevant difference is the aperture of the tracking stations, which was elliptical for the TP. While this will change the absolute acceptance, the results for the relative acceptance carry over between different cross-sections, as it will scale the same way for the different lengths in z considered in this study. This scaling behaviour results from the naturally conical acceptance, as illustrated in Figure 4.1. The acceptance envelope is further studied in Section 4.2.4.

4.1.2 Signal production

$N_2 \to \mu \pi$ is considered as a benchmark channel here, as it is the dominant decay channel for the ranges of hnl masses available at SHiP energies and also has the cleanest experimental signature. For the couplings, the default FairShip hnl couplings are used (see. Eqn. (4.1) for the definition of the couplings):

$$U^2_e : U^2_\mu : U^2_\tau = 1 : 16 : 4.2,$$

(4.1)

where the absolute scale is $U^2_e = 0.447 \times 10^{-9}$ and the subscripts indicate the couplings to the different generations of active neutrinos. The mass hierarchy of the active neutrinos is normal in this case. For further information on the hnl simulation procedure in the SHiP software framework, please see reference [40].

Per tracker position $1 \times 10^5$ hnl are produced from one of two production mechanisms: charm and beauty hadron decays. In both cases the cascade production, as described in reference [54], is included. If not otherwise indicated in the legend, plots show results from charm production, which dominates the production at SHiP when kinematically allowed. Several masses are considered for the hnl: 1, 1.6, 2 and 3 GeV/c$^2$. Note that 2 and 3 GeV/c$^2$ hnl can only be produced via the beauty production channels. If not indicated otherwise, results are shown for $1.6$ GeV/c$^2$ hnl. To allow efficient study of the decays of the long-lived hnl, they are simulated with a uniform decay probability over the range 0–300 m from the production point. The proper decay prob-

*Note that these couplings correspond most closely to the physics beyond colliders study group (pbc) benchmark case 7, as defined in reference [20], but are not the same, as the benchmark cases were defined after this study was completed.
ability is accounted for by weighting the events according to the flight time before their decay.

The simulation is performed using FairShip, in particular using Pythia8 for the production and Geant4 for the simulation of the interactions with the detector. As the study is performed in this common framework, other signal channels can be easily substituted.

### 4.1.3 Reconstruction & selection

The simulated events are reconstructed using the standard FairShip routines. Any decay vertex that is reconstructed and within the decay vessel is selected. The effect of other selection cuts is negligible for the simulated signal considered here. The decay vessel imposed by the selection is an elliptical cylinder defined by the tracker dimensions and the length, which can be defined arbitrarily to study different configurations. However, the effective vessel geometry differs from the elliptical cylinder defined by the selection: All reconstructed events lie in a cone within the cylinder (see Figure 4.1 and also Section 4.2.4 and Figure 4.8).

Note, that Figure 4.1 could be misleading to the eye: Due to the elliptical cross-section of the vessel, the projection of the decay vertex distribution may be more complicated than it seems, as the $x$ and $y$ envelopes are not independent.

As we are considering very rare processes, the expected signal rate is well below one event per spill, such that the effect of multiple signal events interfering is negligible.

The number of events reconstructed per configuration is $O(10^3)$, which are then weighted and normalised as described below.

### 4.2 Results

For this study the acceptance $A$ is defined as the scaled number of selected $\nu_{\mu} \pi$ decays:

$$ A \left( N_{2,3} \rightarrow \mu \pi \right) = e^{-<\tau>} \times \frac{\# \text{selected}}{\# \text{simulated}}, $$

(4.2)

†Where reconstructed decay vertices are defined as vertices where both decay products of $N_2 \rightarrow \mu \pi$ have reconstructed tracks passing track quality cuts in the tracking stations.
Figure 4.1: Plot of the reconstructed $\text{hnl}$ vertices. While the selection imposes an elliptic-cylindrical vessel (selection cuts in red), effectively all reconstructed events lie within a cone within the vessel. The vessel length is about 40 m, and the distance from the target is 70 m.
where $\tau$ is the lifetime and the two scale factors are the following:

- The generator simulates a uniform decay probability over the studied region. To incorporate the effects of the lifetime, the events are weighted according to their flight time $t$ before their decay, imposing an exponential-decay probability distribution.

- To facilitate the comparison of different simulations, the number of selected events is normalised by the total number of generated events.

The typical flight distance $c\gamma \tau$ of hnl SHiP is sensitive to is of order tens of metres to tens of kilometres.

Uncertainties shown in the figures and those quoted are purely statistical, calculated from the Poisson sampling uncertainty and then propagated appropriately. No theoretical uncertainty on the lifetime weighting nor any systematic uncertainties are assumed. Additionally, as the simulations are created in $5\text{ m}$ intervals, this interval is taken as an estimate for the uncertainty in the optimal length in Table 4.1.

### 4.2.1 Behaviour for vessel distance relative to target

Figure 4.2 shows the acceptance as a function of the distance of the vessel from the target. The effect of moving the entire vessel is clear: Moving the vessel closer to the target increases the acceptance approximately exponentially. To understand this increase, consider the interplay of geometrical acceptance and lifetime: For a fixed tracker size and fixed vessel length, the solid angle covered decreases as it moves further away from the target. The volume of the acceptance cone covered however grows, as the cone asymptotically becomes a cylinder. However, as the hnl decay vertices are weighted by their decay probability, which is determined by their flight distance, the volume is weighted. The weight integrated over the volume significantly decreases when moving the vessel away from the target, more than balancing the increase in volume covered.

The feasible distances of the vessel from the target, about $30–60\text{ m}$, are in an approximately linear regime, which gives us the rule of thumb that for each metre moved closer to the target there is a gain of about $1\%$ in acceptance.

Note, that there is a difference in the behaviour for hnl produced in charm and beauty decays. This effect can also be seen in Figure 4.5. While this has not been
Figure 4.2: $\text{hnl}$-acceptance for $M_{\text{hnl}} = 1.6 \text{ GeV}/c^2$ relative to the total reconstructed $\text{hnl}$ as a function of the distance of the vessel from the target for beauty and charm production channels. The data follows falling exponentials, as one would expect considering the simple argument considering geometry and lifetime presented in Section 4.2.1.
studied in further detail, it is probably an effect of the larger mass of beauty hadrons, which results in a larger $p_T$. This larger transverse momentum leads to $\text{hnl}$ production and decay at larger angles from the beam direction, increasing the importance of the detector covering a large solid angle. This improves the gain for short distances from the target significantly for $\text{hnl}$ produced via the beauty hadron decays. However, note also, that the number of $\text{hnl}$ in beauty decays is orders of magnitude fewer than that of $\text{hnl}$ produced in charmed meson decays, when both channels are allowed.

### 4.2.2 Comparison of muon shield configurations

Having studied the acceptance behaviour for different distances from the target, the natural follow-up question is how this affects feasible muon shield configurations. In the following, the acceptance of the $\tau\pi$ design and the $c\bar{d}s$ design resulting from the muon shield optimisation will be compared. The two configurations are described in more detail in Section 3.4.

The effect of the difference in length of $25\,\text{m}$ between the two configurations can be seen in Figure 4.5. The gain is between 25–40\% depending on the production
Figure 4.4: \(a\) cds magnet configuration. Compared to Figure 4.3a the length of the muon shield has been decreased by about 25 m. \(b\) The acceptance corresponding to this configuration is highlighted in red (cf. Figure 4.2).

Table 4.1: Gain and optimal lengths for the \(\tau\) and cds shield configurations for different \(hnl\) masses and the production mechanisms described.

<table>
<thead>
<tr>
<th>production</th>
<th>mass [GeV/c(^2)]</th>
<th>gain [%]</th>
<th>optimal length [m]</th>
</tr>
</thead>
<tbody>
<tr>
<td>charm</td>
<td>1.6</td>
<td>0.24 ± 0.01</td>
<td>52 ± 5</td>
</tr>
<tr>
<td>beauty</td>
<td>1.0</td>
<td>0.43 ± 0.01</td>
<td>42 ± 5</td>
</tr>
<tr>
<td></td>
<td>1.6</td>
<td>0.39 ± 0.01</td>
<td>37 ± 5</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>0.38 ± 0.01</td>
<td>37 ± 5</td>
</tr>
<tr>
<td></td>
<td>3.0</td>
<td>0.29 ± 0.01</td>
<td>42 ± 5</td>
</tr>
</tbody>
</table>

mechanism (see Table 4.1 for the full results). In both cases the optimal length of the vessel is roughly the same at 45 m.

A simple geometric argument explains why it should be of no surprise to find that an optimal length exists for a given distance from the target: For a fixed-size tracker the solid angle covered depends on the distance from the target. The second effect in play is that the decay volume has a higher density of \(hnl\) decays the closer it is to the production point, due to the exponential decay with flight time.
Figure 4.5: (a) Acceptance for different vessel lengths compared between the configurations shown in Figure 4.3a and Figure 4.4a. The configuration considered in the tp [17] is marked in red. Note, that there is a slight difference in the gain for hnl produced in beauty production. Please also note, that the relative normalisation of b- and c-produced hnl is arbitrary. (b) Same as above, but using distance from target instead of length as abscissa. Note, that the curves for the two different vessel distances from the target are shifted relative to each other compared to (a).
For different hnl masses the results are similar (see Table 4.1 and Figure 4.6). There is little difference in the gain or in the optimal length. Only considering the different lifetimes of hnl with different masses, the decay probability should decrease less quickly when moving away from the target, and as a result the gain should be less, for lighter hnl. But on the contrary, the gain observed is largest for light hnl. However, this is not surprising as the lifetime argument neglects a more important effect: The $p_T$ of the hnl. As argued above, the importance of a large geometrical acceptance increases for high-$p_T$ hnl as they will have a wider angular distribution. The results in Table 4.1 show that kinematic effects completely dominate the effect of the lifetime difference.

### 4.2.3 Moving vessel front while keeping tracker position constant

To complete the set of possible variations of the vessel configuration, Figure 4.7 shows the acceptance behaviour when moving the front face of the vessel while keeping the distance of the tracker from the target fixed. Unsurprisingly, when moving the front of the vessel backwards, one can only lose decay vertices, i.e. the acceptance will only fall.

### 4.2.4 Acceptance envelope

To decide on the vessel cross-section and shape, it is important to study the acceptance envelope, which defines which volumes contain which percentage of the total acceptance.

The procedure to calculate the acceptance envelope as a function of the distance along the beam axis $z$ is as follows:

1. Each reconstructed vertex has two reconstructed daughter tracks. For each track and for each slice in $z$ find $|x|$ and $|y|$.

2. Discard signal candidates with the daughter track with the largest $d$ until the desired percent level is reached, where $d$ is dependent on the desired vessel cross-section, and is given by:

   - For a rectangular cross-section $d = |x| + a|y|$, where $a$ is the aspect ratio for a vessel.
Figure 4.6: Comparison of behaviour seen in Figure 4.5a for different masses. Due to no phase-space being available for heavy masses in charm-produced $h\nu\ell$, results are shown for beauty-production only. Note, that the scales of the ordinates differ. The optimal lengths and maximum gain are tabulated for all masses in Table 4.1.
Figure 4.7: Effect of movement of the front of the vessel for a fixed tracker distance from the production point. While it is not surprising that having a shorter vessel further away reduces the acceptance, this plot complements Figure 4.5 to show the behaviour of moving either end while keeping the other fixed.
Figure 4.8: Acceptance envelope for $\text{hnl}$ decays in the channels (a) $N_{2,3} \rightarrow \mu\pi$, and (b) $N_{2,3} \rightarrow \mu\mu\nu$ projected on to the $z$-$x$ (blue) and $z$-$y$ (red) planes. Note the clear conical shape of the acceptance envelope. For 100\% some outliers—and the tracks of their daughters, which are being considered here as well—distort the shape, particularly for the partially reconstructed channel. But already for 99\% and 95\% acceptance, the envelopes are nearly perfectly conical. The procedure for making these plots is described in Section 4.2.4.
• For an elliptical vessel \( d = x^2 + (ay)^2 \) where again \( a \) controls the aspect ratio.

3. Assuming a projection onto the \( z-x \) plane or the \( z-y \) plane is desired, plot \( \max(x) \) or \( \max(y) \) of remaining tracks for each \( z \).

For this study the decay modes \( N_{2,3} \rightarrow \mu\pi \) and \( N_{2,3} \rightarrow \mu\mu\nu \) are considered, as the envelope might differ significantly for partially reconstructed decay modes.

The resulting acceptance envelope can be seen in Figure 4.8. It is not surprising that the envelope for the fully reconstructed channel \( N_{2,3} \rightarrow \mu\pi \) in Figure 4.8a is nearly conical: As decay products being reconstructible determines whether a given \( hnl \) falls within the acceptance, the further one recedes from the tracking detector, the smaller the angle of the decay products has to be with respect to the beam direction. If this angle is smaller than the maximum angle, the vertices may lie further away from the beam line without the decay becoming irreconstructible.

For partially reconstructed decays, \textit{a priori} one would expect a more complicated envelope. Whereas for the fully reconstructed decays, the reconstructible decay products of a single vertex fall within a cone pointing back to the target, due to the missing momentum carried away by the neutrino, it does not necessarily point back. For the case of \( N_{2,3} \rightarrow \mu\mu\nu \) this can clearly be seen in the 100% contours in Figure 4.8b. However, as soon as one lowers the desired acceptance to just 99%, the envelope becomes conical again, with essentially the same dimensions as for the fully reconstructed case.

Finally, the envelope has to have the same symmetries of the tracking station and vessel, such that the cross-section of the acceptance cone will be a projection of the tracker. There is no discernible effect of the spectrometer magnet on the acceptance envelope in either case. One would expect that it introduces a small asymmetry in \( |x| \), as it could bend out decay products after the first tracking stations.

Clearly from a signal perspective, the vessel shape optimally would be conical. For a discussion and optimisation of the vessel shape also taking into account the background, please see Section 5.6.

4.3 Summary

As a motivation for the optimisation of the decay vessel geometry and the muon shield, the acceptance, defined by the fraction of reconstructible \( hnl \) candidates contained
in the fiducial volume, was studied. While $\text{hnl}$ of different masses were used as a benchmark using both fully and partially reconstructible decay channels, the results can be generalised to other $\text{hs}$ channels, as they are dominated by kinematic effects. The optimal length of the vessel for different distances from the target was determined. Furthermore, the acceptance as a function of the distance from the target was studied, in order to determine the expected improvement due to an optimised muon shield, which would allow moving the $\text{hsd}$ closer to the target.

The study of the acceptance envelope shows that,—as one might expect,—the tracker can reconstruct $\text{hnl}$ vertices within a cone. This motivated switching to a conical frustum from the cylindrical vessel of the $\text{tp}$. The acceptance depends on the distance from the target approximately linearly in the range of the feasible distances from the target ($\sim 30$–$50$ m), with about $1\%$ gain in acceptance per metre closer to the target.

Even though this study was performed for $\text{hnl}$, the results are also indicative for many other models, as they are similar kinematically. The main noticeable difference would be due to the production mechanisms, as they determine the initial $p_T$, similar to the effect observed for the different masses of $\text{hnl}$. Finally, for some models the lifetimes of the portal particles may be shorter than for $\text{hnl}$ in the parameter space regions of interest, increasing the benefit of a short muon shield further.
Chapter 5

Background studies

To create and prove a zero background environment, the SHiP strategy is to redundantly reject backgrounds. In addition to the safety margin allowed for by the redundancy in case of hardware failure and other unexpected factors, it also allows relaxing the individual requirements, such that the rejection efficiency of other requirements can be studied with data *in situ* in order to prove that there is no residual background.

The main backgrounds at SHiP are shown in Figure 5.1 and can be split into two categories: *muon*-induced backgrounds and *neutrino*-induced backgrounds. *Muon*-induced backgrounds are comprised of the fake signal candidates due to the random combination of these residual muons (Figure 5.1b), and the fake signal due to *dis* interactions of muons in the cavern and material of the experiment (Figure 5.1a). The neutrino background (Figure 5.1c) is beyond the scope of this thesis, but has been shown to be under control in reference [41]. The background due to cosmogenic particles is negligible compared to the beam-induced backgrounds, as demonstrated in the TP [17]. None of the changes for the *cos* should affect the cosmogenic background, making a new study not necessary at this time.

In addition to these backgrounds, the overall muon flux and the resulting occupancy of muons and other particles* in the detectors needs to be studied as an input to the optimisation of the other muon induced backgrounds and the optimisation of all sub-system downstream of the muon shield.

In this chapter, it is shown that all signal candidates due to these muon induced backgrounds can be redundantly rejected. Furthermore, the residual detector occu-

---

*As these particles are dominated by soft electrons and photons, these are commonly referred to as electromagnetic (EM) background, even though there are also some pions and other particles*
Figure 5.1: Schematic of the dominant backgrounds: (a) Background due to muon Deep Inelastic Scattering (DIS) in the walls of the experimental hall or the material surrounding the fiducial volume. (b) Background due to random combinations of residual muons. (c) Background due to neutrino interactions in the fiducial volume, and due to interactions in the material surrounding it (not covered in this thesis).
pancy due to muons and particles produced by muons in the detector is studied, and it is demonstrated that passive shielding can be effective at reducing the rate of particles produced by muons in the detector.

Unless stated otherwise, “muons” are used as a shorthand for both polarities in the following discussions, and all figures and results are shown for the cds configuration of SHiP.

5.1 Signal selection

In addition to the fiducial cut already described in Chapter 4, there are several kinematic cuts common to all physics channels to reject backgrounds. These standard cuts are presented in Table 5.1.

Currently the selection is cut-based, but will eventually be improved using multi-variate classification algorithms such as boosted decision trees (BDTs) or deep neural networks (DNNs). At this stage, the additional intelligibility of the cut-based techniques outweigh the slightly higher rejection efficiency of multi-variate methods. As multi-variate selection would by construction perform at least as well as the cut-based approach, the selection cuts also offer a lower bound on the rejection efficiency in the general case. Note, that the kinematic cuts differ for fully reconstructed and partially reconstructed final states, but are otherwise common to all signal channels. The only exception is the diphoton final state of the alps, which is studied separately, and backgrounds to that channel are outside of the scope of this thesis, as it is a recent addition to our physics programme.
In addition to the described selection, the reconstructed mass distribution of signal candidates would further improve the discrimination between signal and background: Fully reconstructed signal would manifest in a mass peak, while partially reconstructed signal would result in a threshold, and finally the backgrounds would show a smooth distribution without prominent features, with the exception of $V^0$ decays, which would be visible as peaks at known masses.

5.2 Muon flux

While not a physics background in the sense that they could not produce signal candidates, the flux of muons and associated particles through the sensitive detectors and surrounding structures needs to be studied. The muon flux in particular determines the envelope ultimately available to each sub-system downstream of the muon shield, within which the occupancy is manageable. Additionally, the muon flux is an important input to the study of muon induced backgrounds, which are studied in Section 5.3 and Section 5.4.

Quantifying the flux of muons through the sub-systems is a pre-requisite for optimising the shield, and is used in the definition of the loss function for the optimisation, as described in Section 6.1. Similarly, it is an important input to the optimisation studies of all SHiP sub-systems downstream of the muon shield.

The most computationally intensive part of the muon background simulation is the simulation of the target and hadron absorber. As these are optimised as part of the bdf and rarely change, this simulation is performed separately from the rest of SHiP—but also in FairShip—, and the particles surviving the hadron absorber are saved, such that they can then be transported through the rest of the experiment at a later stage. This simulation is called the pre-production and is also used for the muon shield optimisation. To allow for changes in the magnetisation of the hadron absorber, the pre-production is performed without a magnetic field. When the particles from the pre-production are later transported through the full experiment starting in the target, the desired field can be applied in the hadron absorber.

---

The optimisation of the target and the hadron absorber, in addition to the physics performance, is dominated by considerations of radiation protection aspects and target material properties.
In the simulation, rare muon producing processes, such as resonant $e\mu$ production and charm have been artificially enhanced to ensure coverage of all possible muon phenomenologies.

The statistics available for this study as a result of this pre-production is about $1.7 \times 10^7$ muons, which are weighted up to one spill of the sPS. This weighting is required to combine the different muon pre-production samples, and to correctly scale the contribution of the rare muon processes, whose rates are augmented to ensure coverage of rare events. As the overall scale that is weighted to is arbitrary, one spill is chosen. Per one-second spill $\sim 10^{11}$ muons are expected with about $10^7$ spills over the planned run-time of SHiP. As it is not feasible to simulate significantly more muons than currently available, the following studies use several methods to make the most of the available statistics.

Since this study was initially performed, a larger pre-production of $4.96 \times 10^8$ muons have become available, with a larger contribution from artificially enhanced rare processes, which is used for the muon $\pi$S study.

To allow isolation of changes to the geometry, the analysis was calibrated against previous background studies for the $\tau\pi$, and made reproducible, such that, for the same geometry and random number seeds, the results would always be the same. However, as even small changes to the geometry would change the random number sequence completely, resulting in a fluctuation limit, below which the results of changes to the geometry were indistinguishable from statistical fluctuations.

The envelope of the muons swept out by the muon shield can be seen in Figure 5.2. Occupancy maps for all sub-systems were made, to determine the expected rates per channel, and optimise the sub-system dimensions. These occupancy maps include all particles produced by muons on their passage through material. A strategy for shielding this background is investigated in Section 5.5. A selection of occupancy maps is shown in Figure 5.3, showing muons only and all particles.

As different sub-systems updated their design and in response to more global changes in the geometry, or when more detailed information was required for the optimisation of specific sub-systems, this study was repeated.

Particularly to aid the design of the target complex of the $bdf$ facility, several simulations were performed to study the physics requirements, and implications of different engineering options on the SHiP physics performance. For instance, different magneti-
Figure 5.2: Muon envelope for one spill. The histogram is normalised to unity, so each bin is coloured according to the fraction of the total muon flux.
Figure 5.3: Occupancy maps for three SHiP sub-systems to exemplify the evolution of the detector rates throughout the experiment. Note, that these rates are pre-digitisation and before energy cuts, such that the actual rate would be lower. The rates are all within the design specifications for the subsystems. Each detector is shown twice, once including all hits and once only showing muon hits.
sation options for the hadron absorber and the integration between the target complex and the experimental hall were studied.

5.3 Muon combinatorial background

Combinatorial muon pairs, if both muons are missed by the background taggers, are one of the main backgrounds to the search for hidden sector decays.

The main problem for quantifying this backgrounds is that only 29 muons get past the muon shield and into the acceptance in simulation to produce enough hits to be reconstructed as a proper track, with the sample available at the time of this study. As a result, the number of unique combinations of muons is only \( \binom{29}{2} = 406 \). To adequately study the background a much larger sample is needed. To overcome this problem, the available sample is used to seed a toy-mc that is then used to generate muon pairs directly. This method had been used in the past for the study presented in the addendum to the tr \cite{39} and is improved here.

The measured \( P-p_T \) distribution at the tracking stations is fit with a 2D exponential distribution to interpolate the momentum distribution. To generate muons, this fitted function is sampled to obtain a \( P-p_T \) pair and the position at the tracker is independently sampled from the histogram of hits in the tracking stations.

In order to fully determine the muon direction, the remaining missing piece of information is the \( \phi \)-angle. Previously, for the estimation performed for reference \cite{39}, \( \phi \) was assumed to be uniformly distributed.

To test this assumption, a large sample of muons passing the shield was required. For this purpose, a particle gun sample of \( 10^7 \) muons with momenta up to and beyond the kinematic range (see Figure 5.4 for the momentum and transverse momentum distribution) and uniform \( \phi \) was generated. The \( \phi \) distribution at the tracking stations is seen in Figure 5.5. It is far from uniform, confirming that an improved method of estimating \( \phi \) is required: The results of the particle gun simulation were used to create a histogram in \( P-p_T-\phi \), which fully describes the kinematics of muons at the tracking station. To generate muons, as before the momentum is sampled, but instead of sampling \( \phi \) from a uniform distribution, it is sampled from the \( \phi \) distribution in the \( P-p_T \) bin corresponding to the muon’s momentum.

\footnote{\( \phi \) is defined as the angle in the \( x-y \) plane}
Figure 5.4: Artificial distributions of muons simulated in order to study $\phi$-distribution after the muon shield

Figure 5.5: $\phi$-distribution of muons reaching the tracking stations for the sample shown in Figure 5.4
This new sampling method improves the agreement of the pair kinematics with that of the original pairs, as can be seen in Figure 5.6.

As a cross-check and as a worst-case scenario, combinatorial muons from different generated signal events were also studied, to determine the power of the cuts to reject even the most signal-like combinatorial muons.

Since this study was performed, the larger pre-production and muon sample became available. Using this larger sample, the statistics become sufficient to study the background without resorting to toy simulation. The resulting kinematics, compared with the distributions for two-body signal final states are shown in Figure 5.7. However, as the distribution of muons in the detectors is still sparse, further studies are under way to use a generative adversarial network (GAN)-generated sample to improve statistics.

The estimated number of combinatorial track pairs which are reconstructed in the SHiP spectrometer is $8.5 \times 10^{15}$ for $2 \times 10^{20}$ protons on target. The standard kinematic selection cuts and fiducial volume cut reduce this number to $10^{9}$ pairs. Assuming a timing window of $340$ ps, more than three times the resolution of the timing detector (less than $\sim 100$ ps for either technology option), the expected number of background candidates is further reduced by a factor of $\sim 10^{11}$ to $4.2 \times 10^{-2}$ at 90% CL.

As this background requires at least two charged tracks entering the decay vessel, the background detectors can also be used to reject this background, adding a rejection factor of $10^{4}$, where the tagging inefficiency is assumed to be independent of the kinematic selection cuts.

### 5.4 Muon Deep Inelastic Scattering (DIS) background

Together with the neutrino background, the muon DIS background is expected to be the most dangerous. As muons interacting in the vessel wall could, for instance, produce $K_L$, which have the same signature as the signal and are long-lived enough to easily pass fiducial volume cuts, they could easily mimic the signal. The muon DIS background comprises two separate cases: DIS of muons in the material of the vessel and material upstream of it, and DIS of muons in the cavern, downstream to the muon shield and most of the vessel.
Figure 5.6: Comparison of kinematics between simulated muon pairs \textit{(blue)} and muon pairs generated from toy-mc \textit{(red)}.
Figure 5.7: Comparison of reconstructed final state kinematics for hnl signal (orange) and fake signal candidates from the muon combinatorial background (blue) with no cuts applied. Data courtesy of Alexander Marshall.
In the first case (hereafter referred to as the “vessel” case), there is only a small sample of muons that are not deflected by the muon shield available, while most muons are deflected into the cavern downstream of the vessel (the “cavern” case) resulting in a large sample of muons to study this case. The vastly different size of available samples requires separate treatment in the following.

5.4.1 dis in the vessel

As muon dis events happen relatively rarely and only a small sample of muons ever pass the muon shield in simulation, we force the muons to interact via dis as described in following:

1. Muons from the muon background pre-production described in Section 5.2 are transported through the detector using Geant4, just as for the muon occupancy study.

2. The information for all muons creating hits in the sbt or the first tracking station are saved, and their momentum is extrapolated to give the flight path.

3. For each incident muon, $1 \times 10^4$ muon dis events using the muon sample and protons and neutrons at rest (50:50) are generated using Pythia6.

4. The dis interactions are placed according to the local material density between the $z$ position of the start and the end of the decay vessel, and then passed to Geant4 which simulates their interaction with the SHiP detector.

5. The final rate of dis events is computed as:

$$N_{\text{dis}} = N_{\text{spills}} \cdot \sum_{i} \left( w_{\mu} \cdot w_{\text{dis}} \cdot \sigma_{\text{dis}}(E) \right)_{i}, \quad (5.1)$$

where $w_{\mu}$ is the muon weight, coming from the production of the initial proton interactions, needed to obtain the full statistics of muons per spill and $\sigma_{\text{dis}}$ is the cross-section, taken from Pythia6. The weight of each dis event is calculated as $w_{\text{dis}} = \rho \cdot L$, where $\rho$ is density along the muon trajectory and $L$ is the muon track length.

---

§The successor, Pythia8, does not (yet) implement muon dis scattering, but is used for many other purposes in the SHiP software framework.
Figure 5.8: Momentum versus transverse momentum of muons hitting the decay vessel. A $3 \text{ GeV}/c$ momentum cut is applied.

The number of muons impinging on the vessel is $5.8 \times 10^4$ per spill. Their momentum and transverse momentum is shown in Figure 5.8. They have an average momentum of $\sim 8 \text{ GeV}/c$, resulting in $2.1 \times 10^8$ muon $\text{dis}$ interactions in the tank over the run-time of the experiment for $2 \times 10^{20}$ protons on target, taking into account the encountered density and flight distance of each muon.

The expected total number of $\text{dis}$ events is calculable as

$$N_{\text{dis}} = N_{\text{spills}} \cdot N_\mu \cdot \langle \rho \cdot L \rangle \cdot \sigma_{\text{dis}} \left( \langle E \rangle \right),$$

(5.2)

where $N_\mu$ is the number of muons impinging on the decay vessel, $N_{\text{spills}}$ is the number of spills over five years of run-time, $\sigma_{\text{dis}}$ gives the cross-section for $\text{dis}$ of muons on protons using analytic calculation, taken from reference [55], $\rho$ is the mean mass density along the muon track, $L$ is the muon track length, and $E$ is the muon energy. Average $\rho \cdot L$ and $E$ are taken from FairShip simulation.

A sample corresponding to two years of SHiP data-taking is generated using the procedure described. As the material density along the flight path is taken into account, most of the $\text{dis}$ interactions occur in the vessel walls or the sbr. The interaction vertices are shown in Figure 5.9.
Figure 5.9: The interaction points of all generated ππ events before any cuts are applied, weighted according to the interaction probability.
Table 5.2: Effect of applying the different selection cuts independently on the muon inelastic background events. Numbers correspond to $2 \times 10^{20}$ protons on target. The veto systems comprise the uvt and the sbt. As no events remain after application of the veto requirement, the UL(90%CL) is given instead, as calculated in equation (5.3).

<table>
<thead>
<tr>
<th>Selection cut</th>
<th>Selected events</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Events reconstructed</td>
<td>$1.5 \times 10^6$</td>
<td>1%</td>
</tr>
<tr>
<td>Fiducial Volume cut</td>
<td>$1.78 \times 10^5$</td>
<td>12%</td>
</tr>
<tr>
<td>DOCA $&lt; 1$ cm</td>
<td>$3.7 \times 10^5$</td>
<td>25%</td>
</tr>
<tr>
<td>$ip &lt; 10$ cm</td>
<td>$2.6 \times 10^4$</td>
<td>2%</td>
</tr>
<tr>
<td>$ip &lt; 250$ cm</td>
<td>$4.2 \times 10^5$</td>
<td>29%</td>
</tr>
<tr>
<td>Not vetoed by the sbt</td>
<td>$&lt; 1.7$</td>
<td>0.0001%</td>
</tr>
<tr>
<td>Not vetoed by any veto system</td>
<td>$&lt; 1.7$</td>
<td>0.0001%</td>
</tr>
</tbody>
</table>

To these dis events the standard selection cuts as described in Section 5.1 are each applied on their own. The effect of the different cuts on the background are summarised in Table 5.2.

As the vessel in the cds configuration is now projective, muons traverse more material on average and interaction-products have a higher chance to re-interact hadronically in the vessel compared to the study performed for the tp presented in reference [39]. This increases the probability to reconstruct an hnl candidate, but also leads to a higher multiplicity, almost always leaving a signal in the sbt. The requirement that the sbt is not fired reduces the background to zero. Hence, we can place an upper limit of

$$UL(90\% CL) < -\ln(0.1) \cdot \frac{\langle w_{dis} \rangle \langle \sigma_{dis} \rangle}{w} = 1.7$$

(5.3)

for $2 \times 10^{20}$ protons on target at 90% CL, where $w$ is the scale factor between simulated sample and total number of expected dis events as calculated in equation (5.2), and the average $w_{dis}$ and $\sigma_{dis}$ are taken from the simulated sample.

The residual numbers of candidates after applying all kinematic selection cuts in order is shown in Table 5.3. None of the events passing the cuts have two leptons in the final state, allowing a further reduction of the number of partially reconstructed hnl candidates to zero using the pid. Assuming that the veto efficiency of the incoming muon and decay products is approximately independent of the ip cut, we can apply
Table 5.3: Summary of the muon inelastic background event after applying the offline selections cuts to search for partially and fully reconstructed \( h_{\text{nl}} \) candidates. Numbers correspond to \( 2 \times 10^{20} \) protons on target.

<table>
<thead>
<tr>
<th>Selection cuts</th>
<th># Candidates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fiducial Volume cut</td>
<td>( 1.78 \times 10^5 )</td>
</tr>
<tr>
<td>( \text{DOCA} &lt; 1 \text{ cm} )</td>
<td>3707</td>
</tr>
<tr>
<td>( \text{IP} &lt; 10 \text{ cm} ) (( \text{IP} &lt; 250 \text{ cm} ))</td>
<td>27 (566)</td>
</tr>
</tbody>
</table>


the veto efficiency to the residual events to arrive at an upper limit on the residual background of \( \text{UL}(90\%\text{CL}) < 2.7 \times 10^{-5} \) for fully reconstructed, and \( \text{UL}(90\%\text{CL}) < 6 \times 10^{-4} \) for partially reconstructed events.

5.4.2 dis in the experimental hall

In the case of the cavern walls \( \mathcal{O}(10^8) \) muons are available, as the vast majority of muons is deflected into the cavern walls. Fully simulating the inelastic scattering for this sample would be prohibitively computationally expensive. However, the naïve expectation for the background from the cavern wall is that it should be negligible compared to that from the vessel, as the geometry of the situation makes it highly unlikely for any interaction products to scatter back into the decay vessel and decay in such a way as to pass selection cuts. This is because the majority of muons hit the cavern walls downstream of the tracker, and the cavern walls are 10 m away from the beam-axis in the \( x \) direction, i.e. the bending plane of the muon shield. However, as there will be a very high flux of muons, this should be confirmed. To prove that this background is indeed negligible, an order of magnitude estimate is performed based on factorising the problem, and estimating the individual components using simulation and geometric arguments.

The expected number of incident muons with \( P \geq 3 \text{ GeV}/c \) in cavern over 5 years (see Figure 5.10) is \( 10^{15} \). Focusing on muons that penetrate less than 1 m of concrete before interacting inelastically, we expect \( \leq 5 \times 10^{10} \) muon dis events over 5 years using equation (5.2). As the worst case it is assumed that all decay products leave the concrete without being absorbed or scattered further.

As a proxy for the kinematics of these events, the dis events generated with Pythia 6 for the vessel case (\( 4.45 \times 10^6 \) \( \mu p \) and \( \mu n \) interactions available covering entire momen-
tum range seen in the cavern) are studied to determine typical momenta, angles and γcτ of decay products, and to estimate the resulting background.

**Geometric acceptance:**

To determine the overall geometric acceptance of dis interaction products to enter the vessel, the angle relative to and the angle orthogonal to the muon direction are studied. The angle between the vessel and the incident muon can be defined in two ways, as shown on Figure 5.11. The smaller of these angles is the minimum scattering angle needed to reach the vessel. The distribution and cumulative distribution of this minimum angle in the studied sample are shown in Figure 5.12. While some only need 0.6 rad relative to the muon, the mode of the distribution is at 2.9 rad, resulting in a large angular scattering needed for particles to reach the vessel.

Orthogonal to the muon direction the vessel covers $O(10\%)$ of the possible scattering angles, which, due to rotational symmetry, are uniformly distributed.

**Decay in vessel:**

Out of the particles that can be produced by muons in the cavern walls, $K_L$ is the only $V_0$ particle with sufficient $c\tau$ to reach the vessel. 0.2% of dis events or $1.2 \times 10^4$ in total in this sample (and about $1 \times 10^8$ over 5 years) have $K_L$ with large enough angle to reach vessel, none of which have sufficient flight distance, once the γ-factor is taken into account.
Figure 5.11: Sketch of the geometry for muon interaction products scattering back to the vessel, where \( \alpha_i \) define the minimum scattering angles in the \( z-x \) plane. Not to scale.

Figure 5.12: Minimum angle required to reach vessel, i.e. the minimum angle in the \( x-z \) plane between the muon direction and the vessel, as shown in Figure 5.11, showing both the distribution (blue) and the cumulative distribution (red).
But let us assume some do. To study this case, a phase space of $10^6 K_L$ is generated with most favourable kinematics; the highest observed momentum with 3 GeV/c, and the smallest possible angle with respect to the direction of the original muon geometrically possible in order to reach the vessel, such that the angle of the $K_L$ relative to the acceptance of the tracking station is also as small as possible (but is still relatively large due to the initial backward scattering).

The $K_L \to \pi \ell \nu$ (combined $\sim 70\%$ BR) decay modes are studied as proxy for possible three-body decays\footnote{as the $K_L$ fly backwards, two-body decays cannot result in a reconstructible vertex.}, where $\pi \ell$ is treated as the signal candidate and loose kinematic selection cuts are applied:

The standard momentum cut on candidate tracks requiring that $P_\pi$ and $P_\ell > 1$ GeV/c rejects 75\% of candidates. As the $K_L$ are not fully simulated and thus no vertex location is available, the usual $p_\text{T}$ cut cannot be applied. Instead, as a lower bound on the rejection efficiency, it is only required that the candidate pass the $p_\text{T}$ cut for partially reconstructed signal candidates if placed anywhere in the vessel. None of the candidates fulfil this requirement for any location in the vessel for any combination of momentum and angle considered.

If other initial momenta or angles are chosen for the generated $K_L$, the prospects for background candidates are even worse: For smaller momenta the daughter momentum cut becomes even more effective at rejecting signal candidates, whereas for larger angles relative to the muon direction, the flight direction of daughters becomes sufficient to reject all candidates, as they tend to fly off to $-z$.

Note, that it is not needed to require that the track is in the spectrometer acceptance, and all other kinematic cuts are still available for further redundancy in the rejection.

It is concluded that the decay of $V_0$ particles produced in the cavern walls is a completely negligible background, as it is straightforward to reject even under very generous assumptions.

**Interaction in vessel wall:**

On average, 0.025 particles per dis event (for $1.25 \times 10^9$ in total over 5 years) have a large enough angle with respect to the initial muon, and large enough $\gamma c \tau$ to reach the vessel.
The interaction probability depends on the type of particle, material budget and other factors. But regardless of interaction, a simple energy consideration shows that they cannot create any signal candidates: As seen in Figure 5.13, all the particles observed have energies below $1.6 \text{ GeV}/c^2$, which, due to energy and momentum conservation, cannot result in at least two $1 \text{ GeV}/c$ particles entering the tracking stations, as would be required by the standard kinematic cut on track momentum.

Even assuming that some rare higher energy particles that make their way to the vessel, while the possibility of recoil against a nucleus could reduce the efficiency of the $\ip$ requirement compared to the decay case, recoiling nuclei or initial charged particles could be tagged by the $\sst$, and the fiducial cut would be very effective against interactions in the vessel walls.

Finally, for occupancy and combinatorial background these events are of no concern, as their rate of $1.25 \times 10^9$ over 5 years is much smaller than that of the direct muon flux.

The conclusion is that this type of event is also a negligible background.
5.5 Passive shielding of electromagnetic (EM) background

The distribution of electromagnetic (EM) background in the different sub-systems is discussed in Section 5.2 and shown e.g. in Figure 5.3. Outside of the hotspots due to muons passing through material, the EM background seems nearly uniformly distributed. To study the origin of these particles and to determine whether the SND detector mainly shields this background, or whether this effect is dominated by the new particles produced by muons passing through it, a muon background simulation as described in Section 5.2 was performed without the presence of the SND

This question is of particular interest to understand whether the EM background needs to be considered during the muon shield optimisation, or whether additional passive shielding can be employed to reduce the flux as needed once the muons are swept out.

As can be seen in Figure 5.14, the SND clearly has a shielding effect, showing that in the regions not traversed by muons shielding can be employed. Encouraged by this proof-of-concept result a closer study was initiated in order to optimise the shape and thickness of shielding required.

To determine the available for the introduction of shielding, a large plane of lead is introduced just after the muon shield. Using the procedure developed for the muon flux study, the muon flux through this shield is studied. A rectangular shield with triangular cut-outs, resulting in an hour-glass shape, approximates the space well. The muon flux on the initial shield and the identified muon-free region is shown in Figure 5.15.

Freezing a shape, the thickness is studied. The lead shield is replaced by a sandwich of lead absorbers and sensitive planes, to study the evolution of the background rate throughout the shield. The results can be seen in Figure 5.16. After an initial steep decline in EM particle rates, they level off at around 5 cm and remain approximately constant after. This can be interpreted as a balance between absorbed particles and

\[\text{Note that this study was performed when there was still a dedicated upstream veto tagger, before its role was taken over by the muon tagger of the SND.}\]
Figure 5.14: Comparison of occupancy in the upstream veto tagger with and without the snd present. Without it there is an overall increase in occupancy of factor two, and an increase of $\times 50$ in the central region.
Background studies

(a) Muon rate in the volume between muon shield and snd which could accommodate shielding

(b) Residual muon rate in shield

Figure 5.15: (a) Rate of muons in the volume available for shielding. Close to the beam axis in $y$ the muons are swept out least, reducing the space available. (b) Within the identified free region, the residual rate is minimal. Figures courtesy of Alexandre Grandchamp.
new particles produced by the residual muon flux, as well as particles entering the shield from the sides.

While this shield is not yet included in the \textit{cds} baseline, this study shows that shielding can be used to effectively control the \textit{em} background in absence of muons. Furthermore, the space requirements are \( \mathcal{O}(\text{cm}) \), which can be easily accommodated around the \textit{snd}.

In future, it is planned to optimise the shield thickness and dimensions simultaneously to minimise the rate in the \textit{snd}. As the search space is much smaller than for the muon shield optimisation, a simple grid search would suffice, but using the optimisation framework developed would allow significantly loosening the assumptions, by \textit{e.g.} allowing for more complex shapes and non-uniform thickness.

5.6 Optimisation of the SHiP vessel shape

From the signal studies presented in Chapter 4 it is clear that the acceptance is naturally conical, so the larger volume of the elliptical cylinder of the \textit{tp} offers no benefit.

From the background perspective a conical frustum is also optimal: As muons are swept out of the acceptance in \( x \) to clear a wedge, the muon-free region can be utilised much more efficiently by a projective vessel.
Figure 5.17: Muon hits for the old focus in the upstream veto tagger at the front of the vessel, illustrating the problem of focused muon beams hitting the corners of the front of the vessel.

Figure 5.18: Comparison of hit rate (red) and $\text{hnl}$ acceptance (blue) for different foci in $x$. 
However, the focus of the conical frustum in $x$ needs to be carefully optimised, as the vessel walls have a finite thickness, and, while the cone of the signal acceptance points back to the target, the focus of the muon envelope is not a priori obvious.

From hit maps produced as described in Section 5.2, it is observed that foci too far upstream result in hot spots at the front of the vessel, as seen in Figure 5.17, where the vessel is too wide and touches the focused beams of high-energy muons, which are bent out just enough to miss the tracking stations, but not far enough to miss the front of the vessel. Away from these hot spots the non-muon occupancy is nearly uniform, so changes to the vessel focus will not significantly affect them.

As only one parameter needs to be optimised, we can perform a simple grid-search, and pick an optimum based on observed signal and background evolution over the range of $x$ studied.

The results of the signal and background simulations performed are seen in Figure 5.18. For small changes of the focus in $x$ the muon occupancy drops exponentially before levelling out around $10 \text{ m}$ relative to the target. The occupancy of all particles only decreases slowly, as it is dominated by non-muons, and is approximately proportional to the surface area presented, as their distribution is nearly uniform. The
effect on the acceptance for hnl, as a benchmark for all similar hidden sector decays, is well below the statistical uncertainty. Figure 5.19 shows the hot spots alongside a comparison of the vessel front sizes, visualising the effects of changing the focus. As the muon beams no longer impinge on the front face of the vessel, it is clear why there is no further improvement after $\sim 10 \text{ m}$.

As a result of this study, a focus of $10 \text{ m}$ relative to the target was chosen for the cds and the optimisation of the sub-systems downstream.

### 5.7 Summary

This Chapter introduced the machinery used to study muon induced backgrounds in the SHiP simulation. In addition to the studies presented here, it is central to the optimisation of the muon shield using full simulation presented in Chapter 6.

It was shown, that the combinatorial and muon dis backgrounds are under control, and some of the challenges of the optimisation, particularly the large number of muons in simulation required to have any muons pass the shield, were introduced.

The em background is shown to be independently reducible, which allows separating the optimisation of the muon and em backgrounds, such that the muon shield optimisation can focus exclusively on muons, significantly speeding up the simulation.

Finally, the conical frustum is confirmed as an optimal shape from the perspective of background reduction as well, and an optimal focus is selected, taking into account the effect on both signal and background.
Chapter 6

Optimisation of the muon shield

To achieve the SHiP physics goals, it is clear that an active muon shield is needed: As discussed above, passive shielding prevents having sensitive detectors as close to the target as needed to study very weakly interacting new particles efficiently with such an intense high-energy beam dump.

Earlier experiments usually were either further away from the target or at low enough energies that passive shielding was sufficient.

Notable exceptions here are donut [56] and faser [57]. donut used a system of two magnets to deflect muons. This system however did not succeed in reducing the rate of muons and related backgrounds as much as expected, forcing the experiment to accumulate an order of magnitude fewer protons on target than planned [58]. More recently, the planned faser experiment will use the dipole magnets of the LHC to sweep out charged particles from the LHC collisions. However, in this case, the existing magnets are used parasitically, as they were not designed with such a use in mind.

Due to the vital role of the muon shield for the SHiP experiment it is important that its configuration is optimised. To guarantee that the optimisation results can be trusted, this optimisation needs to be performed with the full simulation in the FairShip software framework, described in Section 3.6, which has been validated with available data, and will be further validated with dedicated measurements like the one presented in Chapter 7.

The muon shield needs to efficiently reduce the flux of muons by at least 6 orders of magnitude for a wide spectrum of muons with all momenta up to \( P \sim 350 \text{ GeV}/c \) and all transverse momenta up to \( p_T \sim 8 \text{ GeV}/c \). This spectrum is shown in Figure 6.1. The simulated muon sample available corresponds to \( 10^{10} \) protons on target, which is
a fraction of a second of the experiment’s planned run-time. Especially in the tails of the distribution the statistics available to the optimisation is very limited, and these rare events cannot be generated efficiently on their own. However, even the available sample comprising $\sim 1.7 \times 10^7$ muons is sufficiently large, that it cannot be feasibly simulated with FairShip for every proposed configuration.

Previously, the optimisation was performed using a fast but approximate stand-in for Geant4. Initially, it was performed in 2D (ignoring the $y$ direction, which is perpendicular to the bending plane), reducing the number of free parameters and simplifying the simulation by reducing the number of dimensions, and a much reduced sample of muons, biased towards those expected to be the most problematic to deflect. It was later extended to include the height of the magnets.

Using the Minuit optimisation package, a robust optimum was found, which was subsequently implemented in the full SHiP software framework and tested extensively using the full simulation. This configuration serves as a baseline for the new optimisation. Similarly, the previous optimisation procedure is built upon and is improved for

---

$^*$There have been several ideas to use machine learning techniques such as GANs or probabilistic programming for fast simulation, but these will require further study before they could be considered for the optimisation.
this study. For further information about the previous optimisation and the resulting configuration, see reference [45].

Even though this configuration works well, there are several reasons to re-optimise the muon shield:

Since the previous optimisation, several geometrical changes had to be made which change the geometry further and further from the configuration the shield was optimised for. These include changes to the magnetised hadron absorber necessitated by the integration into the target complex, as well as changes to the muon shield itself to obey engineering constraints, such as new estimates for the achievable field strength. This highlights the need for a repeatable procedure, such that the muon shield can be updated whenever the requirements change.

Several physics processes were not implemented fully in the fast simulation, which allows some muons to penetrate the shield, when simulated using Geant4. In order to find a configuration that efficiently deflects these muons as well, these processes need to be taken into account during the optimisation by using the full FairShip framework.

So far, the muon shield was only optimised with the hsd in mind. In order to include additional requirements, such as the maximum muon rate in the emulsion of the snd, a robust and repeatable procedure for the muon shield optimisation needs to be developed.

Furthermore, a new optimisation using a global approach could also find other local optima and possibly even a global optimum.

The optimisation however presents several challenges:

• With \( \mathcal{O}(50) \) free parameters, the dimension of the search space is high, even though the freedom in the shape of the individual magnets is already restricted as much as possible.

• For every single configuration that the optimisation algorithm decides to evaluate, a full simulation of the muon shield using FairShip is necessary, which takes about \( \mathcal{O}(10^3) \) CPU hours.

\(^\dagger\) See Section 6.2 for the parametrisation chosen. To significantly reduce the number of parameters, one would need to completely change the approach of parametrising the muon shield. One possible approach is presented in Section 6.8.1.
• Even though millions of muons are simulated, only $O(10)$ muons reach the acceptance. As a consequence, the value of the *loss function*, the metric that allows comparison between configurations, has a statistical uncertainty of about 10%. This uncertainty becomes larger for well-performing configurations, as fewer muons reach the acceptance.

This Chapter describes the design and optimisation of the SHiP muon shield using full simulation and machine learning techniques, and how these challenges are met.

### 6.1 Loss function

One of the key problems for automated optimisation is defining the loss function, the metric which quantifies the performance of a particular muon shield configuration. Deciding what to include in the loss function, and how to quantify these ingredients, will determine which optima are found. A poorly chosen loss function can lead to optima which find surprising ways to minimise the loss function without solving the original problem adequately or, due to lack of convergence, lead to no optima at all.

In the SHiP case, the desired optimum will minimise the rate of muons in the $hsd$, while also minimising length of the muon shield, indirectly maximising acceptance, and minimising its cost.

The length is easy to quantify, as it is analytically calculable from the magnet parametrisation.

To determine the muon rate however, simulation is needed. While the rate is quantifiable in all sub-systems, combining these occupancies into a single loss function is challenging, as different detectors have different constraints. To idealise the problem, a sensitive plane is introduced at a fixed distance from the muon shield —that of tracking station $t1$— and all detectors and the decay vessel are removed from the simulation to ensure they do not interfere. This has the side-effect of speeding up the simulation.

For each muon reaching the sensitive plane with a momentum $\geq 1$ GeV, its position in $x_\mu$ and $y_\mu$ on this sensitive plane is recorded. To incentivise sweeping out muons

---

$\dagger$ When manually optimising, the scientist usually has some implicit loss function, which they might subconsciously tune until the results agree with their expectations.
according to their polarity, the quantity \( \sigma_\mu \) (shown in Figure 6.2) is defined as

\[
\sigma_\mu^\pm = \sqrt{\left( 5.6 \mp \left( x_\mu^\pm \pm 3 \right) \right) / 5.6},
\]

where \( x_\mu \) is in metres. As muons are counted only in the acceptance \((-2.6 \text{ m} < \pm x_\mu < 3 \text{ m} \text{ for } \mu^\pm; |y_\mu| < 5 \text{ m})\), \( \sigma_\mu \) is automatically zero outside of it.

\textit{em} particles are not considered during the optimisation, as it was shown, that their rate can be controlled independently in Section 5.5. As the emulsion of the \textit{snd} is mostly limited by the rate of \textit{em} particles and low-momentum muons, the rate in the emulsion is out of scope of this first optimisation, but will be included in future.

The cost is harder to quantify. At the time of the study, first engineering designs were only just becoming available, without any easily parametrisable costings. So as a rough approximation, the material cost is used for the optimisation, which is proportional to the weight. In future this input to the loss function will be refined, as the engineering designs become more and more concrete (see Chapter 8).

With these ingredients in place, they need to be combined into a loss function. Initially, to allow for direct comparison, the new optimisation was to continue using the loss function pioneered for the first optimisation using the toy-mc, as presented in
Optimisation of the muon shield

equation 3.1 of reference [45]

\[ f(W, L, \sigma_\mu) = \frac{W \times (1 + \sum \sigma_\mu)}{1 - L/100}, \]  

(6.2)

where \( W \) is the weight of the configuration in tonnes, \( L \) is the length of the muon shield in metres. Note, that the sensitive plane was at the location of tracking station 4, while in the new optimisation it was moved upstream to the first tracking station, slightly increasing the acceptance.

However, it soon became clear that this loss function would not work without modifications for this optimisation: With the toy-\( mc \) it was possible to find regions of parameter-space in which no muons reach the acceptance. In this case, the loss function given in equation (6.2) works well, as \( \sum \sigma_\mu = 0 \), and thus there is no uncertainty on the loss function value. In this regime the weight and length can be efficiently optimised —even with conventional algorithms due to the absence of noise—, and whenever the muon shield becomes too light or short to sweep out all muons, the algorithm can return to the muon-less regime by backing away from this configuration.

Conversely, with the full simulation, not a single configuration —considering only configurations with a reasonable weight and length— without any muons reaching the acceptance was observed. This means the optimisation always remains in the regime where the loss-function value has a high uncertainty.

This difference is due to the physics processes considered in the toy- and full simulation. Study of the design optimised using the toy simulation showed that nearly all muons penetrating the shield undergo processes not included or approximated in the toy simulation. Particularly the proper treatment of multiple scattering and energy loss make the full simulation more sensitive to the statistical effects discussed here.

To study the uncertainty of the loss function, two tests were performed: A single configuration was simulated 65 times with different initial random seeds, while another configuration was simulated 88 times varying one parameter by \( \pm 1\% \) each run for each parameter, keeping all other parameters constant and reusing the same seed each time. The result of these tests is seen in Figure 6.3. In both cases, the fractional

\( \chi_\mu \) has since been renamed \( \sigma_\mu \) to avoid confusion of the quantity with the \( \chi^2 \) statistic.

More simulations were started, but initially the distributed computing framework had reliability issues, so once enough runs were successful to have an estimate of the uncertainty, other failed runs were abandoned. The failures were unrelated to the simulation itself, so there is no indication that the resulting sample is biased.
uncertainty is $O(15\%)$, but generally this uncertainty cannot be expected to be the same for all configurations.

Working in the uncertain regime leads to further negative feedback with the weight, as fluctuations in the number of muons passing the shield, are amplified by the weight. As a result, very heavy configurations with a weight factors heavier than the baseline could outperform it and other more realistic configurations because of downward fluctuations in the muon rate, leading to undesirable local optima.

To counteract this, a cut-off in the weight was introduced at 3 kt, which is 50% heavier than the weight of the baseline configuration, and an exponential penalisation of the weight difference to the baseline was introduced.

Further, it was noticed that, for configurations with reasonable weights, the weight and length were highly correlated. To simplify the loss function, it was thus decided to remove the explicit penalty on the length, in favour of the implicit penalty due to the weight.

The resulting loss function after the mentioned changes is

\[
 f(W, \sigma) = \begin{cases} 
 10^8 & \text{if } W > 3 \text{ kt}, \\
 (1 + \exp (10 \times (W - W_0) / W_0)) \times \left(1 + \sum \sigma (x) \right) & \text{otherwise},
\end{cases}
\]

(6.3)
where $W$ is the weight of the evaluated muon shield configuration, $W_0$ is the weight of the baseline configuration, and $\sigma_\mu$ is the weighted position of muon $\mu$ passing a sensitive plane at position $x_\mu$ as defined in equation (6.1).

There are methods to derive optimal loss functions, when multiple parameters need to be optimised simultaneously, such as in this case. However, these typically require running the optimisation many times —orders of magnitude more frequently than possible for this optimisation problem.

Note, that the statistical weights of the muons are not considered. In the sample used they are all of similar magnitude, so each muon is treated with equal weight in the loss function for simplicity. This choice will need to be revisited once the new, larger muon pre-production is used, as there the weights vary over several orders of magnitude.
6.2 Parametrisation of the muon shield

The parametrisation of the magnets needs to enforce invariants, such that any choice of numbers in the bounds results in a physical configuration, but also allow for enough freedom in the shape, that efficient designs can be found.

The key criteria for physical magnets are, that there is sufficient space for the coils, and, to ensure an efficient magnet circuit, a constant field width all around.

These considerations result in the following parametrisation: For each magnet, there are seven parameters. Three define each of the two ends of the magnet:

\[
\begin{align*}
\Delta x &= \text{field width}, \\
\Delta y &= \text{half-height of magnet core} \\
gap &= \text{gap between field and return field}
\end{align*}
\]

while the seventh defines the length. The shape is linearly interpolated between the two ends, and is symmetric in \(x\) and \(y\). Figure 6.4 shows one quarter of a magnet in cross-section.

Between magnets, there is a gap in \(z\) of 10 cm to allow space for the coils and to avoid interference between the magnets. The distance between the hadron absorber magnets and the free-standing muon shield magnets is under study by the \(bdf\) team, but is currently assumed to be 20 cm. This gap is dominated interface between the target complex and the experimental hall, which is subject to strict engineering requirements, particularly for to radiation protection.

The idealised field direction is part of the configuration, but the absolute field strength is a boundary condition of the optimisation. This means that the same configuration can be simulated with different field strengths.

To allow for the coils, a minimum area between the magnet limbs of 200 cm\(^2\), to allow for \(\mathcal{O}(1 \times 10^4)\) Ampère-turns is enforced, which is sufficient to achieve the required field when using grain-oriented steel. For the previous optimisation, space for the coils was cut out in the corners, however it became clear from engineering studies, that the

\[\text{Commonly the criterion of Pareto optimality is used, which describes configurations which cannot further improve a parameter without a trade-off in another. Selecting a suitable family of loss functions then allows one to choose one of these trade-offs.}\]
The muon shield consists of five identically parametrised magnets and a specialised sixth magnet: From studies of pathological muons passing the baseline shield, it became clear that sharp corners at the end of the final, sixth, magnet could result in scattering of muons back into the acceptance. To counter this, these corners were clipped, by hand for the baseline, and in an automated fashion for the optimisation.

Furthermore, support structures were introduced, so that their approximate material budget could be taken into account by the optimisation.

For later runs, the search space is discretised to cm steps and reduced as far as possible without adding any assumptions as to what a solution should look like.

The magnetisation of the hadron absorber was initially included in the optimisation, with the difference that the coil gaps were filled with non-magnetic steel and the magnets were supported by the hadron absorber, so no support structures were needed. However, after initial tests, the hadron absorber was fixed to new bdf designs and removed from the scope of the optimisation, as the strict and complex requirements resulting from the hadron absorber integration with the bdf, particularly due to high radiation, prevent automatic optimisation of its geometry. This reduced the number of free parameters to the final 42 (six magnets with seven parameters each).

The number of magnets, as well as the point at which the field is flipped is a fixed parameter at this point.

### 6.3 The muon sample

At the time of this study, only $1.7 \times 10^7$ muons were available. In future runs, the new pre-production will be used. The pre-production procedure and the two samples are described further in Section 5.2. The $P-p_T$ distribution of the available muons is shown in Figure 6.1.

The true muon distribution produced in the target is not known. While the simulation is tuned to agree with all available data, most available data was taken under conditions unlike those expected at SHiP, e.g. with lighter targets and lower or higher energies.
While a cross-check was being planned, as discussed in Chapter 7, this proof-of-concept optimisation would have to be performed before the results become available. To prevent a surprise *à la donut*, it was decided to resample the available muon sample to intentionally bias the sample towards the most dangerous regions of the muon phase space, and to even out the coverage of the muon sample. As low momentum and low-$p_T$ muons make up the majority of the available sample, this also has the welcome side-effect of reducing the overall number of muons that need to be transported through the muon shield, speeding up the simulation.

The histogram-based re-sampling procedure is the following:

1. In each bin of $P-p_T$, only keep 100 muons (Figure 6.5a), reducing the sample from $1.7 \times 10^7$ to $1.6 \times 10^5$, by capping the well-populated bins.

2. In each bin with at least one muon, sample from the muons in that bin, randomising $\phi$, until there are at least 10 in each populated bin of the sample (Figure 6.5b), adding $1.4 \times 10^4$ muons in total.

3. For each muon, add a copy with the $p_T$ rotated into $p_x$ and a copy rotated into $-p_x$ to maximise the $|p_x|$ in the bending plane. Skip this step for the added muons with random $\phi$ to avoid duplication. This adds $3.15 \times 10^5$ muons (Figure 6.5c).

The resulting sample (Figure 6.5d) comprises $4.75 \times 10^5$ muons, reducing the overall sample size by a factor of $\sim 35$.

A similar procedure was also used to construct the muon sample for the optimisation presented in reference [45]. The main difference is, that in reference [45] additional muons of fixed $p_T = 2.5 \text{ GeV/c}$ and momentum uniformly distributed over $150-350 \text{ GeV/c}$ were added to the sample.

Note, that the resampled muon sample is not a true sub-sample of the original sample, due to the addition of muons in the tails of the $P-p_T$ distribution.

The spatial distribution of muons seen on the scoring plane at tracking station $\tau_1$ is shown in Figure 6.6. The distributions of anti-muons are nearly identical. There is a clear separation into two populations: A contiguous area the muons are swept into above $x = 400 \text{ cm}$ with a bulge at the level of the beam, and an approximately uniform distribution dominated by low-momentum muons that are scattered on their

**The binning is 100 bins each in $P \in [0,350] \text{ GeV/c}$ and $p_T \in [0,6] \text{ GeV/c}$. There are no muons outside of this range in the available sample.**
Figure 6.5: Resampling procedure shown step-by-step. See Section 6.3 for a detailed description of the individual steps. (b) and (c) only show added muons. (Continued on page 107)
Figure 6.5: (Continued) Resampling procedure shown step-by-step. See Section 6.3 for a detailed description of the individual steps. (b) and (c) only show added muons.
exit from the muon shield. Due to the reduction of low-momentum muons in the
resampled muon sample, the uniform contribution is reduced, and the bulge becomes
more important, as it comprises the high-momentum muons swept out just enough to
miss the acceptance.

6.4 Bayesian optimisation

In optimisation, the problem to be solved is to extremise (without loss of generality
let us consider minimisation) some function \( y = f(x) + \epsilon \), where usually \( x \) is an \( d \)-
dimensional vector of parameters with respect to which \( f \) is optimised, and, as in this
case, there might be some statistical noise \( \epsilon \) on every sampling of \( y \).

In high energy physics the standard optimisation and fitting package is Minuit. However, for the problem at hand, it struggles, as it depends on gradients for optimi-
sation. The usual methods of approximating a gradient fail, as even very close points
differ in an unpredictable way due to the inherent uncertainty of the loss function.

Bayesian optimisation \([59–61]\) tries to solve the problem of optimising functions
effectively under the following constraints:

- There is no known analytic expression, \textit{i.e.} no gradient information is known,
- the function is expensive to evaluate, \textit{and}
- the evaluations of the function have a significant uncertainty.

All of these constraints are present in this case. However, we have the additional
constraint of a high-dimensional search space.

At the core of the algorithm is the following iterative process:

1. Use the evaluations \((x_i, y_i)\) of the loss function to build a probabilistic surrogate
model of the loss function.

2. Optimise the so-called acquisition function constructed from the surrogate model
to find the next \( x \) to evaluate. This acquisition function is cheap to evaluate. A
commonly used choice is the expected improvement, which is defined as

\[
EI(x) = \max (y^* - \langle f(x) \rangle, 0),
\]  

(6.4)
Figure 6.6: Comparison of spatial distribution of muons incident on the sensitive plane at the $z$-position of the tracking station $t_1$ for the two muon samples.
where $y^-$ is the minimal $y$ observed so far, and $\langle f(x) \rangle$ is approximated using the surrogate model.

3. Evaluate the next point $x$ and repeat.

Different choices of acquisition function allow controlling the trade-off between exploring the search space and exploiting the acquired knowledge to find an optimum. Asymptotically, the global optimum is found.

As Bayesian optimisation is a global algorithm, the traditional optimisation concepts of step size or starting point have no direct correspondence.

For this study, the open-source scikit-optimize [62] implementation of Bayesian optimisation is used. In future a more specialised or optimised implementation could be considered, but for the proof-of-concept optimisation, this implementation proved sufficiently performant and flexible.

To construct the surrogate models, three different regression algorithms are used in this study: Gaussian processes, random forests and gradient-boosted decision trees. Gaussian processes approximate the loss function by using a collection of random variables, which jointly follow a normal distribution. For Bayesian optimisation, their most important property is, that they have analytic expressions for their mean and variance, allowing the optimiser to take into account the uncertainty when optimising the expected improvement. When using common priors, the convergence properties of Bayesian optimisation using Gaussian processes are also well-understood. However, the prediction of points to evaluate when using Gaussian processes scales badly with a high number of samples to fit, so the two decision-tree-based algorithms are also used, as they scale better, at the cost of sacrificing the uncertainty information.

In the following subsection, some of the properties of Bayesian optimisation will be highlighted using a simple example. See reference [63] for a more formal introduction to Bayesian optimisation.

### 6.4.1 An example problem

Consider the function

$$f_n(x, y) \equiv (\sin 5x \times (1 - \tanh x^2)) \times (\sin 5y \times (1 - \tanh y^2)) + \epsilon, \quad (6.5)$$
shown in Figure 6.7a for \( x, y \in [-2, 2] \), and where \( \epsilon \) is normally distributed noise with zero mean and a width of 0.1. This function is chosen, because it has several degenerate local minima, which are not very deep compared to the level of noise. Traditional optimisation algorithms such as Minuit struggle with this function and do not generally find an optimum.

Figure 6.7b shows a typical evolution of the optimisation for Bayesian optimisation, and for comparison random (uniform) sampling of the search space.

It is clear immediately, that the convergence behaviour is unlike that of traditional algorithms where the loss function decreases nearly monotonously. This is because the Bayesian optimisation algorithm always keeps exploring in order to reduce the uncertainty of its estimate of the loss function. It only asymptotically converges when the global optimum is found and the uncertainty everywhere is low enough that there is no doubt about the optimum. In all but the simplest one-dimensional cases, one would not expect to reach this point.

A more useful concept is the cumulative optimum—here minimum—of the loss function, i.e. the most optimal loss function value seen up to the iteration in question. It allows monitoring the progress of the optimisation, but one can never be certain, that no further improvement is possible.

Note, that the Bayesian optimisation requires initialisation to bootstrap the surrogate model, usually by initially sampling the loss function uniformly. Even after this initialisation is complete, it may take a significant number of iterations until the loss function is approximated sufficiently well to outperform random sampling. As seen in Figure 6.7b, the optimisation proceeds in jumps: The loss function value oscillates around a central value, which abruptly changes as better points are evaluated. As Bayesian optimisation is a stochastic algorithm, the actual improvement will differ from the expected improvement predicted by the current model. Note as well, that in this function with two degenerate global minima, only one is found. This is evident from the evaluations shown in Figure 6.7a, which as the optimisation progresses are tightly clustered about the minimum near \((-0.29, 0.29)\).

The scaling behaviour to large numbers of dimensions does not yet seem to be well-studied in literature.
Optimisation of the muon shield

Figure 6.7: (a) Contours of $f_n$ (noise not shown). The markers show the evaluations of the loss function, clearly showing the convergence to one of the degenerate minima. (b) Convergence of optimisation using Bayesian optimisation using Random Forest for regression (left) and random sampling (right). The transition from random initialisation to optimisation is indicated using the dashed red line.
6.5 Computing architecture

For the optimisation, a large number of simulations would have to be performed in a time critical manner. To fulfil this requirement, a specialised job management framework, disneyland, was developed to make use of the Yandex Skygrid, a distributed computing platform used by SHiP that is based around the concept of job containers. This framework allows direct submission and control of simulation jobs to a dedicated cluster of $1.6 \times 10^3$ worker nodes.

The simulation jobs are run in Docker containers, to ensure that the execution environment is reproducible, and all random number seeds were set in the job description, such that the same configuration would always result in the same loss function value. In addition to ensuring reproducibility per se, this also allows caching of evaluations, such that every point only needs to be computed once, even if requested multiple times, be it in separate optimisation runs or within a single run.

The initial state of the Bayesian optimisation is also controlled to guarantee reproducibility.

Each evaluation is parallelised between 16 worker nodes, for up to 100 evaluations in parallel. Further parallelisation of evaluations would lead to diminishing returns, as the constant costs of the simulation and job-submission framework begin to dominate. Increasing the number of parallel evaluations would increase the time needed to predict new points to evaluate significantly, as it does not scale linearly. To predict several points, scikit-optimize uses the algorithm presented in reference \[64\]. The algorithms are robust to individual failed jobs. Parallelised like this, the cluster can evaluate up to $\sim 2 \times 10^3$ points per day, assuming full cluster availability.

The cache of results can also be used to speed up the initialisation of the optimisers by giving them access to the entire evaluation history, i.e. the entire available knowledge of the loss function, and even replay the entire optimisation run to e.g. analyse its evolution in more detail or to recover from technical issues.

A modular Python 3 framework was built on-top of the disneyland-client library, while FairShip is entirely confined to the container images. The modular design allows

---

\[64\] Yandex is a Russian search engine and internet company that is very active in machine learning research. Their School of Data Analysis is supporting LHCb and Comet, and are also supplying a large part of the SHiP computing infrastructure and is a member of both LHCb and SHiP.
replacing the scikit-optimize optimiser for cross-checks and potential future studies using different optimisation approaches.

### 6.6 Optimisation runs

Even when using Bayesian optimisation, the muon shield optimisation is at the limit of what is feasible. To develop an understanding of the behaviour, several test optimisation runs are performed.

In particular, the convergence and dependence on the size of the search space are studied. For this study, no automated stopping criteria are defined: Once the cumulative minimum of the loss function ceases to improve, which it usually does after $\mathcal{O}(5 \times 10^3)$ iterations, the optimisation is halted. As the optimisations are reproducible, each optimisation run could in theory be resumed at a later point.

Recall the example in 2D, shown in Figure 6.7b. A similar transition is observed in the full 42-dimensional case as shown in Figure 6.8. This demonstrates, that the surrogate model starts making predictions which are significantly better than the initial random exploration, but also shows that in 42 dimensions, this initialisation phase requires $\mathcal{O}(1 \times 10^3)$ evaluations of the loss function. For clarity, hereafter only the evolution of the cumulative minimum is shown.

A strong dependence on the size of the search space is observed. When reducing the number of free parameters or limiting the size of the search space, the cumulative minimum of the loss function improves much sooner.

In addition to these tests to study the convergence behaviour and the optimisation procedure in general, two main optimisation runs were performed, which are described in detail below.

The best configurations found prior to termination for each run are then studied further, to determine the performance on the full sample and the rates in the individual sub-systems of SHiP. The results of these studies are shown in Section 6.7.
Figure 6.8: Evolution of loss function for a run using random forests performed to study convergence behaviour and the effect of the weight cut-off. Note, that there is a clear transition after about $1.2 \times 10^3$ evaluations of the loss function, similar to that observed in Figure 6.7b.

6.6.1 Continuous optimisation

The first optimisation run had the goal of finding the baseline configuration or a similar configuration under approximately the same conditions as used for the optimisation presented in reference [45]. For this, a field strength of 1.8 T was set and the previous hadron absorber design was used.

The search space for this run is continuous, and constrained between an artificially constructed minimal configuration and the maximum of the parameters for known constructible configurations.

The evolution of the optimisation is shown in Figure 6.9. As one of the points used to initialise the optimisation is the baseline, the algorithm rapidly explores the region around it and manages to find many configurations with better loss function values, and ceases to improve after about $1.5 \times 10^3$ iterations. The best found configurations are cross-checked with the full sample, and the second best configuration observed performs best on the full sample, and is within the uncertainty of the best configuration on the reduced sample.
Figure 6.9: Evolution of the cumulative minimum of the loss function for the continuous optimisation run.

See Section 6.7 for the performance of this configuration and further studies. Due to the constraints in search space and initialisation of this run, the configuration resembles the baseline, but is ~25% lighter.

6.6.2 Low-field optimisation

The purpose of the optimisation run was to determine whether a configuration could be found that performed acceptably well with only 1.7 T field in the magnets, and with the new hadron absorber design with 1.6 T nominal field.

Compared to the previous optimisation run, the search space is discretised to a grid of 1 cm steps, and the boundary of the search space has been defined independently of any previous configuration, to minimise the influence of any prior assumptions.

The evolution of the optimisation can be seen in Figure 6.10.

For the surrogate model Gaussian processes are used initially, and, once their evaluation becomes expensive, random forests are used instead, based on previous experiences during the test runs. After ~1.4 x 10^3 iterations, the cost of computing new Gaussian process predictions started to dominate over simulation time, so an optimiser using random forests was used with the same history for the remainder.
The coverage of the search space is sparse throughout, but without any obvious holes, as expected for successful exploration of such a high-dimensional space.

After about $4 \times 10^3$ iterations, no further progress was apparent, so the search space was reduced to the vicinity of the best found configurations by imposing that all parameters are within their parameters $\pm 10\%$ and within the original bounds. Since most evaluations up to then were outside of this subspace, it was feasible to switch back to Gaussian processes for the rest of the optimisation run.

### 6.7 Results

As some configurations are observed to have significantly different performance between the reduced and full data set, it is necessary to check promising candidates. To evaluate their performance, they were simulated with several times with different random seeds and both the reduced and full data sets. This also allows an estimation of the statistical uncertainty on the performance.

The best configuration discovered in the low-field optimisation run (called low-field optimum from now on) solves the optimisation problem well for the reduced sample, but performs a factor $O(10)$ worse on the full sample. It is shown in Figure 6.11. For previously studied configurations at 1.8 T, the performance on the full sample was
only a factor of a few worse than on the resampled muons. The loss values for all mentioned configurations are shown in Table 6.1.

As a cross-check of the optimisation result, the baseline configuration was scaled in length to keep the overall $\int B dl$ along the muon path constant. This naïvely scaled configuration is called scaled baseline hereafter. Its performance for 1.7 T is as expected comparable to that of the baseline for 1.8 T. This confirms that an acceptably light and short configuration is possible with 1.7 T field independently of the optimisation.

As the result of the low-field optimisation run does not perform well enough on the full sample, the runner ups in the optimisation and configurations from previous optimisations are considered as well. However, the runner ups do not perform any better so are not considered further.

Somewhat surprisingly, both the baseline and the previous optimum (shown in Figure 6.12) perform better than the low-field optimum and scaled baseline on the full sample at 1.7 T, an environment for which they were not optimised. The previous optimum also performs well on the small sample, outperforming the low-field optimum. To confirm this the previous optimum is rerun with 10 seeds each for both samples. Additionally the closest point on the discrete grid is also run 10 times each for statistically indistinguishable results. In the tables they are not differentiated. This also shows that the configuration is robust to changes of $O(0.5 \text{ cm})$ to the parameters.

Note, that the parameter values of the baseline and previous optimum do not lie entirely within the bounds defined for the low-field optimisation run, due to the change in search space between the two runs.

The different effects of the change from 1.8 T to 1.7 T on the different configurations and muon samples is interpreted with the following hypothesis: The small sample is—by design—biased to high (transverse) momentum muons. The reduced integrated field due to the lower field further increases the importance of high-$p$ muons. This effect results in the optimisation being over-trained for the high-$P$ muon range, as it is given too much importance. When checked with the full sample, the configuration then encounters orders of magnitude more low-$p$ muons, some of which it cannot handle, as they were not important enough in the optimisation sample.

A possible solution might be to weight the muons in the resampled muon sample by their actual abundance in the full sample. This way the coverage of the entire parameter
Figure 6.11: Low-field optimum from (a) the side, (b) from above.

Figure 6.12: CDS muon shield configuration from (a) the side, (b) from above.
Table 6.1: Resulting loss for the discussed configurations with 1.8 T and with 1.7 T and the updated hadron absorber respectively. Estimates of uncertainty are indicated for configurations that were evaluated multiple times. As a rule of thumb ± 20% seems to be a good estimate if one can extrapolate between configurations. This is in agreement with the measured uncertainty of the loss function (see Section 6.1 and Figure 6.3).

<table>
<thead>
<tr>
<th>Configuration</th>
<th>Length [m]</th>
<th>Weight [kt]</th>
<th>1.8 T baseline</th>
<th>1.8 T resampled 1.7T</th>
<th>1.8 T full 1.7T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Previous optimum</td>
<td>34.60</td>
<td>± 7.2</td>
<td>34.82 ± 1.28</td>
<td>34.82 ± 1.28</td>
<td></td>
</tr>
<tr>
<td>Low-field optimum</td>
<td>36.08</td>
<td>± 7.7</td>
<td>34.46 ± 1.46</td>
<td>34.46 ± 1.46</td>
<td>34.46 ± 1.46</td>
</tr>
<tr>
<td>Scaled baseline</td>
<td>37.0</td>
<td>± 5.7</td>
<td>36.08 ± 7.7</td>
<td>36.08 ± 7.7</td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>37.0</td>
<td>± 5.7</td>
<td>36.08 ± 7.7</td>
<td>36.08 ± 7.7</td>
<td></td>
</tr>
</tbody>
</table>

Optimisation of the muon shield
space is evened out, without having to bias the distribution that the muon shield is optimised for.

The low-field optimisation run demonstrated that the previous optimum found in the continuous run is robust to the changes to the hadron absorber and the lower field. As a result, it was frozen as the configuration for the comprehensive design study (cds) for the study of the other sub-systems, backgrounds and sensitivities. An engineering design was also implemented in computer-aided design (CAD) (see Figure 6.13) to compute realistic field maps using the OPERA finite element modelling package‡‡. The realistic field maps and engineering designs are discussed further in Chapter 8.

While this configuration performs nearly as well with 1.7 T as with 1.8 T, it is also \( \sim 25 \% \) lighter at comparable length. The corresponding muon and EM rates in the different HSD sub-systems are within the statistical uncertainty of the baseline rates.

The performance of the cds configuration has also been checked independently of FairShip by the CERN radiation protection unit using FLUKA. The simulated flux is shown in Figure 6.14 and is consistent with the flux simulated with FairShip.

Using the new pre-production the pathologies of the muons passing through the shield can be studied. Figure 6.15 shows the original momentum distribution, showing that some muons are focused into the acceptance. Many of these muons pass the shield by scattering to the wrong side early in the shield, and are subsequently bent into the

‡‡ There is no connection to the experiment of the same name.
Figure 6.14: Performance of muon shield as evaluated with FLUKA at the beam axis. Taken from reference [41].
Figure 6.15: Original momentum and transverse momentum of muons that penetrate the new shield configuration when simulated with the new pre-production.
acceptance. To combat this effect, one could consider adding a term to the loss function to explicitly penalise this behaviour in the future.

The rate of low energy muons and \( \text{em} \) particles in the emulsion of the \( \text{snd} \) and possible countermeasures are under study. Potential shielding can be optimised now that the muon shield is frozen. The results of this study could lead to the inclusion in the loss function and a further optimisation run.

However, there are two puzzles that remain and will require further study: We need to understand under which circumstances the correspondence between full sample and resampled muon sample holds. While in most cases, performance on resampled muons and the full sample very correlated, this correspondence sometimes breaks down. The behaviour observed between the two data sets could be analogous to over-training in machine learning, where the reduced and full data sets take the roles of \( \text{train} \) and \( \text{test} \) data sets respectively. Cross-validation and other common methods to counter this phenomenon in machine learning might be adaptable to our problem, but this will need to be studied further.

Of lesser importance, but still worthwhile understanding, is the question, why the configuration from the previous optimisation run is as robust to changes as it is. Robustness indicates further optimisation potential, \textit{i.e.} that a configuration is not yet optimal, as for an optimal configuration any change would reduce its performance. But robustness is also desirable, to take into account uncertainties in the muon spectrum, and to allow for manufacturing tolerances and other safety factors.

### 6.8 Conclusion

Being able to optimise the SHiP muon shield in response to changing requirements is critical for the interplay with the other sub-systems and further iterations of the engineering design.

In this Chapter, a procedure was developed to optimise the shield, resulting in a robust new baseline configuration for the \( \text{cds} \) studies.

For this first proof-of-concept optimisation, the scope was kept intentionally narrow: Only the rate of muons in the \( \text{hsd} \) was considered, and the fields in the muon shield were all idealised.
With the procedure in hand, the optimisation can now be repeated to include further constraints, such as the rates in the SND. Similarly, now it is conceivable to iterate the parametrisation to reflect the improving understanding of the engineering requirements, and to use the realistic field maps to parametrise the field of the magnet segments realistically.

6.8.1 Future research

While all muon shield designs and the optimisations so far have been following the same concept of two magnet systems, this may not yet be the optimal solution. Now that the machinery exists to solve this high-dimensional optimisation problem, one could give the algorithms more freedom by imposing fewer \textit{a priori} assumptions on how the magnets should be arranged. The engineering design currently favours using smaller modules of grain-oriented steel (see Chapter \[8\]). Parametrisating the muon shield in terms of these smaller modules might allow for more freedom, while also automatically enforcing more of the engineering constraints.

To solve the correspondence problem when extrapolating the performance between muon samples, there is a promising approach to combine samples of different fidelities to construct the surrogate model presented in reference \[65\]. This would allow using a small sample to explore the loss function as before, but in critical regions, the algorithm could improve its estimate of the muon shield performance by simulating a large sample automatically. This could be tuned to balance optimisation speed and confidence in the performance of simulation results, and effectively automates the manual procedure of optimising with the small sample and checking the performance of configurations on the full sample.

There are some other non-conventional optimisation approaches, which could offer advantages over Bayesian optimisation, and should be evaluated: \textit{Evolutionary algorithms} might offer better scaling behaviour, as the parallelisation of the Bayesian optimisation to the entire cluster has several difficult-to-avoid bottle-necks. However, evolutionary algorithms usually require careful tuning of the hyper-parameters, which require many optimisation runs. This limitation might be overcome by performing this hyper-optimisation on a simpler but comparable problem, but whether a suitable problem can be found is not guaranteed.
A novel approach to optimisation, which could work for this difficult optimisation problem, could involve the use of Reinforcement Learning, a field of techniques that is rapidly developing currently. Mapping the optimisation of the shield into a solvable Reinforcement Learning problem is the key challenge: The actor must be given quick feedback and actions that allow it to find a solution to the problem. The move to the smaller grain-oriented steel modules might benefit this approach.

Finally, an interesting approach was presented in reference [66], developed for the optimisation of a nuclear fusion experiment, where the evaluations of the loss function involved physical experiments: The Optometrist Algorithm is designed to combine a stochastic algorithm with human expert intervention, for problems with hard to specify loss functions. In the SHiP muon shield optimisation, the performance of the shield configurations with respect to different sub-systems has complicated interdependencies, and, as a result, is hard to quantify. However, in many cases, comparing e.g. the occupancy maps of two configurations allow the physicist to easily pick a more promising configuration. So this approach might be applicable to the presented optimisation problem.
Chapter 7

Measurement of the muon flux

The SHiP background estimates and in particular the muon shield optimisation depend on the knowledge of the muon spectrum emanating from the SHiP target. While the simulation is tuned to available experimental data, there is no data available for the conditions expected at SHiP. Furthermore the target environment is challenging to simulate due to the large multiplicity of particles and re-interaction of particles within the long, dense target. To cross-check the simulation, an experiment was performed to replicate the conditions as closely as possible and directly measure the muon spectrum.

The design, data taking and analysis of this experiment are outlined in this Chapter. As the analysis is still in progress, all numbers and plots in this Chapter are preliminary. For the proposal for this measurement, see reference [53].

7.1 Design of experiment

The layout of the experiment is shown in Figure 7.1.

A replica of the SHiP target with the same configuration of TZM and tungsten is exposed at the H4 beamline at the CERN SPS. This allows using the same beam energy as expected at SHiP. Additional concrete and iron shielding allow intensities of up to $2 \times 10^7$ protons on target per spill, allowing to integrate $\mathcal{O}(5 \times 10^{11})$ protons on target over three weeks, a sample comparable to the available statistics from simulation.

A hadron absorber following the target reduces the rate of non-muons seen at the detectors. It is followed by a magnetic spectrometer comprising four tracking stations using OPERA drift-tubes and the Goliath dipole magnet, which bends charged particles
in the $z-x$ plane. The first two tracking stations have $u$ and $v$ stereo layers respectively, and also surrounded by the $s_2$ scintillators, which together with the beam counter $s_1$ form the trigger, while the tracking stations after the magnet do not have stereo layers.

RPCs with iron filter walls follow the fourth tracking station and act as muon identification system.

### 7.2 Online and offline computing

The data acquisition (DAQ) system of the measurement is distributed, i.e. the tracking stations and muon identification system have independent local DAQ systems, which send their data to the central event builder. The run control and communication between the DAQ systems and event builder uses the ControlHost messaging protocol.

The local data of the tracking stations and the built events are converted into the FairShip analysis format online, using the same procedure as for the later offline conversion, allowing near-realtime monitoring of the data quality during data taking.

The data conversion is based on the modular FairRunOnline framework of Fair-Root.

The tools developed for data taking, processing and analysis are reused by the feasibility study for the charm cross-section measurement proposed in reference [67], which took data for a week following the muon flux measurement.

---

**Figure 7.1:** Experimental set-up for the muon flux measurement. Taken from reference [41].
7.3 Data taking

The data taking can be broadly separated into three types of runs, distinguished by the Goliath field used for the run.

With the full field of Goliath, $5.3 \times 10^{11}$ protons on target were collected. In addition, $2.7 \times 10^{10}$ protons on target were collected with no field for alignment purposes. Furthermore, $4.8 \times 10^{10}$ protons on target were collected with a reduced field strength to increase the acceptance for low-momentum muons. In total, $O(6 \times 10^{11})$ protons on target were collected, surpassing the goal. The normalisation is still being calibrated, as efficiency and dead-time of the trigger scintillators and trigger are studied more closely.

7.4 Analysis

For accurate tracking, the calibration of the drift times is crucial. The readout of the \texttt{tdcs} used for the drift tubes is triggered by a delayed master trigger. When triggered, the preceding window of $\sim 2.3 \, \mu s$ is read out. So that the drift time can be reconstructed without clock jitter, the master trigger is also sent to \texttt{tdc4}. The raw signals from the trigger scintillators and triggers are read out by the \texttt{tdcs} as well, to allow for later analysis of the efficiencies.

As time differences on the same \texttt{tdc} cancel out the jitter, the accurate drift time for a drift tube read out by \texttt{tdci} can then be reconstructed as

$$t_{\text{drift}} = \Delta t_{\text{delay}} - t_i + t_{\text{raw}},$$

(7.1)

where $\Delta t_{\text{delay}} = t_4 - t_{\text{MT}}$, $t_i$ are the delayed triggers on \texttt{tdci}, and $t_{\text{MT}}$ is the master trigger. This calibration is performed for each hit during the conversion to the \texttt{FairShip} data format.

The resulting drift time spectrum is shown in Figure[7.2] It can be used to measure the space-to-drift time ($rt$) relation. The $rt$-relation depends on the gas in the drift tube. The $rt$-calibration is performed during the reconstruction for groups of four drift tubes along the gas flow for a few spills at a time, to ensure approximately constant gas-properties.
Measurement of the muon flux

A coarse alignment using survey data and biased residuals from tracking is performed as described in reference [68]. A fine alignment using the Millepede software package is in progress.

To calibrate the absolute momentum, decays of known particles to a dimuon final state will be studied.

A large simulated sample is used in analysis to understand efficiencies and tracking performance.

7.5 Summary

The optimisation of the muon shield and the background studies rely on accurate simulation of the muon flux emanating from the SHiP target. As the SHiP target is very dense, the cascade of interactions and re-interactions is challenging to accurately simulate. The muon flux measurement successfully took data, which will result in an accurate muon spectrum that can be used to check and correct the simulation for any mis-modelling of the physics processes in the target. The resulting muon spectrum can
also be directly used to test the performance of the current muon shield. Furthermore, it may be used to re-optimise the muon shield independently of the target simulation.
Chapter 8

Muon shield technology and prototyping

The optimised muon shield is idealised. While the design of the optimisation constraints was based on early engineering studies as presented in reference [69], the engineering design needed to be concretised for the comprehensive design study (cds).

With the new baseline configuration for the cds resulting from the optimisation presented in Chapter 6, a possible engineering design is studied. This Chapter summarises some of these studies and further plans for prototyping.

8.1 Muon shield technology and engineering challenges

The complicated shapes required by the muon shield severely restrict the space available for the magnet coils and cooling. To achieve the high fields using warm magnets, grain-oriented (go) steel is a natural option. In grain-oriented steel, the grains are aligned in one dimension, significantly reducing the magnetic reluctance. This allows producing higher fields with reduced currents, and without dedicated cooling, which significantly decreases the profile of the coils needed. In industry, grain-oriented steel is commonly used for transformer magnets, but a magnet system of the size and complexity of the SHiP muon shield is unprecedented.

The key limitation of grain-oriented steel is the flip-side of its advantage: due to the aligned grains, joints between go steel segments need to be carefully engineered to preserve the low reluctance of the magnetic circuit. From the manufacturing per-
spective, assembly of large structures is challenging, as go steel is usually available as laminated sheets, 300–500 μm thin, which necessitate complicated support structures.

Recent material science advances could potentially lift some of these limitations: The results presented in reference [70] suggest that it might be possible to construct more complex go steel structures using Selective Laser Melting in a process that is reminiscent of 3D-printing for plastics.

### 8.2 Realistic field maps

Using the CAD model of the CDS muon shield, realistic field maps are computed using the OPERA finite element modelling package for a design field of 1.7 T. The resulting field is shown in Figure 8.1. In the critical regions, which will see the largest flux of high-momentum muons, an approximately uniform field of nearly 1.7 T is achieved.

The field maps were subsequently implemented in FairShip and the performance is compared to that of the idealised field. As can be seen in Figure 8.2, the performance is consistent with that of the idealised field.
8.3 Assembly

The steel sheets will be laser-cut to size. Up to 150 sheets at a time can be welded together into packs of 5 cm depth. Figure 8.3 shows how these packs are then bolted together into rectangular modules of 50 cm depth, which can then be used to construct the magnets.

A possible assembly of the muon shield from the resulting modules is shown in Figure 8.4.
8.4 Prototyping and plans

In laboratory tests, the magnetic properties of different joint techniques have been studied. The results can be seen in Figure 8.5. While welded joints with annealing cannot recover the performance of perfectly aligned Co steel, it significantly reduces the current density required compared to untreated welded joints.
To fully evaluate the effects of the stacking factor, determined by the achieved density of stacking of the sheets, and joints, it is necessary to produce a realistic prototype, as they are difficult to accurately model.

A 1:4 scale prototype of one half of a muon shield magnet is currently being planned.

An alternative design that would involve no joints is the unicore design, which is based around the concept of wrapping the go steel sheets around the coils to allow the grain direction to always approximately coincide with the field direction. A smaller prototype using a unicore design is also being planned.

The properties of these prototypes will be studied by measuring induced currents. Test with a muon beam are also being considered.

8.5 Implications for the optimisation

The currently used frustum shaped magnets are difficult to manufacture and assemble using go steel. Instead, it might be easier to re-parametrise the muon shield in terms of the rectangular go steel modules directly. As the number of modules is much larger than the number of magnets currently, the degrees of freedom per slab will have to be reduced, to maintain $O(50)$ degrees of freedom in total. One could envision not parametrise each module individually, but making use of their number to approximate smooth functions.

Alternatively, if the Reinforcement Learning approach proves feasible, the number of degrees of freedom is less important, and the freedom the modules give the actor could result in completely novel designs of the muon shield, while still ensuring feasibility of construction.

Finally, the unicore design would require a completely different approach to parametrising the muon shield as it becomes clear, what kind of shapes can be constructed this way.
Chapter 9

Conclusion

This thesis comprises many studies performed for the optimisation of the SHiP experimental design.

The signal acceptance was studied for a benchmark channel to establish the dependence on the position, length and geometry of the SHiP decay vessel.

The backgrounds induced by muons were studied, showing that the muon dis in the vessel and experimental cavern, as well as the muon combinatorial background are under control. Furthermore, it was demonstrated, that the em background can be reduced independently in the muon-free region, allowing the optimisation to focus on muons. Finally, a procedure was developed to reproducibly study the em and muon rates in all SHiP sub-systems. Thus, they can be repeated in response to any updates of the experimental layout.

These signal and background studies were used to optimise the geometry of the SHiP decay vessel.

Informed by these studies, the muon shield was re-optimised using full simulation and machine learning techniques. As a result of this, a new baseline configuration was found for the cds. Even though this particular configuration will probably be superseded as a result of the current prototyping and the concretisation of the engineering design, the more important result is the optimisation procedure itself: An algorithm has been developed, which allows us to re-optmise the muon shield whenever necessary by

1. Parametrising the muon shield design,
2. Adjusting the loss function, if required,
3. Automatically optimising the muon shield, and finally,

4. Manually correcting for any observed focusing effects.

This method has been shown to work reliably, and is implemented in a flexible software framework, so that it can be extended and improved.

A measurement of the muon flux has been performed to cross-check the simulation which the optimisation is based on. The results are not yet available, but regardless of whether they confirm or disagree with the simulation, this can be fed back into the optimisation to update the muon shield design.

Finally, the technological and engineering challenges of the muon shield and associated prototyping were summarised.

The work presented in this thesis is a key part of the SHiPcds. The overall status of these studies is presented in reference [41].
Bibliography


[5] Belle collaboration, Measurement of the branching ratio of $\bar{B} \rightarrow D^{(*)}\tau^-\bar{\nu}_\tau$ relative to $\bar{B} \rightarrow D^{(*)}\ell^-\bar{\nu}_\ell$ decays with hadronic tagging at Belle, Phys. Rev. D92 (2015) 072014 [1507.03233].


[7] Belle collaboration, Measurement of the branching ratio of $\bar{B}^0 \rightarrow D^{*+}\tau^-\bar{\nu}_\tau$ relative to $\bar{B}^0 \rightarrow D^{*+}\ell^-\bar{\nu}_\ell$ decays with a semileptonic tagging method, Phys. Rev. D94 (2016) 072007 [1607.07923].

[8] LHCb collaboration, Measurement of the ratio of branching fractions $B(\bar{B}^0 \rightarrow D^{*+}\tau^-\bar{\nu}_\tau)/B(\bar{B}^0 \rightarrow D^{*+}\mu^-\bar{\nu}_\mu)$, Phys. Rev. Lett. 115 (2015) 111803 [1506.08614].

[9] LHCb collaboration, Measurement of the ratio of the $B^0 \rightarrow D^{*-}\tau^+\nu_\tau$ and $B^0 \rightarrow D^{*-}\mu^+\nu_\mu$ branching fractions using three-prong $\tau$-lepton decays, Phys. Rev. Lett. 120 (2018) 171802 [1708.08856].

with a large range of luminosities and radii, from NGC 4605 ($R = 4$ kpc) to UGC 2885 ($R = 122$ kpc), \textit{Astrophys. J.} \textbf{238} (1980) 471.


[27] XENON collaboration, *Constraining the spin-dependent WIMP-nucleon cross sections with XENON1T*, 1902.03234.


List of figures

2.1 Dominant HNL production (a), (b) and decay (c) channels in the energy-range probed by SHiP.

2.2 Sketch of an HNL decay in the SHiP detector.

2.3 SHiP sensitivity to HNL coupled preferentially only to muon neutrinos. The solid and dot-dashed line indicate the uncertainty over the $B_c$ contribution at SHiP energies. The kaon contribution is not included, so below the kaon threshold the sensitivity is underestimated (dashed line). Experimental limits are shaded in grey, while theoretical limits are indicated in the lighter shaded regions and described in the text. Figure taken from reference [40].

3.1 Diagram of the SHiP layout as resulting from the comprehensive design study (CDS).

3.2 Overview of the BDF facility. Taken from reference [41].

3.3 Layout of the target complex. Figure from reference [41].

3.4 Illustration of the muon shield concept. The muon shield is shown in a $z$–$x$ section along the bending plane, and the two colours indicate the opposite field directions. Note that the fields are fully contained in material.
3.5 A single magnet of the SHiP muon shield. The magnet yoke is shown in blue, green and grey to differentiate the regions with approximately uniform field directions and strengths. The coil is shown in copper and field lines of the magnetic circuit are indicated in red. The stray field outside of the magnet is negligible. Figure 3.4, Figure 3.6a, Figure 3.6b and Figure 3.7 show the muon shield configurations as sections through the magnet yokes in $z-x$ and $z-y$. ................................. 42

3.6 Diagrams of the $\text{t}_{\text{p}}$ and previous optimisation muon shields. As in Figure 3.4 the colours indicate the idealised field direction. Sections in $z-x$ and $z-y$ are shown. Note, that the $\text{t}_{\text{p}}$ magnet starts only after the hadron absorber, $\sim 5$ m later than the configuration shown in (b) ................................. 43

3.7 Diagram of $\text{c}_{\text{d}_{\text{s}}}$ muon shield design. As in Figure 3.4 the colours indicate the idealised field direction. Sections in $z-x$ and $z-y$ are shown. ................................. 45

3.8 Schematic of the Scattering and Neutrino Detector ($\text{s}_{\text{n}_{\text{d}}}$). Figure taken from reference [41]. Note, that the magnet yoke is cut away to show the detector within in the side view. ................................. 46

3.9 Schematic diagram of the Hidden Sector Detector ($\text{h}_{\text{s}_{\text{d}}}$) (not to scale). For illustration of a signal candidate, an $\text{h}_{\text{n}_{\text{l}}}$ decaying in the vessel is shown in red. ................................. 48

4.1 Plot of the reconstructed $\text{h}_{\text{n}_{\text{l}}}$ vertices. While the selection imposes an elliptic-cylindrical vessel (selection cuts in red), effectively all reconstructed events lie within a cone within the vessel. The vessel length is about 40 m, and the distance from the target is 70 m. ................................. 55

4.2 $\text{h}_{\text{n}_{\text{l}}}$-acceptance for $M_{\text{h}_{\text{n}_{\text{l}}}} = 1.6 \text{ GeV}/c^2$ relative to the total reconstructed $\text{h}_{\text{n}_{\text{l}}}$ as a function of the distance of the vessel from the target for beauty and charm production channels. The data follows falling exponentials, as one would expect considering the simple argument considering geometry and lifetime presented in Section 4.2.1 ................................. 57

4.3 (a) Magnetic field configuration detailed in the Technical Proposal [17]. $z$ is the distance from the target, while $x$ is the distance from the beam-axis in the bending plane and $y$ is vertical. The acceptance corresponding to this configuration is highlighted in red (cf. Figure 4.2). ................................. 58
4.4 (a) CDS magnet configuration. Compared to Figure 4.3a, the length of the muon shield has been decreased by about 25 m. (b) The acceptance corresponding to this configuration is highlighted in red (cf. Figure 4.2).

4.5 (a) Acceptance for different vessel lengths compared between the configurations shown in Figure 4.3a and Figure 4.4a. The configuration considered in the TP [17] is marked in red. Note, that there is a slight difference in the gain for HNL produced in beauty production. Please also note, that the relative normalisation of $b$- and $c$-produced HNL is arbitrary. (b) Same as above, but using distance from target instead of length as abscissa. Note, that the curves for the two different vessel distances from the target are shifted relative to each other compared to (a).

4.6 Comparison of behaviour seen in Figure 4.5a for different masses. Due to no phase-space being available for heavy masses in charm-produced HNL, results are shown for beauty-production only. Note, that the scales of the ordinates differ. The optimal lengths and maximum gain are tabulated for all masses in Table 4.1.

4.7 Effect of movement of the front of the vessel for a fixed tracker distance from the production point. While it is not surprising that having a shorter vessel further away reduces the acceptance, this plot complements Figure 4.5 to show the behaviour of moving either end while keeping the other fixed.

4.8 Acceptance envelope for HNL decays in the channels (a) $N_{2,3} \rightarrow \mu \pi$, and (b) $N_{2,3} \rightarrow \mu \mu \nu$ projected on to the $z-x$ (blue) and $z-y$ (red) planes. Note the clear conical shape of the acceptance envelope. For 100% some outliers—and the tracks of their daughters, which are being considered here as well—distort the shape, particularly for the partially reconstructed channel. But already for 99% and 95% acceptance, the envelopes are nearly perfectly conical. The procedure for making these plots is described in Section 4.2.4.
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Schematic of the dominant backgrounds: (a) Background due to muon Deep Inelastic Scattering (DIS) in the walls of the experimental hall or the material surrounding the fiducial volume. (b) Background due to random combinations of residual muons. (c) Background due to neutrino interactions in the fiducial volume, and due to interactions in the material surrounding it (not covered in this thesis).</td>
</tr>
<tr>
<td>5.2</td>
<td>Muon envelope for one spill. The histogram is normalised to unity, so each bin is coloured according to the fraction of the total muon flux.</td>
</tr>
<tr>
<td>5.3</td>
<td>Occupancy maps for three SHiP sub-systems to exemplify the evolution of the detector rates throughout the experiment. Note, that these rates are pre-digitisation and before energy cuts, such that the actual rate would be lower. The rates are all within the design specifications for the subsystems. Each detector is shown twice, once including all hits and once only showing muon hits.</td>
</tr>
<tr>
<td>5.4</td>
<td>Artificial distributions of muons simulated in order to study $\phi$-distribution after the muon shield.</td>
</tr>
<tr>
<td>5.5</td>
<td>$\phi$-distribution of muons reaching the tracking stations for the sample shown in Figure 5.4.</td>
</tr>
<tr>
<td>5.6</td>
<td>Comparison of kinematics between simulated muon pairs (blue) and muon pairs generated from toy-mc (red).</td>
</tr>
<tr>
<td>5.7</td>
<td>Comparison of reconstructed final state kinematics for hnl signal (orange) and fake signal candidates from the muon combinatorial background (blue) with no cuts applied. Data courtesy of Alexander Marshall.</td>
</tr>
<tr>
<td>5.8</td>
<td>Momentum versus transverse momentum of muons hitting the decay vessel. A 3 GeV/c momentum cut is applied.</td>
</tr>
<tr>
<td>5.9</td>
<td>The interaction points of all generated DIS events before any cuts are applied, weighted according to the interaction probability.</td>
</tr>
<tr>
<td>5.10</td>
<td>Muon occupancy in cavern walls.</td>
</tr>
<tr>
<td>5.11</td>
<td>Sketch of the geometry for muon interaction products scattering back to the vessel, where $\alpha_i$ define the minimum scattering angles in the $z-x$ plane. Not to scale.</td>
</tr>
</tbody>
</table>
5.12 Minimum angle required to reach vessel, *i.e.* the minimum angle in
the $x$–$z$ plane between the muon direction and the vessel, as shown in
Figure 5.11 showing both the distribution (*blue*) and the cumulative
distribution (*red*). ......................................................... 85

5.13 Energy of particles that could reach vessel. ........................................ 87

5.14 Comparison of occupancy in the upstream veto tagger with and without
the *snd* present. Without it there is an overall increase in occupancy of
factor two, and an increase of $\times 50$ in the central region. ...................... 89

5.15 (a) Rate of muons in the volume available for shielding. Close to the
beam axis in $y$ the muons are swept out least, reducing the space avail-
able. (b) Within the identified free region, the residual rate is minimal.
Figures courtesy of Alexandre Grandchamp. ........................................... 90

5.16 Evolution of the particle rate throughout the shield. Figure courtesy of
Alexandre Grandchamp ............................................................ 91

5.17 Muon hits for the old focus in the upstream veto tagger at the front of
the vessel, illustrating the problem of focused muon beams hitting the
corners of the front of the vessel. .................................................. 92

5.18 Comparison of hit rate (*red*) and *hnl* acceptance (*blue*) for different
foci in $x$. .......................................................... 92

5.19 Same as Figure 5.17 but with the approximate outlines of the front face
of the vessel indicated in *red*. The solid outline corresponds to an $x$-focus
of $-5$ m, while the dashed outline corresponds to $10$ m. ......................... 93

6.1 Distribution of all muons available weighted to one spill. ..................... 96

6.2 Loss function component $\sigma_\mu$ for muons and anti-muons. .............. 99

6.3 Measured uncertainty of loss function for different random seeds and
small geometry changes. Note, that this uncertainty will vary for each
configuration, but the order of magnitude is used to estimate the general
uncertainty. The uncertainty can be approximated with a Gaussian
distribution. The Poisson uncertainty based on the number of muons is
significantly smaller. ................................................................. 101
6.4 Diagram of one quarter of a magnet, indicating the three measurements that define the cross-section. The magnet segment outlines are indicated. A minimum gap for the coils is enforced, if the gap would be too small to accommodate them. The field is symmetric in $x$ and anti-symmetric in $y$, and is taken to be uniform in each magnet segment.  

6.5 Resampling procedure shown step-by-step. See Section 6.3 for a detailed description of the individual steps. (b) and (c) only show added muons. (Continued on page 107)  

6.6 Comparison of spatial distribution of muons incident on the sensitive plane at the $z$-position of the tracking station $T_1$ for the two muon samples.  

6.7 (a) Contours of $f_n$ (noise not shown). The markers show the evaluations of the loss function, clearly showing the convergence to one of the degenerate minima. (b) Convergence of optimisation using Bayesian optimisation using Random Forests for regression (left), and random sampling (right). The transition from random initialisation to optimisation is indicated using the dashed red line.  

6.8 Evolution of loss function for a run using random forests performed to study convergence behaviour and the effect of the weight cut-off. Note, that there is a clear transition after about $1.2 \times 10^4$ evaluations of the loss function, similar to that observed in Figure 6.7b.  

6.9 Evolution of the cumulative minimum of the loss function for the continuous optimisation run.  

6.10 Evolution of the low-field optimisation. Transitions to random forests and to the smaller search space are indicated by vertical lines.  

6.11 Low-field optimum from (a) the side, (b) from above.  

6.12 CDS muon shield configuration from (a) the side, (b) from above.  

6.13 CAD drawing of the resulting shield design including the coils. The colour of the volumes indicates the idealised field direction. Blue: $\pm y$; green: $\pm x$, grey: $\mp y$.  

6.14 Performance of muon shield as evaluated with FLUKA at the beam axis. Taken from reference [41].
6.15 Original momentum and transverse momentum of muons that penetrate the new shield configuration when simulated with the new pre-production. ........................................... 123

7.1 Experimental set-up for the muon flux measurement. Taken from reference [41]. ........................................... 128

7.2 Typical drift time spectrum taken from a single spill of collected data. A constant background of noise is visible. The shape, and in particular the start and end of the spectrum allow measurement of the \(rt\)-relation. ........................................... 130

8.1 Realistic field maps as computed by opera. Taken from reference [67]. 134

8.2 Performance comparison between the ideal (blue) and realistic (red) field maps. Figure from reference [41], simulated and analysed using the procedure developed in Section 5.2. 135

8.3 A conceptual design of the magnet assembly: Welded packs of grain-oriented steel sheets are bolted together (a) into 50 cm thick rectangular modules (b). These modules are then combined to form the muon shield magnets (c). Taken from reference [67]. ........................................... 135

8.4 The \(cds\) muon shield approximated using the rectangular module-assembly presented in Figure 8.3. Taken from reference [67]. 136

8.5 Measured magnetic properties for different joint treatment techniques. Untreated go samples measured parallel and perpendicular to the grain orientation are compared with welded joints without annealing, and after annealing at 500 °C and 700 °C. Taken from reference [67]. 136
List of tables

2.1 Possible portal interactions .............................................. 27
2.2 Table of final states .................................................... 31

4.1 Gain and optimal lengths for the tp and cds shield configurations for different hnl masses and the production mechanisms described .................................................. 59

5.1 Standard selection cuts used for the study of background rejection and signal sensitivity for the cds studies ............................................................................. 69

5.2 Effect of applying the different selection cuts independently on the muon inelastic background events. Numbers correspond to $2 \times 10^{20}$ protons on target. The veto systems comprise the uvt and the sbt. As no events remain after application of the veto requirement, the UL(90%CL) is given instead, as calculated in equation (5.3). ...................................................... 82

5.3 Summary of the muon inelastic background event after applying the offline selections cuts to search for partially and fully reconstructed hnl candidates. Numbers correspond to $2 \times 10^{20}$ protons on target .............................................. 83

6.1 Resulting loss for the discussed configurations with 1.8 T and with 1.7 T and the updated hadron absorber respectively. Estimates of uncertainty are indicated for configurations that were evaluated multiple times. As a rule of thumb $\pm 20\%$ seems to be a good estimate if one can extrapolate between configurations. This is in agreement with the measured uncertainty of the loss function (see Section 6.1 and Figure 6.3). .............................................. 120
### List of acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALP</td>
<td>Axion-like particle</td>
</tr>
<tr>
<td>BAU</td>
<td>Baryon asymmetry of the universe</td>
</tr>
<tr>
<td>BBN</td>
<td>Big-bang nucleosynthesis</td>
</tr>
<tr>
<td>BDF</td>
<td>Beam Dump Facility</td>
</tr>
<tr>
<td>BDT</td>
<td>Boosted decision tree</td>
</tr>
<tr>
<td>BEH</td>
<td>Brout-Englert-Higgs</td>
</tr>
<tr>
<td>BSM</td>
<td>Beyond the Standard Model</td>
</tr>
<tr>
<td>CDM</td>
<td>Cold dark matter</td>
</tr>
<tr>
<td>CDS</td>
<td>Comprehensive design study</td>
</tr>
<tr>
<td>CERN</td>
<td>European Organization for Nuclear Research (Conseil europeen pour la recherche nucleaire)</td>
</tr>
<tr>
<td>DAQ</td>
<td>Data acquisition</td>
</tr>
<tr>
<td>DIS</td>
<td>Deep Inelastic Scattering</td>
</tr>
<tr>
<td>DNN</td>
<td>Deep neural network</td>
</tr>
<tr>
<td>DOCA</td>
<td>Distance of closest approach</td>
</tr>
<tr>
<td>ECAL</td>
<td>Electromagnetic calorimeter</td>
</tr>
<tr>
<td>EM</td>
<td>Electromagnetic</td>
</tr>
<tr>
<td>ESPPU</td>
<td>Update of the European strategy for particle physics</td>
</tr>
<tr>
<td>GAN</td>
<td>Generative adversarial network</td>
</tr>
<tr>
<td>HCAL</td>
<td>Hadronic calorimeter</td>
</tr>
<tr>
<td>HNL</td>
<td>Heavy Neutral Lepton</td>
</tr>
<tr>
<td>HS</td>
<td>Hidden Sector</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>HSD</td>
<td>Hidden Sector Detector</td>
</tr>
<tr>
<td>IP</td>
<td>Impact parameter</td>
</tr>
<tr>
<td>(\Lambda_{\text{cdm}})</td>
<td>Lambda cold dark matter. Standard/concordance model of cosmology</td>
</tr>
<tr>
<td>LDM</td>
<td>Light dark matter</td>
</tr>
<tr>
<td>LFV</td>
<td>Lepton flavour violation</td>
</tr>
<tr>
<td>LHC</td>
<td>Large Hadron Collider</td>
</tr>
<tr>
<td>MC</td>
<td>Monte Carlo</td>
</tr>
<tr>
<td>NP</td>
<td>New Physics</td>
</tr>
<tr>
<td>(\nu_{\text{MSM}})</td>
<td>Neutrino Minimal Standard Model</td>
</tr>
<tr>
<td>PBC</td>
<td>Physics beyond colliders study group</td>
</tr>
<tr>
<td>PID</td>
<td>Particle identification</td>
</tr>
<tr>
<td>PNGB</td>
<td>Pseudo-Nambu-Goldstone boson</td>
</tr>
<tr>
<td>QCD</td>
<td>Quantum chromodynamics</td>
</tr>
<tr>
<td>QED</td>
<td>Quantum electrodynamics</td>
</tr>
<tr>
<td>QFT</td>
<td>Quantum field theory</td>
</tr>
<tr>
<td>SBT</td>
<td>Surrounding background tagger</td>
</tr>
<tr>
<td>SHiP</td>
<td>Search for Hidden Particles</td>
</tr>
<tr>
<td>SM</td>
<td>Standard Model</td>
</tr>
<tr>
<td>SND</td>
<td>Scattering and Neutrino Detector</td>
</tr>
<tr>
<td>SPS</td>
<td>Super Proton Synchrotron</td>
</tr>
<tr>
<td>TP</td>
<td>Technical proposal</td>
</tr>
<tr>
<td>UVT</td>
<td>Upstream veto tagger</td>
</tr>
<tr>
<td>WIMP</td>
<td>Weakly interacting massive particle</td>
</tr>
</tbody>
</table>