Machine learning based background simulations and R-parity violating neutralinos at SHiP

and

Contributions to the angular analysis of $B^0 \to K^{*0}(\to K^+\pi^-)\,\mu^+\mu^-$ at LHCb

By

Alexander Mclean Marshall

School of Physics
University of Bristol

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the Faculty of Science School of Physics.

March 2021

Word count: 32,940

Abstract

The SHiP (Search for Hidden Particles) experiment is a proposed general purpose facility at the new Beam Dump Facility at the Super Proton Synchrotron (SPS) at CERN. The experiment is designed to search for new physics by hunting for as of yet unobserved extremely weakly interacting particles. Hidden particles are predicted by a large number of beyond the Standard Model (SM) theories that attempt to explain inconsistencies between the Standard Model and experimental observations. The SPS will provide SHiP with a high intensity proton beam creating an ideal environment to study rare processes. The experiment is designed to both maximise acceptance to a wide variety of possible signals and to provide an almost background free environment for new physics searches. This thesis presents several studies of the SHiP experiment and its simulation: a fast sampling approach for background simulation is developed using machine learning methods, techniques to suppress the combinatorial muon background rate at SHiP are evaluated and the sensitivity of the SHiP facility to a collection of R-parity violating Supersymmetry signal scenarios is estimated.

In the SM the $b \to s\ell^+\ell^-$ processes can only occur via rare loop diagrams. Measurements at the LHCb experiment of such processes have in recent years revealed tantalising hints of anomalies to the SM. One such example is the angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay, where unaccounted for new particles can contribute to the decay and distort the angular distribution. This thesis presents contributions to the q^2-binned angular analysis of $B^0 \to K^{*0}\mu^+\mu^-$ at LHCb. These contributions allow the simultaneous fitting of all the CP-averaged and CP-asymmetric observables for the first time. Toy studies are carried out to investigate the performance of the simultaneous fit with the data yields expected from 9fb^{-1} of data.

Dedication and acknowledgements

First and foremost I would like to thank my exceptional supervisor, Kostas Petridis. I have always appreciated both your patience and your enthusiasm. You have taught me so much, you've always provided invaluable advice and you truly have made my PhD special. Thank you so much.

I would like to directly thank some members of the SHiP collaboration. Oliver Lantwin for valuable help, especially while I was new to SHiP. Nico Serra for fruitful discussions throughout my PhD and most importantly, introducing me to GANs. Thomas Ruf, Fedor Ratnikov, Alessandra Pastore and Elena Graverini for their contributions to the internal review of the GAN paper, the comments from which markedly improved the study.

To Mitesh Patel, Mark Smith and the rest of the $K^{*0}\mu^+\mu^-$ group who welcomed me late in my PhD and helped me quickly catch up and get to grips with LHCb and this analysis, thank you.

To the very friendly particle physics group at Bristol, thank you for having me. Special thanks to Jonas Rademacker and the rest of the group that contributed to the Graphcore IPU study, and to Graphcore themselves.

Thanks to the Data Intensive CDT for organising my participation in valuable training courses and workshops. I have thoroughly enjoyed being a part of the CDT, the meetup events and my industry placement at EDF Energy.

Finally, I would like to dedicate this thesis to my friends and family. To those that have been a part of my many special memories from the last 4 years and those that couldn't be.

Author's declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

Alexander Mclean Marshall

December 9, 2021

Table of Contents

List of Figures

Figure **Page**

List of Tables

Table **Page**

Abbreviations

ATLAS A Toriodial LHC ApparatuS experiment.

BDF Beam Dump Facility.

BDT Boosted decision tree.

BSM Beyond the Standard Model.

CMS Compact Muon Solenoid experiment.

DIS Deep-inelastic scattering.

DM Dark Matter.

DOCA Distance of closet approach.

DS Decay spectrometer.

FoM Figure of merit.

GAN Generative Adversarial Network.

HNL Heavy Neutral Lepton.

HS The Hidden Sector.

IP Impact parameter.

IPU Graphcore's Intelligence Processing Unit.

KDE Kernel density estimator.

LHCb The Large Hadron Collider beauty experiment.

LLP Long lived particles.

LSP Lightest supersymmetric particle.

MIMD Multiple instruction, multiple data.

NP New physics.

PDF Probability density function.

PID Particle identification.

POT Protons on target.

ROC AUC Area under the receiver operating characteristic curve.

RPC Resistive plate chamber.

RPV R-parity Violating.

SBT Surrounding background tagger.

SHiP The Search for Hidden Particles experiment.

SND Scattering and neutrino detector.

SPS Super Proton Synchrotron.

SST Strawtube spectrometer tracker.

SUSY Supersymmetry.

TD Timing detector.

UBT Upstream background tagger.

VAE Variational autoencoder.

Chapter 1

Introduction

With the discovery of the long proposed Higgs boson at the Large Hadron Collider (LHC) all particles predicted by the Standard Model (SM) have now been observed [1–3]. This observation, along with countless other increasingly precise measurements, continues to add further credence to the SM and its predictions. The SM is however known to be an incomplete description of nature. Observational evidence has highlighted the following three major inconsistencies: Dark Matter (DM), the Baryonic Asymmetry of the Universe (BAU) and the existence of neutrino masses. Neither theory nor experiment has been able to provide any full explanation of these phenomena. While other unorthodox explanations do exist, the most likely case is that some, as of yet, undiscovered beyond the Standard Model (BSM) particles are responsible for these phenomena. The SM, along with some of these observed inconsistencies, and some BSM models are outlined in Chapter 2.

While ongoing experimental measurements conducted by the particle physics community continue to tighten constraints on various proposed sources of new physics (NP), there is no single answer to a question of where best to look. Generally, to search for NP particle physics experimental efforts are each designed to advance one of the following: the *energy frontier*, the *precision frontier* or the *intensity frontier*.

The *precision frontier* is being expanded by experiments such as LHCb. LHCb is designed to exploit the large production cross section of beauty and charm hadrons at the LHC. Precise investigations into rare processes such as $b \to s\ell^+\ell^-$, a process forbidden at tree-level in the SM, can provide sensitivity to NP. For decays so suppressed in the SM any small NP effects can make contributions of similar order to those of the SM. Branching fractions of rare decays can be significantly affected by NP, where such effects could be nonuniversal with NP potentially coupling to each lepton flavour differently. Any NP effects could also manifest in measurable distortions of the angular distributions of decay products [4, 5]. Measurements of $b \to s\ell^+\ell^-$ processes at LHCb have already produced deviations from SM predictions of angular observables and have revealed hints of the violation of the lepton flavour universality assumption present in the SM [6–10]. These hints come from theoretically clean measurements and appear more significant with each iteration of analysis. Looking forward, the next data taking period and the future high-luminosity LHC (HL-LHC) will significantly increase the size of the data sets available for these analyses [11, 12].

The LHC has the highest center-of-mass energy of any particle accelerator at $\sqrt{s} = 13$ TeV. Measurements at CMS and ATLAS are searching for the direct production of BSM particles pushing the *energy frontier*. The sensitivities of these searches are limited by complex backgrounds, including those originating from Quantum Chromodynamics (QCD) [13]. Currently no searches for BSM physics have observed a significant excess over these expected backgrounds [14, 15]. The proposed Future Circular Collider (FCC), if commissioned, will provide sensitivity to even higher mass scales.

Searches for lower mass BSM physics the discovery of which could so far have been hidden behind extremely weak interactions with the SM are just as necessary. The production of feebly coupled low mass NP particles in the lab is most effectively carried out at a fixed target facility. As opposed to a collider experiment like the LHC, fixed target facilities can offer significantly higher intensities, although notably this intensity comes at lower centre of mass energies. It is with this approach that the SHiP experiment is designed to probe the *intensity frontier* for a wide range of NP signals.

The proposed SHiP experiment is a fixed target facility planned for construction at the new

Beam Dump Facility (BDF) at the Super Proton Synchrotron (SPS) at CERN. The components of the facility are discussed in Chapter 3. The aim is to produce long-lived hidden NP particles predominantly through decays of beauty and charm hadrons produced in proton collisions with a dense target. If produced, these particles will travel unimpeded through the facility into a large vacuum vessel where they may decay into SM particles. In order to maximise the production of these heavy hadrons, it is planned that the SPS will provide SHiP with a massive 2×10^{20} protons-on-target (POT) at an energy of ~ 400 GeV over a 5 year run time [16]. SHiP is designed to be a virtually zero background experiment. This is achieved with a variety of techniques, some of which will be explored in this thesis. The most unique of these approaches to background suppression at SHiP is the development of a sophisticated magnetic muon shield, designed to sweep muons out of the acceptance of downstream detectors.

Huge simulation campaigns have been undertaken to verify the performance of this muon shield and to estimate achievable levels of background suppression. Studies have demonstrated that the expected backgrounds at SHiP can be controlled as claimed given some assumptions. These studies are currently limited by background sample size as they rely on the computationally expensive simulation of 400 GeV protons on the dense SHiP target. Chapter 4 describes the development of a machine learning based fast simulation technique for modelling the kinematics of background muons produced in the SHiP target from the 400 GeV SPS proton beam. The speedup achieved can enable more comprehensive background studies. This work was published in Ref. [17]. The introduction to this technique is followed with a detailed study of the combinatorial muon background rate expected at SHiP in Chapter 5. Some of this work was published in Refs. [18] and [19].

Chapter 6 details a sensitivity study of the SHiP experiment to a variety of benchmark R-parity violating neutralino scenarios, both established benchmarks and some previously unexplored scenarios are included. These sensitivities are compared to similar estimates of the sensitivity of other proposed facilities looking to search for similar NP.

Chapter 7 presents some benchmarking results of a new hardware option, the Graphcore® IPU, against a CPU and a GPU option. Benchmarks are carried out on a variety of generative networks taken from particle physics literature. Performance for both training and inference are

investigated. This work was published in Ref 7.1.

Chapter 8 introduces improvements to the fast simulation approach developed in Chapter 4. The new architecture presented shows clear improvements in performance and flexibility. Examples in this chapter show how this new architecture can be useful for a new optimisation of the SHiP muon shield.

Chapter 9 presents an introduction to the q^2-binned angular analysis of $B^0 \to K^{*0}\mu^+\mu^-$ at LHCb, a decay involving the rare $b \to s\ell^+\ell^-$ transition. Modifications made to the fitting procedure in order to simultaneously fit for every angular CP-averaged and CP-asymmetric observable are described. The performance of the fit and the sizes of any biases are assessed using SM toy studies.

Finally, Chapter 10 summarises the work in this thesis and lays out some conclusions.

Chapter 2

Physics background

Following an introduction to the Standard Model (SM) this chapter provides a description of some models of physics beyond the SM (BSM) that are accessible at the SHiP experiment.

2.1 An overview of the Standard Model

This section is written with reference to [20–22].

The Standard Model of particle physics is a rigorously tested quantum field theory (QFT), based on a $\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ gauge group. The construction of the SM relies on the assumption of local gauge invariance, this is that the SM Lagrangian is invariant under any local transformation of the gauge group. The SM describes all the experimentally observed fundamental particles and three of the four fundamental forces, the *strong*, *weak* and *electromagnetic* interactions. The SM has no description of the gravitational force, therefore the SM acts as an effective field theory. Fundamental particles in the SM are split into the spin 1/2 fermions, the building blocks of ordinary matter, and the integer spin bosons, the mediators of the fundamental forces. The strong force has the massless gluon, the electromagnetic force the

Quarks

Particle	q	I^3	Y
d_R	$-1/3$	0	$-2/3$
d_L	$-1/3$	$-1/2$	$+1/3$
u_R	$+2/3$	0	$+4/3$
u_L	$+2/3$	$+1/2$	$+1/3$

Leptons

Particle	q	I^3	Y
ν_{eR}	-	-	-
ν_{eL}	0	$+1/2$	-1
e_R	-1	0	-2
e_L	-1	$-1/2$	-1

Generation 1	*Generation 2*	*Generation 3*
d	s	b
u	c	t

Generation 1	*Generation 2*	*Generation 3*
ν_e	ν_μ	ν_τ
e	ν	τ

Table 2.1: The fermions of the SM with electric charge q, weak hypercharge Y and weak isospin I^3.

massless photon and the weak force has the massive $W^\pm$ and Z^0 bosons. The properties of the SM fermions are displayed in Table 2.1.

The fermions are split into two groups, *quarks* and *leptons*. Each group has three *flavour generations*. Particles in successive generations are identical to the previous generation, just with higher masses.

For the simplest example of what local gauge symmetry looks like, lets start with the Dirac Lagrangian for a free field ψ,

$$\mathcal{L} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi. \tag{2.1}$$

Then consider a local U(1) transformation of the form,

$$\psi \to \psi' = e^{iq\theta(x)}\psi \tag{2.2}$$

and make this substitution into 2.1. We observe a lack of invariance under this transformation,

$$\mathcal{L} \to \mathcal{L}' = \mathcal{L} - q\bar{\psi}\gamma^\mu\psi\theta(x). \tag{2.3}$$

To address this, instead the covariant derivative may introduced into equation 2.1,

$$\partial_\mu\psi \to D_\mu\psi = (\partial_\mu - igA_\mu)\psi, \tag{2.4}$$

where A_μ is a new gauge field and g is a coupling constant. This adds an interaction term to $\mathcal{L}$,

$$\mathcal{L} = i\bar{\psi}\gamma^\mu\partial_\mu\psi + qg\bar{\psi}\gamma^\mu A_\mu\psi - m\bar{\psi}\psi. \tag{2.5}$$

and reinstates local gauge invariance under U(1) transformations. The gauge field A_μ turns out to have the same properties as the photon, coupling to electric charge q with a coupling strength equal to the electron charge, $g = e$. This brings us to the complete Lagrangian of Quantum Electro-dynamics (QED),

$$\mathcal{L} = -\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} + \bar{\psi}(i\gamma_\mu D^\mu - m)\psi, \tag{2.6}$$

where $F^{u\nu}$ is the electromagnetic field strength tensor,

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu. \tag{2.7}$$

Quantum Chromo-dynamics (QCD) is the description of the gluon mediated strong force. The strong force affects only particles with one of three *colour* charges (C) (designated red, green and blue). Local symmetry in QCD is therefore based on the $SU(3)_C$ group, where an SU(3) group is a 3×3 matrix U with $det(\mathrm{U}) = 1$ and $\mathrm{UU}^\dagger = \mathbf{I}$. As a consequence of SU(3) being a non-abelian group, terms arise in the QCD Lagrangian that allow gluon self-interactions. The field strength tensor for example is,

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\mu G^a_\mu + g f^{abc} G^b_\mu G^c_\nu. \tag{2.8}$$

The colour charge is carried by both quarks and the 8 gluonic fields, that arise from the 8 generators of the SU(3) group. Gluon self-interactions confine quarks to bound states, mesons and baryons, all of which must have neutral colour. This implies that individual free quarks cannot exist.

The $W^\pm$ bosons of the weak force carry electric charge, it is however not the weak force but the electromagnetic force of QED that commands interactions between electrically charged particles. This points to some unification. Electroweak Theory is the complete description. The electroweak component of the SM is invariant under transformations of the product group $SU(2)_L \times U(1)_Y$ where L refers to left-chirality and Y to hypercharge, such a group has $3 + 1$ generators. This will lead to 4 vector bosons, with fields W^a_μ, with $a = \{1, 2, 3\}$, and B_μ. The

electroweak Lagrangian is not parity invariant, it is observed that the weak force is maximally parity violating and interacts only with chiral left-handed particles and antiparticles, where chirality and helicity are equivalent for massless particles. Left-handed particles then transform as doublets, χ_L, under $\mathrm{SU(2)}_L\mathrm{U(1)}_Y$,

$$\chi_L = \begin{pmatrix} \nu_L \\ e_L \end{pmatrix}, \qquad \chi_L \rightarrow \chi'_L = \underbrace{e^{iT^a\epsilon^a}}_{\mathrm{SU(2)}_L} \underbrace{e^{iY_L\alpha(x)}}_{\mathrm{U(1)}_Y} \chi_L, \tag{2.9}$$

where the generators of the SU(2) group, T^a, are 2×2 matrices related to the Pauli matrices as $T^a = \frac{1}{2}\sigma^a$. Right-handed particles then transform as singlets, e_R, under $\mathrm{SU(2)}_L\mathrm{U(1)}_Y$. This leads to the following covariant derivatives,

$$\begin{aligned} D_\mu \chi_L &= (\partial_\mu - igW^a_\mu T^a - i\frac{g'}{2}Y_L B_\mu)\chi_L, \\ D_\mu e_R &= (\partial_\mu - i\frac{g'}{2}Y_R B_\mu)e_R. \end{aligned} \tag{2.10}$$

This produces the following four massless bosons $W^\pm_\mu$, W^3_μ and B_μ. The neutral bosons then mix to give the physical Z^0 and A bosons as

$$\begin{aligned} Z_\mu &= cos\theta_W W^3_\mu - sin\theta_W B_\mu, \\ A_\mu &= cos\theta_W B_\mu + sin\theta_W W^3_\mu, \end{aligned} \tag{2.11}$$

where θ_W is the weak mixing Weinberg angle. It is the 3$^{\mathrm{rd}}$ SU(2) generator T^3 that contains the physical weak isospin charges I^3, listed in Table 2.1. Then, QED can be recovered for right- and left-handed particles using the relation,

$$q = \tfrac{1}{2}Y + I^3, \tag{2.12}$$

leading to definitions of weak hypercharge, Y, also listed in Table 2.1.

The masses of the $W^\pm$ and Z^0 bosons are obtained via the Higgs mechanism and spontaneous symmetry breaking (SSB). We add ϕ, as a doublet of complex scalar fields with a potential $V(\phi)$ of

$$V(\phi) = \mu^2(\phi^\dagger\phi) + \lambda(\phi^\dagger\phi)^2, \tag{2.13}$$

which for values of $\mu^2 < 0$ and $\lambda > 0$ has a non-zero minima, or vacuum expectation value, v, a requirement for SSB. Note, in order to maintain Lorentz invariance a non-zero v is only allowed

for scalar fields. This potential has a degenerate minima in the complex plane of ϕ occurring at

$$\phi = \frac{\mu}{\sqrt{2\lambda}} = \frac{v}{\sqrt{2}}, \qquad \text{where} \quad v = \frac{\mu}{\sqrt{\lambda}}. \tag{2.14}$$

Applying the unitary gauge simplifies the problem and avoids the massless Goldstone bosons that arise in solutions without this choice of gauge. Now, including a perturbation, H,

$$\phi = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} + H \end{pmatrix}, \tag{2.15}$$

ϕ can enter kinetic terms of the Higgs Lagrangian of the form $(D_\mu \phi)^\dagger D^\mu \phi$. From this and 2.13 one can extract mass terms for 3 massive bosons as,

$$m_W = \frac{gv}{2}, \qquad m_Z = \frac{v\sqrt{g^2+g'^2}}{2} = \frac{m_W}{cos\theta_W}, \qquad m_H = \sqrt{2\mu^2} = v\sqrt{2\lambda}. \tag{2.16}$$

The masses of fermions are provided similarly, by inserting 2.15 into Yukawa terms such as, for leptons,

$$-Y_e^{fg} \bar{\chi}_L^f \phi e_R^g + h.c. \qquad \text{with} \quad \chi_L = \begin{pmatrix} \nu_L \\ e_L \end{pmatrix}, \tag{2.17}$$

where f and g run over flavours and generations. The Yukawa couplings are the only part of the SM for which particle flavour is not degenerate, elsewhere the SM has flavour universality.

At tree-level the weak interaction allows quarks to change flavour only via charged current interactions mediated by the $W^\pm$ boson. The Yukawa terms for three generations of quarks are

$$-Y_D^{fg} \bar{Q}_L^f \phi D_R^g - Y_U^{fg} \bar{Q}_L^f \phi^c U_R^g + h.c., \tag{2.18}$$

where the 3×3 matrices Y_D^{fg} and Y_U^{fg} are in general not diagonal, this is allowed as $\bar{Q}_L$, U_R^g and D_R^g are flavour eigenstates not physical mass eigenstates. Diagonalising these matrices leads to the Cabibbo–Kobayashi–Maskawa matrix V_{CKM} that encodes the strength of flavour changing currents between physical quark states,

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{tb} & V_{ts} \end{pmatrix}. \tag{2.19}$$

There are only 4 physical parameters of V_{CKM}, three mixing angles and one complex phase. This complex phase is the only place CP-violation is included in the SM.

So putting everything together the full Lagrangian of the SM before SSB is as follows

$$\mathcal{L} = \mathcal{L}_{bosons} + \mathcal{L}_{fermions} + \mathcal{L}_{Yukawa} + \mathcal{L}_{Higgs}, \tag{2.20}$$

where

$$\begin{aligned}
\mathcal{L}_{bosons} &= -\frac{1}{2}TrG_{\mu\nu}G^{\mu\nu} - \frac{1}{2}TrW_{\mu\nu}W^{\mu\nu} - \frac{1}{2}TrB_{\mu\nu}B^{\mu\nu}, \\
\mathcal{L}_{fermions} &= \sum_{\psi=\chi_L,e_R,Q_L,U_R,D_R} i\bar{\psi}\gamma^{\mu}D_{\mu}\psi, \\
\mathcal{L}_{Yukawa} &= \sum_{f,g}\left[-Y_e^{fg}\bar{\chi}_L^f\phi e_R^g - Y_D^{fg}\bar{Q}_L^f\phi D_R^g - Y_U^{fg}\bar{Q}_L^f\phi^c U_R^g\right], \\
\mathcal{L}_{Higgs} &= (D_{\mu}\phi)^{\dagger}D_{\mu}\phi - \lambda\left(\phi^{\dagger}\phi - \frac{v}{2}\right).
\end{aligned} \tag{2.21}$$

2.2 Beyond the Standard Model (BSM) observations

2.2.1 Dark Matter (DM)

There are many astronomical observations that infer the presence of non-baryonic dark matter (DM) in our Universe [23]. For example, the presence of DM can explain the disparity between measured galactic rotation curves and Keplarian models that account for all the visible matter in galaxies [24]. Observations of gravitational lensing have produced measurements of mass to light ratio of galaxies, which indicate a ~ 5 to 1 excess [25]. Observations of anisotropies in the cosmic microwave background (CMB) provide the most precise measurements for the DM abundance indicating that $\sim 85\%$ of matter in our Universe is made up of DM [26, 27]. All these streams of evidence for DM investigate only gravitational effects. The particle properties and the extent to which DM interacts with ordinary matter via the other forces remains unclear. Since we observe DM in the Universe today it is accepted that DM is stable or is very long-lived due to some feeble coupling to the SM. Whether DM self-interacts or not remains unknown. However, observations of mass distributions of galactic collisions, such as the bullet cluster [28], can provide limits on the DM self-interaction cross section.

The abundance of DM observed in the Universe today may have been dictated by the conditions during *freeze-out*. In the hot early Universe DM production and annihilation occurred

at equal rates. As the Universe expanded two things happened, the Universe cooled ceasing DM production and concurrently the probability for DM annihilation reduced. The timing of freeze-out and produced relic density of DM depend on both the DM mass and strength of any coupling to the SM [13, 29].

A wide variety of well motivated new physics models contain particles which could be candidates for DM. These candidates span large mass ranges and require vastly different experimental designs to detect. Efforts fall into three categories. Indirect detection, which involves the detection of the products of DM annihilation events. Experiments look for gamma ray signals from regions of high DM density [30]. Direct detection, which involves large experiments that are hyper-sensitive to nuclear recoils caused by collisions with DM particles passing through large fiducial volumes [31, 32]. Finally, DM could be produced in particle accelerators. The LHC is searching for DM via missing transverse energy techniques, the LHC with a high center of mass energy is pushing the energy frontier [13]. The SHiP experiment on the other hand would be searching for DM candidates at the intensity frontier, looking for lower mass candidates at unexplored weaker couplings.

2.2.2 Baryonic asymmetry (BAU)

Within the SM there exists minimal matter-anti-matter asymmetry, the only source of CP-violation comes from the complex phase of V_{CKM}. However, the Universe we observe today is matter dominated. To produce an asymmetry of this scale the Sakharov conditions must be satisfied. These are: the presence of baryon number violation, the presence of both C-violation and CP-violation, and that the interaction responsible for the baryon-asymmetry must occur out of thermal equilibrium.

The amount of CP-violation observed within the quark sector is not nearly enough to account for the asymmetry observed in the Universe [33–37]. Indications of CP-violation within the neutrino sector could provide the missing source of CP-violation. Future neutrino experiments will be able explore this further [38, 39].

2.2.3 Neutrino masses

Observations of propagating neutrinos oscillating between flavours can only be explained with the presence of neutrino masses [40]. Individual neutrino masses remain unmeasured, however limits have been placed on the sum of their masses and the mass gaps between neutrinos Δm^2. The next generation of neutrino experiments will provide new measurements constraining Δm^2 and the hierarchy of neutrino mass gaps [38, 39].

Neutrinos are massless in the SM. The Dirac mass terms of the SM require both left and right handed chiral states. A right handed neutrino can be added to the SM Lagrangian but it must be *sterile*, interacting weakly enough with the SM such that it has so far evaded discovery [41]. Unlike the other fermions of the SM, neutrinos have no electric charge. This also allows mass to enter via Majorana mass terms, where Majorana particles are their own anti-particles. However, neither mechanism would provide an explanation of the large difference in mass between the neutrinos and the charged fermions of the SM. Various *seesaw* mechanisms introduce one or more right handed neutrinos and both Dirac and Majorana mass terms [42–44]. The smallness of the SM neutrino mass can be obtained in models with much heavier right-handed neutrinos, see Section 2.3.3.

2.3 The Hidden Sector

This section will introduce various possible extensions to the SM that can be searched for at the SHiP experiment however the signal studies presented in this thesis will focus on R-parity violating neutralinos introduced in Sec. 2.3.5. Hidden Sector (HS) is an umbrella term for a set of proposed BSM particles that cannot interact directly with the SM. HS particles interact, very feebly, with the SM via portals [45].

2.3.1 Vector portal

The introduction of one or more additional U(1) symmetries into the SM gauge group would introduce new heavy vector states. For scenarios in which these states couple strongly to the SM the LHC has already placed strong constraints on existence. The possibility of light (GeV

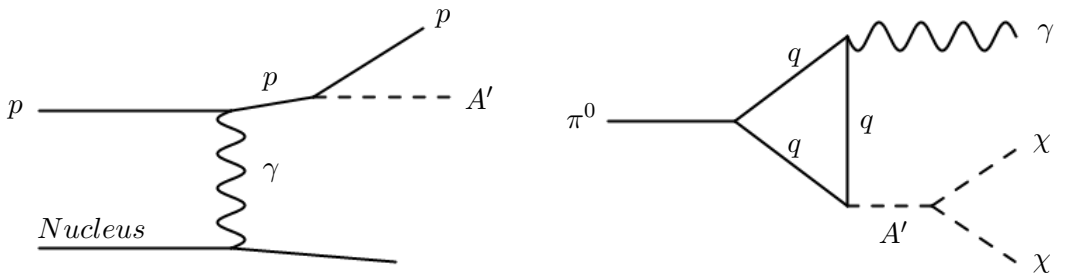


Figure 2.1: Dark photon production via proton Bremsstrahlung and neutral meson decay.

scale) vector state(s) existing with weak couplings to the SM is relatively unexplored. As an example, a model with a single additional U′(1) symmetry could introduce a new gauge boson, the dark photon A' [46]. The dark photon can acquire mass if the U′(1) symmetry is broken via a Higgs-like mechanism. With a field strength tensor F', A' could be produced at SHiP via kinetic mixing with SM photons produced as Bremsstrahlung radiation or in meson decay with an interaction term like,

$$\epsilon F'_{\mu\nu} F_{\mu\nu}, \tag{2.22}$$

where ϵ is a dimensionless coupling. See Fig. 2.1 for diagrams of these production modes. The dark photon then could then decay to SM particles detectable at SHiP for example,

$$A' \to e^- e^+, \qquad A' \to \mu^- \mu^+, \qquad A' \to \text{hadrons}, \tag{2.23}$$

where the decay to electrons dominates for $m_{A'} < 2m_\mu$. The dark photon could also decay to a light dark matter particle, χ, which could be detected via A' mediated elastic scattering, see Ref. [47].

2.3.2 Scalar portal

The discovery of the Higgs boson has motivated the search for other scalar particles. For example a heavy scalar Hidden Sector singlet, S, that could couple to the SM via the square of the Higgs field with an interaction term like,

$$(\alpha_1 S + \alpha S^2) H^\dagger H, \tag{2.24}$$

where α_1 and α are coupling constants. Such a scalar particle can be light and extremely weakly coupled to the SM, and therefore potentially detectable at SHiP [46]. At SHiP the dominant

channels for S production will be processes like,

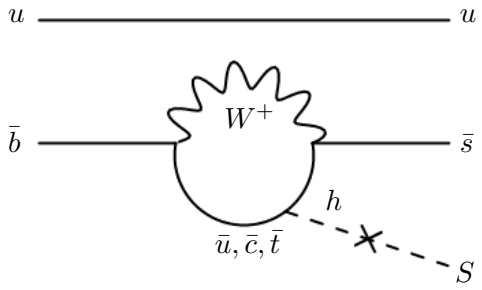


Figure 2.2: Production of S via $B \to S + K$.

$$K \to S + \pi, \qquad D \to S + \pi, \qquad B \to S + K, \tag{2.25}$$

occurring via the α_1 coupling, see Fig. 2.2. The α_1 coupling dominates when the Higgs is off-shell [48]. Of the channels in equation 2.25 the B production channel is dominant at SHiP as the D channel is CKM suppressed and although SHiP will produce $\mathcal{O}(10^{20})$ kaons, these are for the most part stopped in the hadron absorber. The width of the D channel looks like,

$$\Gamma(D \to S + \pi) \sim m_b^2 |V_{cb}^* V_{ub}| \tag{2.26}$$

which is heavily suppressed by the size of V_{ub} [49]. This production occurs via a loop process, similar to that of Fig. 2.2, hence the m_b dependence in equation 2.26. Decays occur to final states such as $e^- e^+$, or $\mu^- \mu^+$, $\pi^- \pi^+$, $K^- K^+$ for heavier scalars [48].

2.3.3 Neutrino portal

The neutrino portal introduces three (or one in the most minimal model) right handed massive neutrinos or Heavy Neutral Leptons (HNLs). The HNLs, N_i, are gauge singlet fermions, and couple to the SM as

$$F_{\alpha i}(\bar{L}_\alpha \cdot \phi^\dagger) N_i + h.c. \,, \tag{2.27}$$

where L_α is an $\mathrm{SU}(2)_L$ lepton doublet, $F_{\alpha i}$ is the corresponding Yukawa coupling and ϕ is the Higgs doublet, $\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}$. Then, after SSB, the active SM neutrino mixes with N_i.

The HNL mass term, just like the SM neutrinos in the presence of right handed chiral partners, can be Dirac or Majorana. An extension to the SM with two HNLs can provide a

solution to the BAU, through the process of leptogenesis [50, 51]. With three HNLs, {N_1, N_2, N_3}, one can also obtain a DM candidate. The SHiP experiment is sensitive to a scenario where the N_2 and N_3 have degenerate masses of between $1 - 10$ GeV/c^2 and N_1 is significantly lighter with a mass of $\mathcal{O}(10)$ keV/c^2, where N_1 is the stable DM candidate. This is well motivated, see Refs. [52–54]. While N_2 and N_3 are feebly coupled to the SM, N_1 is even more so, in accordance with DM observations [41, 45, 55]. Individual interaction strengths relative to the SM neutrinos depend on the mixing angles U_e, U_μ and U_τ, where $U \ll 1$ [18]. The extension of the SM in this way with three HNLs is the Neutrino Minimal SM (νMSM).

At SHiP the dominant production channels of HNLs would be through weak decays of heavy mesons. Accessible HNL decay channels depend on the mass of $N_{2,3}$ but the most relevant at SHiP are for example,

$$N_{2,3} \to \nu e^+ e^-, \qquad N_{2,3} \to \mu^- \pi^+, \qquad N_{2,3} \to \nu \pi^0. \tag{2.28}$$

Figure 2.3 presents diagrams for both production and decay processes with the coupling to the Higgs vacuum expectation value v included.

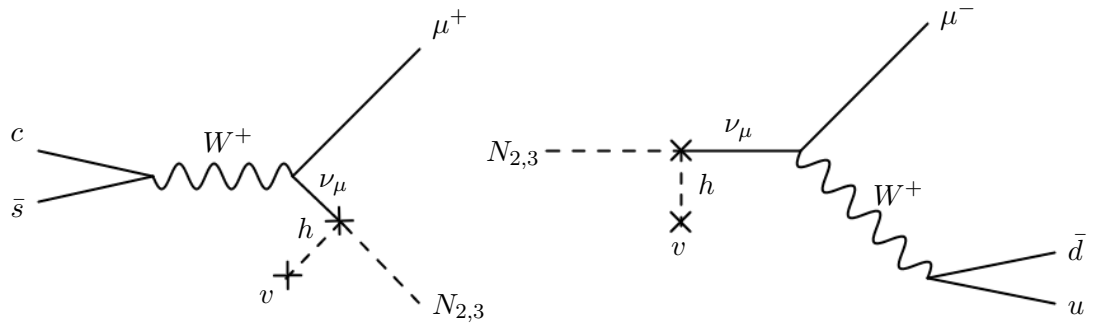


Figure 2.3: HNL production via $D_s^+ \to N_{2,3} + \mu^+$, and decay via $N_{2,3} \to \mu^- + \pi^+$.

2.3.4 Pseudo-scalar portal

Another extension to the SM is the pseudo-scalar portal. This involves the addition of the Axion-like particle (ALP), arising as the pseudo-Nambu-Goldstone bosons of a non-exact spontaneously broken U(1) symmetry. Both the mass and the couplings of ALPs to the SM are suppressed by the scale of the spontaneous symmetry breaking. Thus, for some values of this scale ALPs are

detectable at SHiP. Production could occur at SHiP from two photons via the Primakoff effect, and decay also also possible to two photons, $ALP \rightarrow \gamma\gamma$ [56].

2.3.5 Supersymmetric portal

Supersymmetry (SUSY) is a theoretically motivated and popular extension of the Standard Model able to address some of its shortcomings [57–60]. Supersymmetry adds a fermionic superpartner for every SM boson, and a bosonic superpartner for every SM fermion. Lack of observations at SM masses suggest SUSY would be a broken symmetry and the superpartners of SM particles must have much higher masses. To fix the hierarchy problem, the lack of explanation for the fine-tuning required in the SM to acquire the observed Higgs mass, superpartners would be expected in the 100 GeV/c^2 to few TeV/c^2 range [46].

However, all searches for SUSY so far have shown no evidence of its existence. Since Run 2 of the LHC, direct searches have provided mass limits of the order of $\mathcal{O}(1)$ TeV/c^2 for strongly interacting squarks and gluinos [61, 62]. In contrast, the parameter space for alternative SUSY models with highly suppressed couplings to the SM remains largely unexplored, particularly for SUSY particle masses well below the TeV scale.

Most SUSY models have a stable lightest supersymmetric particle (LSP), the lightest neutralino. In these models the LSP can act as a DM candidate [63]. Stability is provided by the conservation of a new quantum number, R-parity, introduced as

$$R_P = (-1)^{3B+L+2S}, \tag{2.29}$$

where L is lepton number, B is baryon number and S is spin. Every SM particle has $R_P = +1$, and every supersymmetric partner has $R_P = -1$. Therefore, in R-parity conserving SUSY sparticles must be produced in pairs. This maintains proton stability, in line with observational evidence and ensures stability of the LSP.

An LSP with a mass less than 10 GeV/c^2 is allowed, however to avoid cosmological bounds such neutralinos must decay through R-parity violating couplings [64, 65]. The LSP could therefore no longer act as a DM candidate. A SUSY model that contains a DM candidate is attractive, but not required. Other models such as ALPs could still provide a DM candidate. In

order to avoid existing experimental constraints RPV neutralinos of mass less than 10 GeV/c^2 must be purely bino-like, $\tilde{B}^0$ [63].

In contrast to R-parity conserving supersymmetric theories, R-parity violation introduces an extra term in the superpotential that can be written as [46, 66]

$$W_{RPV} = \lambda_{ijk} L_i L_j \bar{E}_k + \lambda'_{ijk} L_i Q_j \bar{D}_k + \lambda''_{ijk} \bar{U}_i \bar{D}_j \bar{D}_k + \kappa_i L_i H_u. \tag{2.30}$$

Here the SU(2)$_L$-doublet fields L, Q and H denote the lepton, the quark and the Higgs chiral-superfields. The SU(2)$_L$-singlet fields E, U and D denote the lepton, up-type quark and down-type quark chiral-superfields. The factors λ, λ' and λ'' denote fully-leptonic, semi-leptonic and hadronic coupling strengths respectively. The indices i, j and k run over generation space. For instance, for the semi-leptonic coupling strength λ', the i index indicates lepton generation, and j and k indicate quark generations; for charged mesons this corresponds to a quark content of $(u_j, \bar{d}_k)$ and for neutral mesons $(d_j, \bar{d}_k)$.

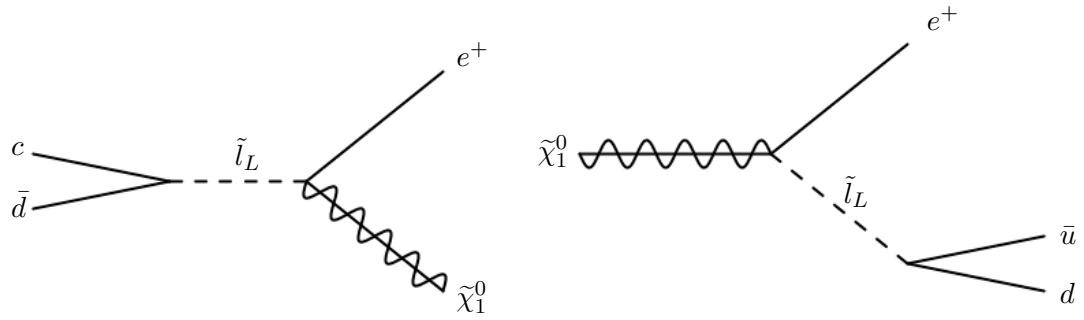


Figure 2.4: Neutralino production via $D^+ \rightarrow e^+ + \tilde{\chi}_1^0$ occurring through the coupling λ'_{121} and decay via $\tilde{\chi}_1^0 \rightarrow e^+ + \pi^-$ occurring through the coupling λ'_{112}.

To maintain the stability of the proton, fully hadronic λ''_{ijk} terms are forbidden and baryon number is conserved. The other two terms, $L_i L_j \bar{E}_k$ and $L_i Q_j \bar{D}_k$, do however introduce lepton number violation, so couplings must be sufficiently small in order to evade experimental constraints [57].

The semileptonic coupling $L_i Q_j \bar{D}_k$ with coupling strength λ'_{ijk}, allows purely bino-like neutralinos to be produced through $M^\pm \rightarrow \tilde{\chi}_1^0 \ell^\pm$ and $M^0 \rightarrow \tilde{\chi}_1^0 \bar{\nu}_\ell$ transitions, see Fig. 2.4, where $M^{\pm,0}$ denotes a charged or neutral meson, see Ref. [67] for formalism. In turn, the coupling $L_l Q_m \bar{D}_n$ with strength λ'_{lmn} enables the neutralino to decay through $\tilde{\chi}_1^0 \rightarrow M'^\pm \ell^\mp$

and $\widetilde{\chi}_1^0 \to M'^0 \nu_\ell$ transitions, again see Fig. 2.4. Additionally, the leptonic coupling, $L_i L_j \bar{E}_k$ with strength λ_{ijk}, results in neutralinos decaying through the process $\widetilde{\chi}_1^0 \to \ell^+ \ell^- \nu_\ell$. Some benchmark scenarios of non-zero λ'_{ijk} and λ_{ijk}couplings are explored in Chapter 6 in which the SHiP has sensitivity to multiple orders of magnitude weaker couplings than those excluded by existing bounds presented in Ref. [68].

Chapter 3

The SHiP experiment

The SHiP experiment [16] is part of a new general purpose fixed target facility proposed at the CERN Super Proton Synchrotron (SPS) accelerator to search for long-lived exotic particles with masses between a few hundred MeV/c^2 and a few GeV/c^2. These feebly interacting particles are predominately expected to be produced in the decays of heavy hadrons. The SHiP facility is therefore designed to maximise both the production of charm and beauty mesons and the detection efficiency of their decay products, while maintaining the lowest possible background rate. The 400 GeV proton beam, extracted from the SPS, will be dumped on a high density target with the aim of accumulating 2×10^{20} protons on target during 5 years of operation. The charm production at SHiP will exceed that of any existing or planned facility.

The SHiP detector incorporates two complementary detector apparatuses, the Scattering and Neutrino Detector (SND), and the HS Decay Spectrometer (DS). The SND will be used to search for light dark matter particles, and to perform tau neutrino measurements. The DS is designed to observe the decays of hidden sector particles by reconstructing their decay vertices within a 50 m long decay volume. The DS comprises a magnetic spectrometer, background vetoing systems and particle identification detectors.

Such a setup will allow the SHiP experiment to probe a variety of models that contain light

long-lived exotic particles at an unprecedented sensitivity. This chapter presents a summary of the main components and subsystems of the SHiP facility before introducing the SHiP simulation suite and expected backgrounds to HS searches.

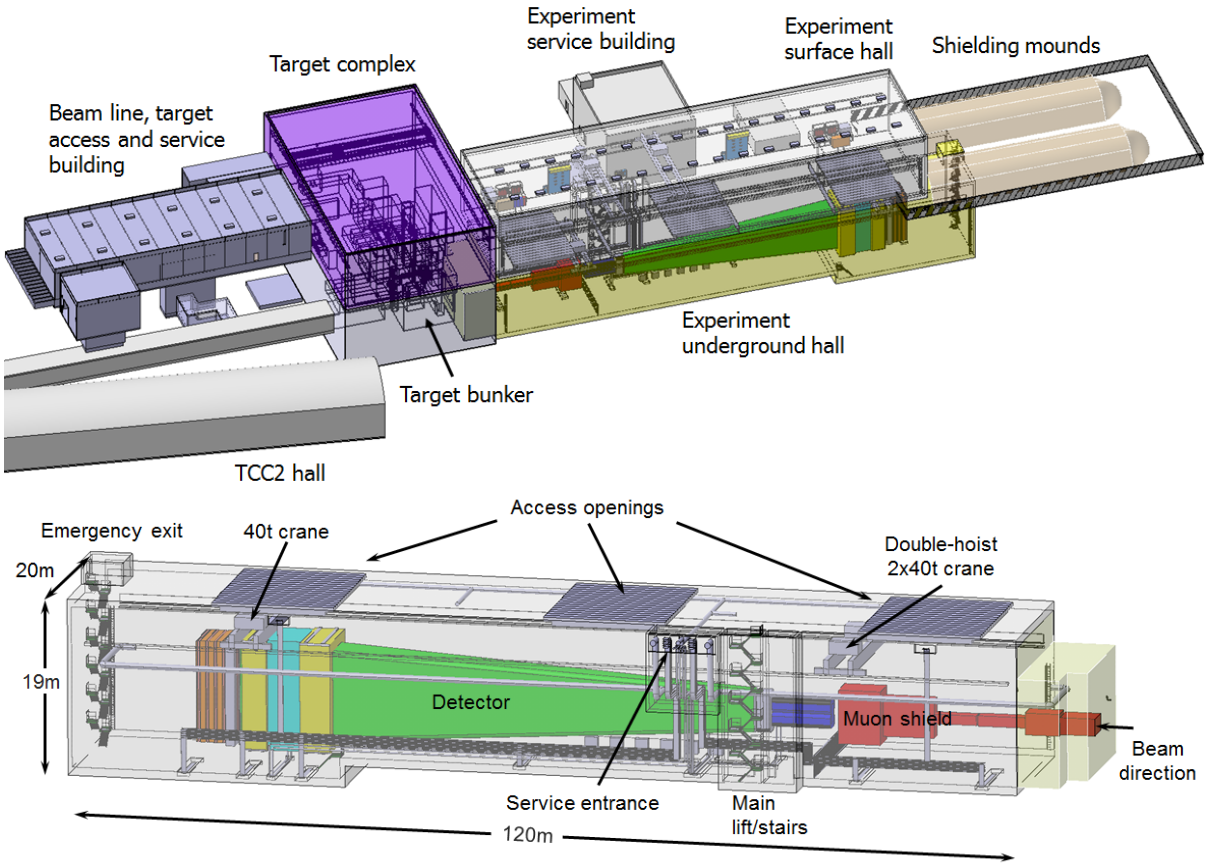


Figure 3.1: Engineering schematics of the SHiP experiment, presented in Ref. [18].

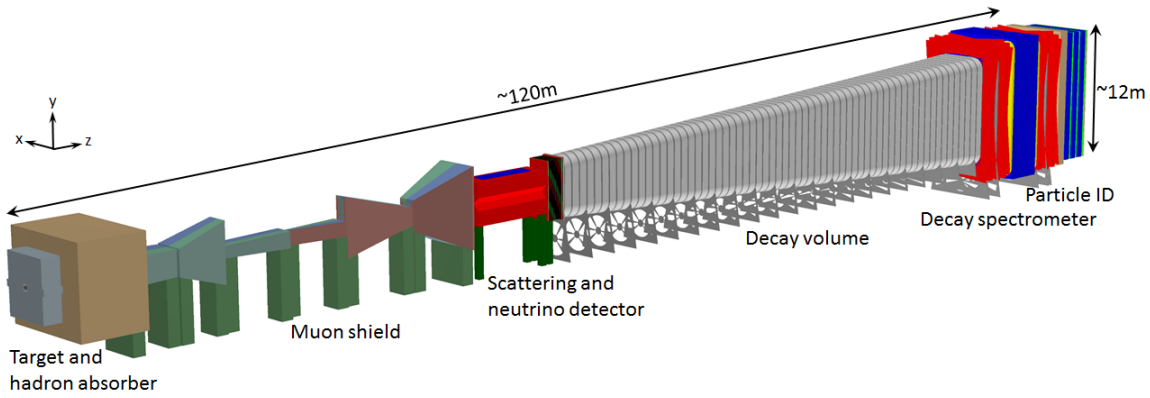


Figure 3.2: Layout of the SHiP experiment, presented in Ref. [18].

3.1 Components of the SHiP experiment

3.1.1 The Beam Dump Facility (BDF) and the target complex

The proposed BDF, planned for construction in the north area of CERN, will house the SHiP experiment [69]. A schematic of the SHiP experiment is provided in Fig. 3.2 and some engineering drawings of the wider facility are provided in Fig. 3.1. The design of the facility is based on both introducing minimal modification to the existing SPS facility and providing maximum reusability of the facility for future experiments to come after SHiP. The BDF includes all of the new beamline, connections to the existing beamline that already supplies the north area and the SHiP target complex. Figure 3.3 gives a sense of where the BDF and SHiP will fit into the existing North Area at CERN. The SHiP facility hall is only $\mathcal{O}(10)$ m deep, this makes radiation safety extremely important. Figure 3.4 displays the design of the shielding that surrounds the SHiP target.

Protons are taken from the SPS via a slow extraction. The goal of this approach is to create spills of uniform intensity, reducing the probability for the occurrence of dangerous muon combinatorial background events. Current expectations are that the BDF will indeed be able to deliver roughly flat-topped 1 s spills of 4×10^{13} 400 GeV protons over 7.2 s cycles [18]. Physics studies of the SHiP experiment assume a full yield over 5 years of 2×10^{20} protons.

At the centre of the SHiP target complex is a dense target constructed from titanium-zirconium doped molybdenum alloy (TZM) blocks and pure tungsten blocks. The target is water cooled via 5 mm wide channels passing between target blocks. The materials used and geometry of the target itself are chosen to maximise the production of charm and beauty hadrons in accordance with the physics goals of SHiP. The current target design is twelve interaction lengths long and the cross-sectional area is large enough to contain full hadronic showers with minimal leakage. These properties enable a large boost in the expected heavy meson production at SHiP from production from cascade interactions within the target. A factor of $\times 2.3$ for charm and $\times 1.7$ for beauty mesons is estimated [70].

To reduce peak heating of the target, and to avoid premature failure, the impact of the SPS proton beam is diluted via circular magnetic sweeping. The radius of this sweeping (50 mm) and

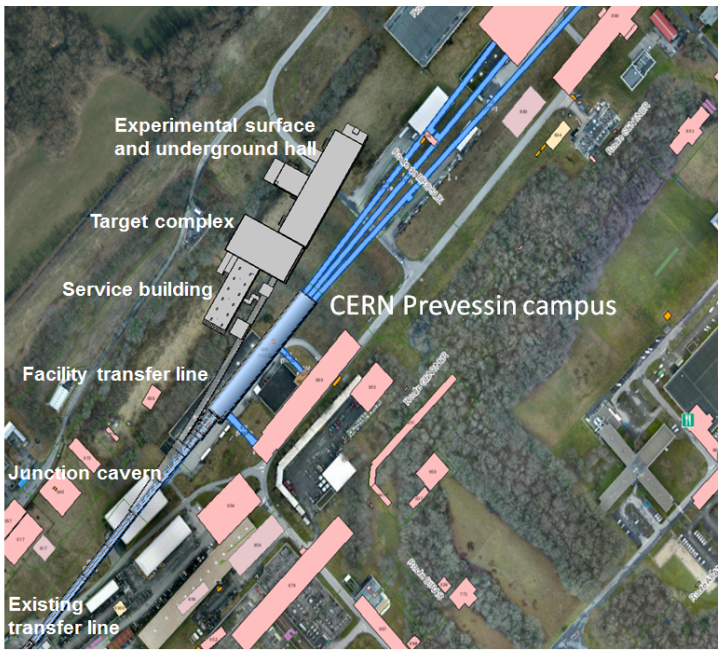


Figure 3.3: Overview of the Prevessin campus, the North Area at CERN, with the new installments planned for SHiP experiment. Figure from Ref. [18].

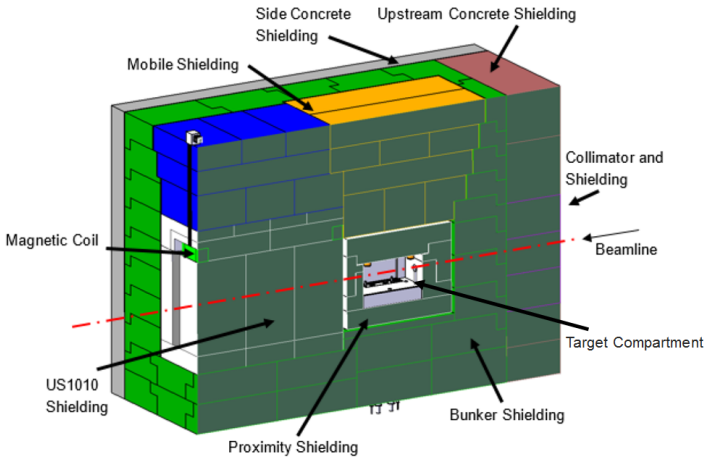


Figure 3.4: Schematic of the target complex including the magnetised hadron absorber and the design of the radiation shield. Figure from Ref. [18].

designed beam width (8 mm) are chosen to maximally reduce impact on the target whilst also observing restrictions of upstream beam-line elements. The resulting temperature distribution after a spill of the SPS is depicted in Fig. 3.6. The aforementioned uniform distribution of beam intensity, achieved via slow extraction from the SPS, also aids reducing peak stress on the target.

Directly down stream of the target is a 5 m long iron hadron absorber. While both the hadron

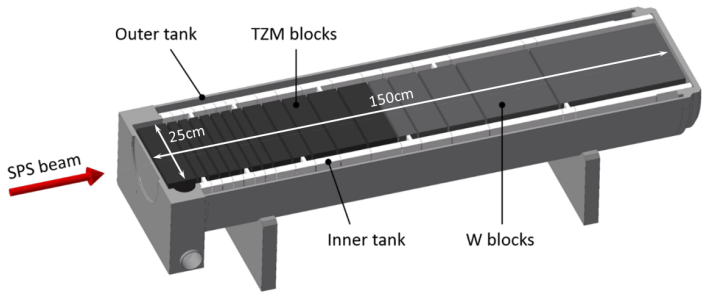


Figure 3.5: Cross-sectional view of the SHiP target. Figure from Ref. [18].

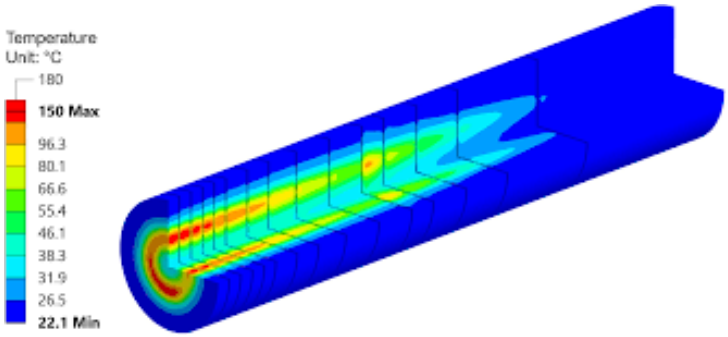


Figure 3.6: Temperature and smearing Figure from Ref. [71]. Water cooling slits are visible.

absorber and dense target are designed to absorb background particles inevitably huge numbers of muons and neutrinos escape. Particularly important then is the softening of these light mesons (π, K) before they decay to background inducing muons. This softening aids the muon shield in reducing the number of muons entering the acceptance of the downstream detectors.

The final 4 m of the hadron absorber is magnetized with a field of $\sim$ 1.6 T. The hadron absorber acts as the first element of the all important active muon shield, discussed in Section 3.1.2.

3.1.2 Active muon shield

Since signal particles originating in charm and beauty meson decays are produced with significant transverse momentum with respect to the beam axis, to maximise acceptance the detector should be placed as close as possible to the target [16]. A critical component of the optimisation of the SHiP experiment is therefore the muon shield, which is by far the largest component between the target and decay vessel [72]. The shield is designed to sweep the high flux of muons

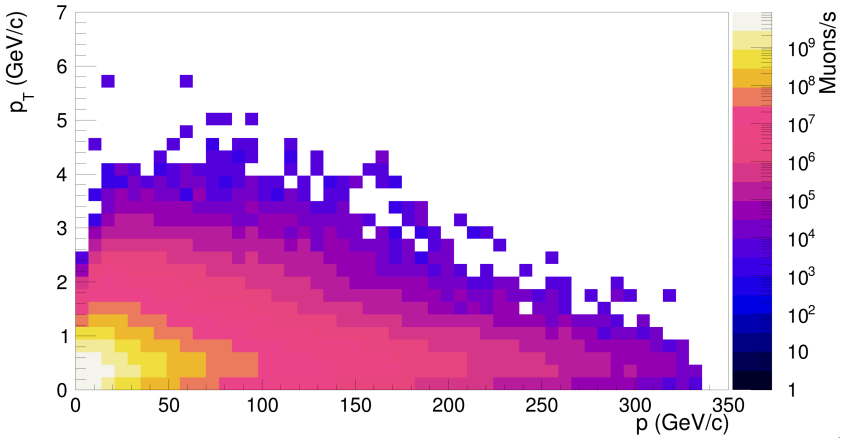


Figure 3.7: Muons produced in the target as a function of momentum and transverse momentum as produced by `Pythia` [72, 73].

produced in the target away from the down stream detectors. The shield must operate with an extremely high muon reduction rate as muons represent a serious background for hidden particle searches. Its task is complicated by the large range of kinematics of muons produced in the target, the shield must be able to clear the vast majority of muons with $0 < p \lesssim 350$ GeV/c and $0 < p_{\mathrm{T}} \lesssim 8$ GeV/c, see Fig. 3.7.

Initially the SHiP collaboration investigated a passive shielding approach using solid iron blocks. However, simulation studies showed that this approach would require too much space to achieve the required performance. A more compact but more complicated solution is the active muon shield.

The active muon shield is a ~ 30 m long configuration of large electromagnets. The magnetised portion of the hadron absorber mentioned in Section 3.1.1 acts as the first element of the shield. Downstream of this there are six free standing elements shown in Fig. 3.2 and 3.8. The general concept of the arrangement is to have the first half of the magnets of one polarity to sweep muons out of acceptance. Then, as some muons will be bent back into acceptance by return fields, the later magnets have the opposite polarity to again sweep away any returning muons. A small charge asymmetry is expected in muons produced, of muons that pass the hadron absorber in `Pythia` simulation $\sim 55\%$ have positive charge.

The shield has been designed assuming that employing large 1.7 T warm (not superconducting) magnets constructed from Grain-Orientated (GO) steel is achievable, a realistic field

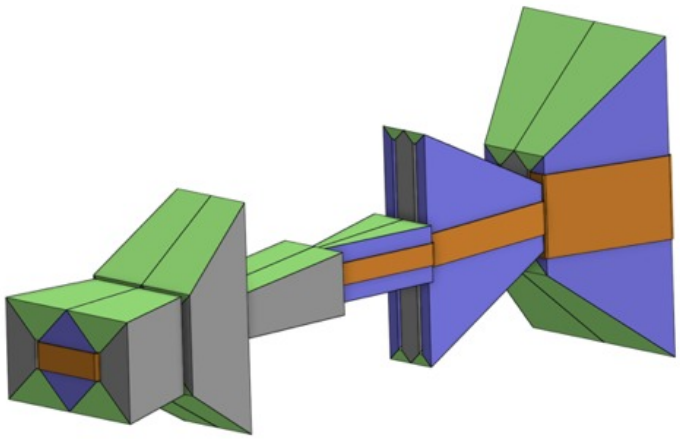


Figure 3.8: CAD model of the optimized active muon shield, regions of contrasting magnetic field polarisation are indicated by the blue, green and grey segments and the coils of each magnet are coloured orange. Figure from Ref. [18].

map is provided in Fig. 3.9. GO steel, over regular steel, offers a higher magnetic flux density for the same current. This comes at the extra cost of the GO steel and added complications of joining GO steel which is manufactured in sheets, the practicalities of this are being investigated. It is important the sheets, which are of size $\mathcal{O}(10)$m are welded together precisely to obtain the maximum stacking factor and minimal loss of magnetic flux and field strength. Prototyping by the collaboration has shown required fields are achievable, although the magnets will ultimately have to assume step-like modular shapes due to engineering constraints.

For the shield to achieve the required performance whilst minimising the total length and weight (a simplistic proxy used for the total cost of the shield) an initial optimization campaign was undertaken [74]. This involved parameterising a baseline configuration where the system was described with $\mathcal{O}(50)$ parameters that encoded a number of the characteristics of the shape of each of the six magnet elements. Bayesian optimization techniques were used to iteratively improve this baseline configuration. For each iteration, an enhanced distribution of realistic muon kinematic distributions was simulated passing through the shield. Muons that enter the downstream detector the shield were then fed into a bespoke loss function. This optimization and subsequent background studies have been carried out with simulations employing idealised fields. Realistic field maps have been created and when compared the background muon flux within the detector acceptance did not show significant differences. A new optimisation is underway, as

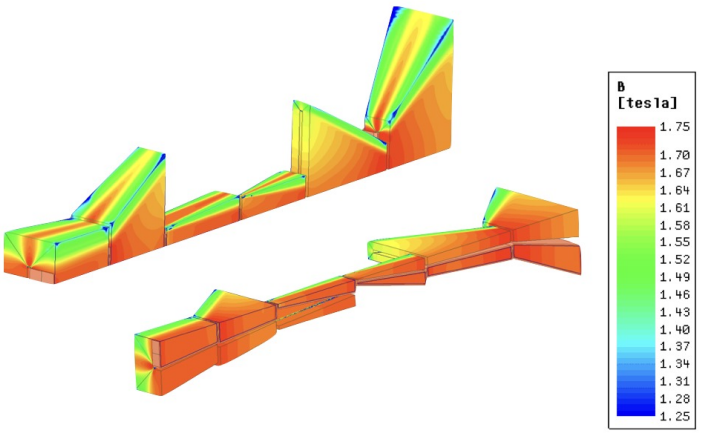


Figure 3.9: Realistic field map of the active muon shield. Figure from Ref. [18].

described in Chapter 8.4.2.

3.1.3 Scattering and neutrino detector

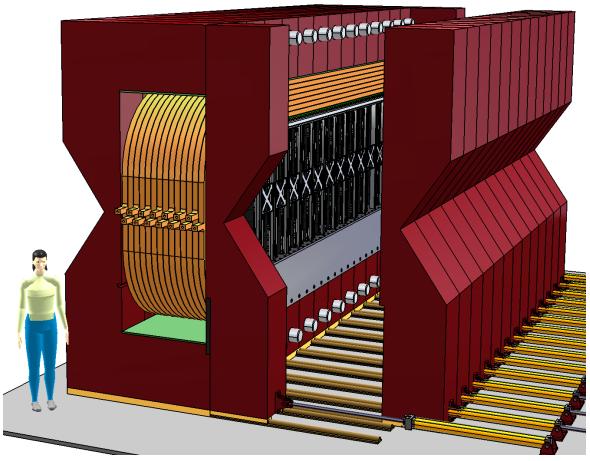


Figure 3.10: Visualisation of the SND apparatus. The magnet core and coils are coloured red and orange respectively. The layers of emulsion, lead and tracker planes sit in the centre, coloured in black. Taken from Ref. [75].

Positioned directly behind the muon shield, see Fig. 3.2, is the first of two complementary detector apparatuses at the SHiP facility, the scattering and neutrino detector. The SND is an emulsion based detector designed to perform both tau neutrino physics and light dark matter (LDM) searches. A visualisation of the sub-detector apparatus is provided in Fig. 3.10. The active muon shield, discussed in Section 3.1.2, plays an important role in protecting the emulsion

films from overexposure that would otherwise be induced by charged backgrounds. Figure 3.11 shows the area cleared of muon flux for the SND apparatus. The shape of this region and a radial decline in the neutrino flux away from the centre makes the case for a longer and thinner target.

The design consists of a ~ 7 m spectrometer magnet encasing both a ~ 3 m emulsion target with a mass of ~ 8 tonnes and some additional downstream trackers inside a large and uniformly magnetized volume. It is especially important for HS searches in the downstream DS that the SND spectrometer magnet has minimal stray field to avoid perturbing the streams of muon flux already swept out of the acceptance of the decay vessel. A full description of the design that achieves these requirements can be found in Ref. [75].

The emulsion target of the SND is comprised of alternating layers of nuclear emulsion modules and target tracker planes. The emulsion modules are constructed by interleaving AgBr crystal emulsion layers and lead layers which act as passive absorbers. This design is based on that developed by the OPERA collaboration [76]. Nuclear emulsion films have sub-micrometric position and milli-radian angular resolution allowing for accurate measurement of the momenta of charged particles. Estimates of background flux show films will require replacement every ~ 6 months. The design of the target tracker layers are yet to be finalised but the current design is closely related to that of a scintillating fibre detector (SciFi) developed for LHCb. The SND apparatus is closely followed by a muon identification system, the Upstream Background Tagger (UBT), introduced in Section 3.1.4. The UBT provides additional information about any muons created in neutrino interactions and the products of τ decays within the SND.

3.1.4 Decay volume and vetoing systems

Background suppression is of paramount importance at SHiP. After the muon shield the decay volume and its associated vetoing systems are the next components of the SHiP facility with features designed to control background rates. The decay vessel, depicted in both Fig. 3.2 and Fig. 3.12, is a ~ 50 m long pyramidal frustum shaped volume placed after the SND within the region most evacuated of background muon flux. The shape of the vessel is optimised for maximal acceptance of hidden sector particles whilst also avoiding diverted muon flux streams. The vessel

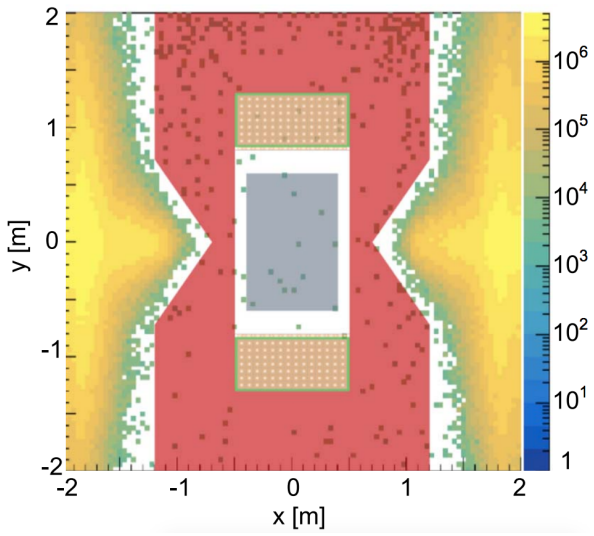


Figure 3.11: Cross sectional view of the SND in the plane transverse to the beam axis. The heatmap indicates the flux of background muons which have been swept out of detector acceptance by the upstream muon shield. The SND magnet core, coils and emulsion target are overlayed in red, orange and grey respectively. Taken from Ref. [75].

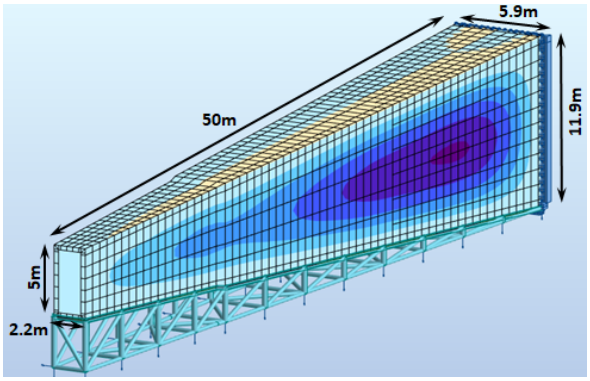


Figure 3.12: The dimensions of and the stresses existing across the structure of the vacuum decay vessel assessed via a Finite Element Method analysis. Taken from Ref. [19].

widens from its narrowest point closest to the SHiP target to its widest point downstream where the vessel incorporates the hidden sector spectrometer discussed in Section. 3.1.5.

Over 5 years the production of 5×10^{18} ν and 3×10^{18} $\bar{\nu}$ is expected within the decay volume acceptance. Neutrino interactions with the downstream material of the SHiP facility can produce long-lived neutral SM particles (V^0) such as K_L. The decay products of such particles could be reconstructed and mimic HS signal characteristics. This is a close to an irreducible background if the neutrino interaction occurs within the decay volume itself. To suppress this background

the decay volume is operated under a vacuum of 1 mbar. Furthermore, the walls and support frame of the decay vessel are designed to be as light and thin as possible to reduce the number of neutrino interactions with the vessel structure. Background events in the wall of the vessel are more easily suppressed with cuts on the location of any reconstructed vertex location.

Two vetoing sub-systems are placed around the decay vessel, the surrounding background tagger (SBT) and the upstream background tagger which covers the upstream entrance to the vessel. These are designed to detect any ionising background particles entering the decay vessel which could then go on to produce or contribute to a signal candidate. This includes any charged particles produced from muon deep inelastic scattering events in the cavern walls and muons that beat the active muon shield. The SBT is a liquid-scintillator[1] detector completely surrounding the side walls of the decay vessel. The liquid-scintillator option is chosen as it can provide large detector efficiency, coverage and sensitivity at a low cost. The UBT is composed of 12 resistive plate chambers (RPCs) each of two gas gaps, depicted in Fig. 3.13. Each RPC layer is separated by layers of 10 cm thick iron passive layers.

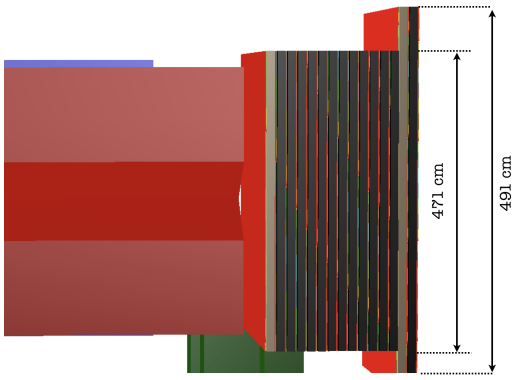


Figure 3.13: Depiction of the UBT RPC layers positioned downstream of the SND (left side of figure). The final two larger layers cover the full entrance window into the decay vessel. Taken from Ref. [18].

[1]The design is not finalized and a plastic scintillator option is also being studied.

3.1.5 Hidden sector detectors

The spectrometer strawtube tracker (SST) consists of four tracking stations, two either side of a large warm magnet, see Fig. 3.14. The SST is designed to accurately measure the momenta of charged particles passing through. From pairs of tracks the SST will reconstruct the momentum of the signal candidate particles, their decay vertices and the impact parameter of reconstructed signal candidates with respect to the SHiP target. These measurements provide enough information to achieve significant background discrimination, see Chapter 5. To accurately accomplish this task the SST is incorporated into the vacuum decay vessel and the tracking stations are designed to be as light as possible to minimise any multiple scattering of passing particles.

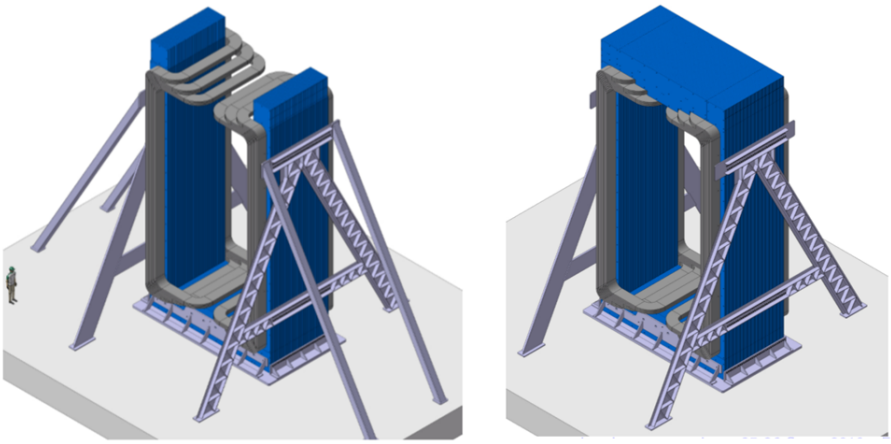


Figure 3.14: CAD model of the HS SST magnet and its supporting structure. Taken from Ref. [19].

A dedicated timing detector (TD) is installed immediately downstream of the aluminium end cap of the vacuum vessel. This cap itself is again designed to be a thin as possible to minimise any impact on downstream calorimetric performance. The TD will provide a resolution of better than 100 ps. This high resolution is a requirement to reduce combinatorial di-muon backgrounds, again see Chapter 5. Tracks made in the SST are extrapolated downstream to the TD where matches can be made. Two technologies for the TD are being studied in parallel, a plastic scintillator option and a time measuring resistive plate chambers option (MRPC).

The electromagnetic calorimeter employed is the SplitCal, a $25X_0$ long lead longitudinally segmented sampling calorimeter, depicted in Fig. 3.15. Along with 40 sampling layers there are

3 wider high precision sampling layers, these are positioned at optimal depths to be able to reconstruct the angle of any incoming photon to a few mrad of precision. This high precision is required to reconstruct $ALP \to \gamma\gamma$ decays. The segmentation within the SplitCal provides sufficient electron/hadron discrimination to remove the dedicated hadronic calorimeter present in the original SHiP design.

The muon system is the final sub-detector of the SHiP facility, it is comprised of four active stations interleaved with iron filters. The baseline technology consists of scintillating tiles with direct silicon photomultiplier readout providing good timing resolution of < 300 ps per station.

The combined effect of the hidden sector detectors is to suppress backgrounds using pointing and timing information and to provide particle identification (PID) information to aid discrimination between a wide variety of HS signals.

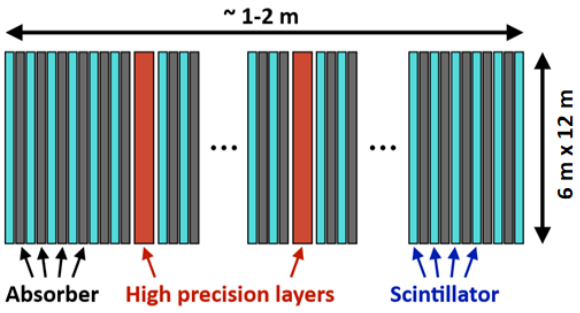


Figure 3.15: Cross-sectional view of the SplitCal with the associated dimensions. Taken from Ref. [19].

3.2 SHiP simulation suite

The simulation of the various physics processes and the response of the SHiP detector are handled with the `FairShip` simulation suite, which is based on the `FairRoot` software framework [77]. Within `FairShip`, primary collisions of protons are generated with `Pythia8` and the subsequent propagation and interactions of particles is simulated with `GEANT4`. Neutrino interactions are simulated with `GENIE` [78], while heavy flavour production and inelastic muon interactions with `Pythia6` [79] and `GEANT4` respectively. Secondary heavy flavour production in cascade interactions of hadrons originated by the initial proton collision with the SHiP target has been implemented [70]. The pattern recognition algorithms used to reconstruct tracks from the hits in

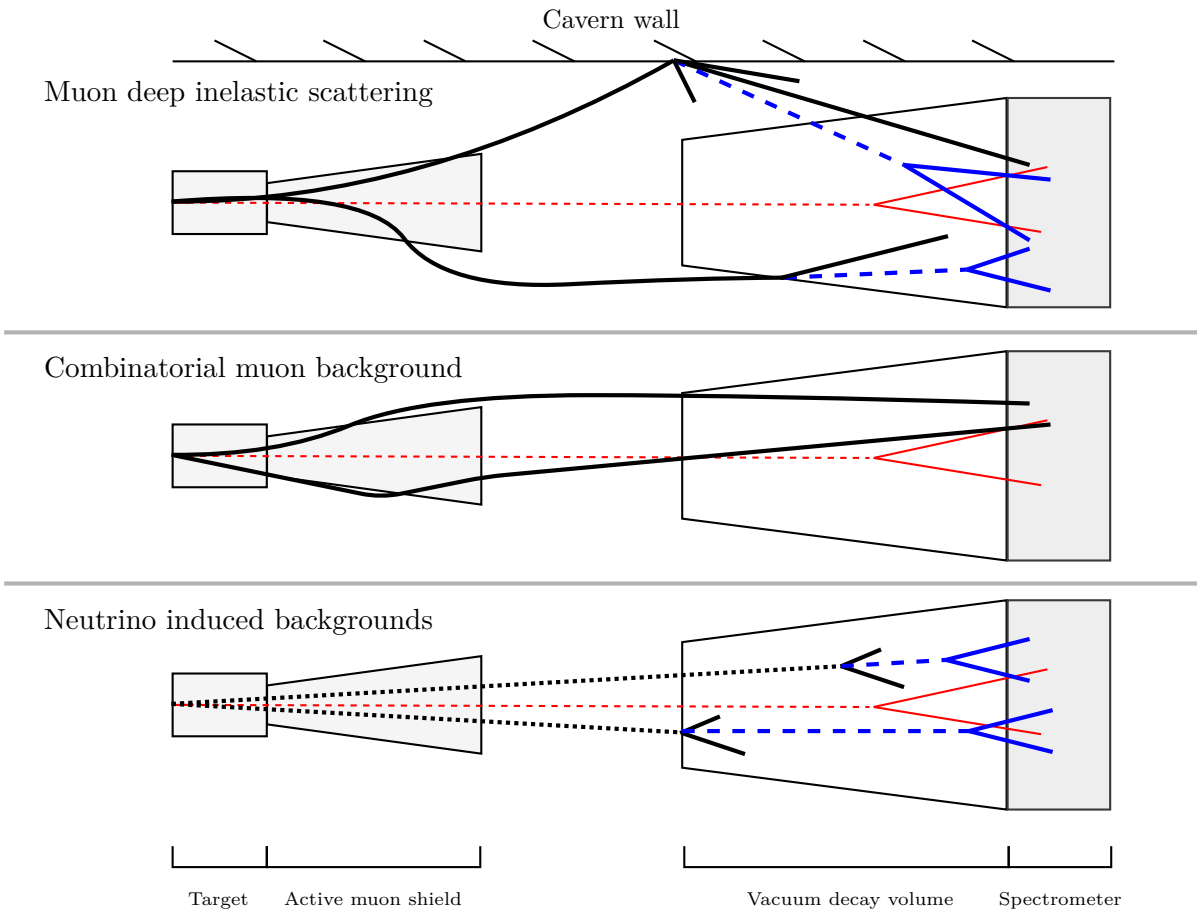


Figure 3.16: Schematics of various backgrounds to HS searches at SHiP. An example signal vertex is provided in red.

the strawtubes of the DS are described in Ref. [80], and the algorithms for particle identification are presented in Ref. [81].

The coordinate system in the `FairShip` has the beam axis at ($x = 0$, $y = 0$), with more positive values of the z coordinate as downstream. The point $z = 0$ is defined at the centre of the decay vessel. The most upstream point, the start of the target, therefore has a negative z- coordinate, this value will change based on the version of the `FairShip` being used. Any variation will predominately be caused by the choice of muon shield configuration. Values for positional coordinates are handled in units of centimetre.

3.3 Backgrounds in the HS detector

Depicted in Fig. 3.16 the following backgrounds are the main considerations for NP searches using the hidden sector detector, it has been shown that these sources can be suppressed to < 0.1 events over the lifetime of the experiment [19]:

- the random temporal coincidence of two background particles reaching the HS detectors, the main component of which is the muon combinatorial background. The estimated rates and suppression of this background are investigated in Section 5.

- the SHiP facility is only $\mathcal{O}(10)$ m underground so a significant fraction of cosmic muon flux will reach the detector. Studies show the majority of muons contributing have momenta > 100 GeV/c which means any signal candidate can easily be cut with pointing and vertex location requirements [16]. More relevant is the deep inelastic scattering of cosmic muons.

- background muons from the initial proton collisions in the target or from cosmic sources can induce deep inelastic scattering (DIS) events in downstream material. Particles such a K_L can be produced in DIS events, these are long-lived enough to enter the decay vessel a mimic the topology of HS decays. These events can be split into two classes. Firstly, muons can scatter in the cavern wall. For this component a large simulated sample exists as the muon shield sweeps the vast majority of muons into the wall. Secondly, muons can scatter in downstream elements of the SHiP detector itself, the majority of which scatter on the walls of the decay volume. These muons are less well understood, only a small subset ($\mathcal{O}(10^4)$) of the collaboration muon background sample survive the shield and make hits in the walls of the decay vessel. Vetoing events with hits in the SBT, requiring pointing back to the target and requiring that a reconstructed vertex be fiducial is enough to suppress these backgrounds [18].

- neutrino interactions in the material of the decay vessel walls and cavern walls results in backgrounds that have a similar topology to μDIS events, and are therefore reduced in a similar manner. Neutrinos can also enter the decay volume undetected and interact with the gas that fills the vessel. As mentioned in Sec. 3.1.4, the decay vessel is operated

under vacuum to reduce the number of events occurring within the vessel itself. It has been demonstrated that it is possible to suppress this background such that it is negligible. This was shown for more than the total run time of SHiP [18].

Chapter 4

Fast simulation of muon backgrounds

Accurate event generation is a crucial component of modern particle physics experiments. Large samples of simulated physics processes, including event generation and detector response, are required in order to optimise the design of new facilities, develop reconstruction algorithms, understand detector efficiencies and model the impacts of various physics based selections. Experiments at the LHC simulate billions of events every year, each event taking up to $\mathcal{O}$(min) to simulate [82, 83]. This results in simulation campaigns consuming up to 70% of experiment computing resources [84, 85]. Even then, the limited size of simulated samples are often a source of major uncertainty in analyses [86].

Newly proposed facilities and experiment upgrades will continue to demand a rapid increase in the number and the quality of simulated events [16, 87, 88]. These challenges have catalysed efforts to develop faster simulation and event generation technologies. The ongoing optimisation and parallelisation of traditional event generation software will at best result in an order of magnitude performance enhancement [89, 90]. This improvement is not sufficient to meet the ever increasing simulation demand. Estimates forecast a 4-fold shortfall of simulation capacity within the next 10 years without significant new investment [91, 92]. This is before taking into account the growing computational complexity of future generator software.

This section details the development of a fast sampling method to produce muon kinematics. The goal is to emulate muon production in interactions of the 400 GeV SPS proton beam with the target of the SHiP experiment. To achieve its physics goals the SHiP detector needs to operate under ultra-low background conditions, ensuring the facility is capable of creating such an environment requires large samples of muon induced background processes. Through the use of Generative Adversarial Networks (GANs) it is possible to imitate the otherwise computationally intensive process of producing large muon background samples. For the simulation requirements of the SHiP experiment, the GANs developed in this chapter offer a speed increase of a factor of $\mathcal{O}(10^6)$.

The methods discussed in this chapter can be generalised and applied to modelling any non-discrete multi-dimensional distribution.

The majority of the contents of this chapter is an extended version of Ref. [17], the content of which is my work. This is enhanced with details from my work within Ref. [93].

4.1 Existing background sample and motivation

In order to optimise the design of the active muon shield at SHiP, develop the reconstruction and selection algorithms of the SHiP experiment, and verify the facilities background reduction performance, a large representative sample of background muons is required. Of all the constituent parts of the SHiP simulation, the simulation of proton interactions with the target and the computation of the resulting particle showers is by far the most computationally intensive step. It therefore made sense for the collaboration to carry out a large target simulation campaign, from which a sample of muon kinematics leaving the target could be recorded and reused for each background study.

Using the SHiP simulation suite, `FairShip`, introduced in Section 3.2, 400 GeV proton on target collisions were simulated by the collaboration. The initial interactions were simulated with `Pythia8`, and the subsequent propagation of particles through the dense target was simulated with `GEANT4`. For this background production campaign any particle produced with a momentum $p < 10$ GeV/c was dropped, this increased the efficiency of the computation with minimal impact

on the results. The value of this momentum cut was guided by an assumption that a muon with $p < 10$ GeV/c has a close to zero probability of leaving the hadron absorber. In addition, if it did, it would not be likely to induce background. Cascade muon production from decays of heavy hadrons in the target was simulated by `Pythia6` [70]. Cascade events are mixed into the full sample with weights to account for their physical production fractions. In order to enhance this sample with muons that are more likely to enter the DS the cross section of muon production processes from decays of ρ^0, ω, η and η' mesons was enhanced by a factor of 100. Similarly, the cross section for photon conversions into muon pairs was also enhanced by the same factor [94]. For this background simulation campaign only the target and an un-magnetised hadron absorber were included, this geometry is laid out as a schematic in Fig. 4.1. Muons that reach the scoring plane at the end of the hadron absorber have their initial kinematics at production recorded.

The campaign produced a background sample corresponding to approximately 10^{11} protons-on-target (POT), for the enhanced channels the sample represents significantly more POT, $\sim 10\%$ of a spill. The simulation of this sample required months of computation on dedicated CPU farms, however corresponds to just a fraction of the 4×10^{13} protons-on-target of a single SPS spill. Muon production is the overwhelming bottleneck in the simulation of muon background processes at SHiP.

Using this background sample the studies of two background processes at SHiP remain statistically limited. These are the combinatorial background, where the estimates of rates require significant factorisation assumptions, and the μDIS. Specifically it is μDIS events occurring in the material that makes up the vacuum vessel and other elements of SHiP downstream of the muon shield which are statistically limited. The component of the μDIS background originating from scattering events in the concrete walls surrounding the facility is much better understood, as the muon shield sweeps muons into this region and therefore the kinematical distributions of muons involved in these events are much more populated. The high efficiency of the muon shield almost paradoxically makes studies of the impact of background processes involving muons that pass the shield less certain. The statistical power of the background samples will be reduced further still if the second round of muon shield optimisation can improve on the current performance of the shield.

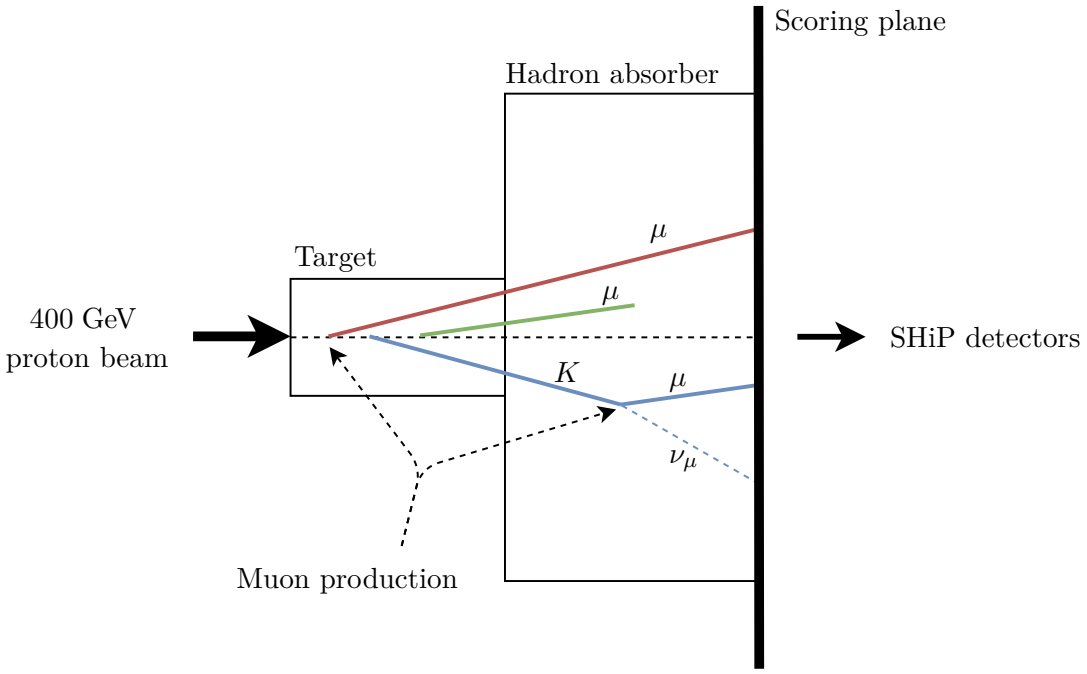


Figure 4.1: Schematic of the geometry employed in the fixed target muon background simulation campaign. This schematic is just illustrative and is not to scale.

Generating muon background samples corresponding to the 2×10^{20} SPS protons on target expected during the lifetime of the experiment is practically impossible using conventional simulation methods. However, larger samples of background muons are needed. Generative adversarial networks can offer an alternative approach, capable of producing orders of magnitude larger samples with minimal expense to the fidelity of the generated muons.

4.2 Generative adversarial networks

GANs are a class of machine learning (ML) algorithm designed to sample from intractable multidimensional functions. Well trained GANs map a given randomly distributed input latent vector to a high quality sample that cannot be distinguished from a training sample. In the ML community GANs have been shown to work well across a spectrum of tasks. The most common task being the generation of fake images [95–101]. Each image in a training set is made up of a multitude of pixels, and so each image essentially corresponds to a single data point in a high

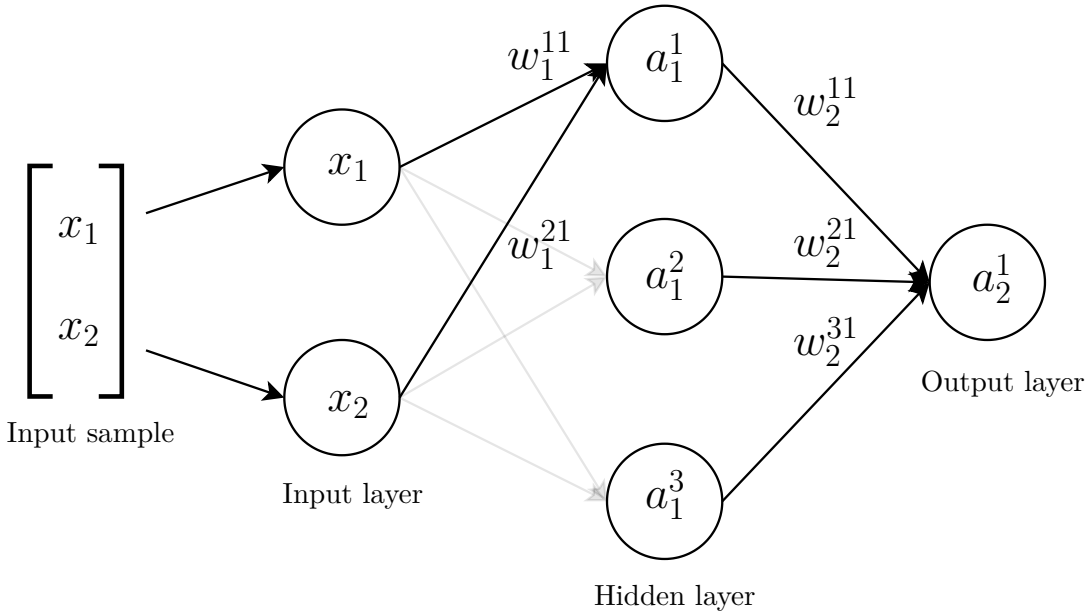


Figure 4.2: Schematic of a simple network.

dimensional space. Within this space, the underlying features of the training set of images are encoded through dependencies between pixels. GANs attempt to model the characteristics that define a set of training images. Trained models can then be used to generate images that are faithful emulations of the examples in the original training set.

Recent developments in the ML community, catalysed by hardware improvements, have improved generative neural networks to the point that they can feature as viable tools within particle physics computation. Applications of GANs within particle physics are constantly appearing. GANs have been applied in both event generation [17, 82, 102–107] and detector modelling [85, 108–117].

Neural networks are *trainable* computational objects that approximate functions, mapping an n-dimensional input space into an m-dimensional output. A basic network is built up of multiple *layers*: an input layer, some intermediate hidden layers, and an output layer, see Fig. 4.2. Layers are comprised of many individual *nodes*, a pattern of connections then joins nodes in adjacent layers. Each node has an associated *bias* term, b, acting as an activation threshold of the node, and each connection has an associated *weight*, w, representing the strength of the connection. The weights and biases of the network are tunable parameters.

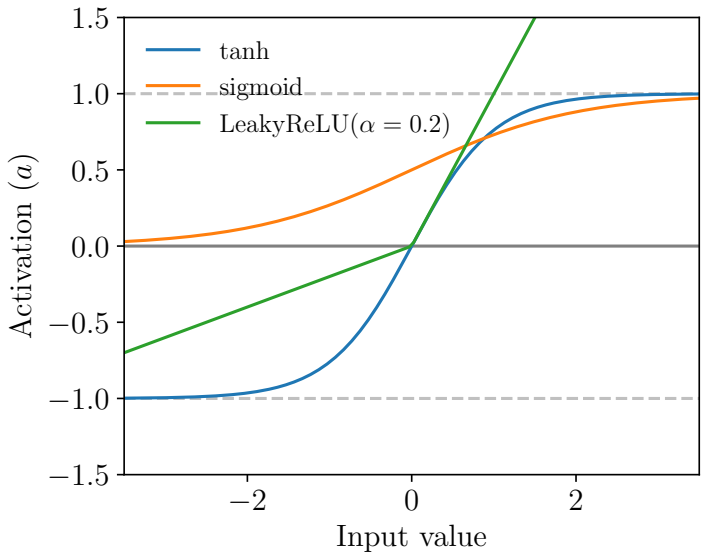


Figure 4.3: Activation functions employed in this chapter.

The simplest pattern of connections between layers is one where the nodes in each layer are *fully connected* with nodes in the adjacent layers. Samples are propagated forwards through the network and in the fully connected configuration the output value of each node is calculated as a weighted sum of the outputs from the last layer. This value is then shifted by the bias term and the result is passed through an activation function. The activation function, $\mathcal{A}$, modulates the output activation, a, of a node and provides some non-linearity. Alternatively, in equation form a is computed as

$$a_L^i = \mathcal{A}\left(\sum_j w_L^{ji} a_{L-1}^i + b^i\right), \tag{4.1}$$

where the index L refers to the layer number, i is the node index and j is the node index of the previous layer. Depending on the layer that a particular node belongs to, different types of activation functions are used. For instance, hidden layers often make use of the so called "leaky rectified linear unit" function [118]. Output layers often use a sigmoid function, this constrains the output to values between 0 and 1. This choice may be appropriate for a binary classification network, the output of which would be an assessment of whether the input sample originated from one of two sample classes. The activation functions used in this chapter are displayed in Fig. 4.3.

A network must be trained in order for it to accurately carry out a task. This training process

involves the tuning of the weight and bias parameters, with *supervised learning* being the most intuitive approach to training. In the example of a binary classification problem, labelled data are first passed through the network. Output values are then compared to the true labels via a *loss function*. The loss provides a quantitative measure of the performance of the network on a set of input training samples. A large value for the loss indicates that the network is currently struggling to distinguish between the two classes of samples. The loss is used in the process of *back-propagation* to systematically step backwards through the network updating the weight and bias parameters in an effort to improve network performance [119]. The extent to which the loss affects network parameters is controlled by the *learning rate*. Networks are trained in steps, in each of which a small *batch* of training data is fed through the network. The loss function is then evaluated and the trainable parameters are updated using only this batch of data.

GANs employ two competing networks, one acting as a generator and the other as a discriminator [120]. The generator $\hat{G}$ is trained to map an input vector of random noise z to an output generated vector $\hat{G}(z;\theta_g)$, where θ_g are the parameters of the network. The discriminator, $\hat{D}$, with trainable parameters θ_d, is trained to map an input vector x to an output prediction $\hat{D}(x;\theta_d)$, which is constrained to a value between 0 and 1. This value represents the probability that x originated from the training sample. A value of $\hat{D}(x;\theta_d)$ closer to 0 indicates that $\hat{D}$ expects the sample to have been generated by $\hat{G}$, whereas if $\hat{D}(x;\theta_d)$ is close to 1 then $\hat{D}$ is predicting that the sample originated from the training data. For this application $\hat{G}(z;\theta_g)$ and x represent vectors of muon kinematics.

The discriminator and generator networks are trained in an iterative approach. Firstly, the discriminator is trained to distinguish between generated and training samples via a binary crossentropy loss function L_d. This is a common loss function for training classifier networks and is defined as

$$L_d = \frac{1}{n}\sum_{i=1}^{n} -[y^i_{\text{true}}\log(y^i_{\text{pred}}) + (1-y^i_{\text{true}})\log(1-y^i_{\text{pred}})], \tag{4.2}$$

where y_{true} takes the values of 1 or 0 for the training or generated label of the sample respectively, and y_{pred} is the predicted label by the discriminator $y_{\text{pred}} = \hat{D}(x;\theta_d)$. The value of this loss function increases rapidly the further y_{pred} is from the y_{true}, see Fig. 4.4. Large values of the

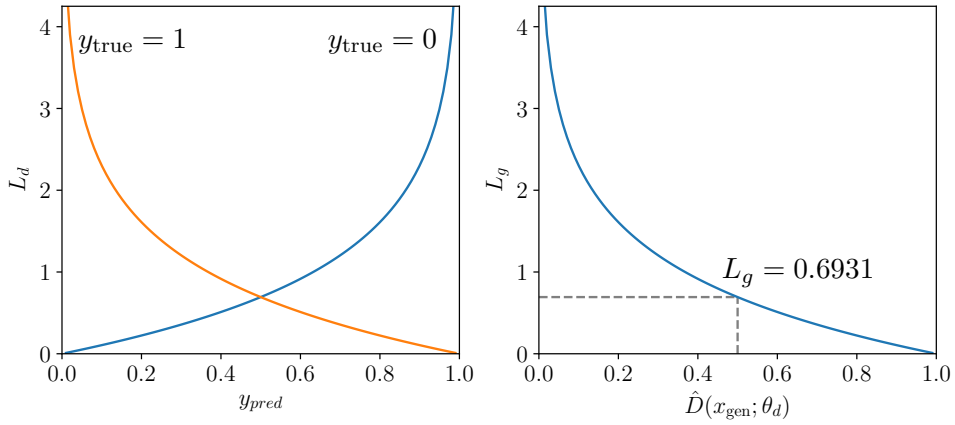


Figure 4.4: The L_d and L_g loss functions.

loss function bring significant changes to θ_d.

The generator network is then trained in a stacked model which directly connects the output of $\hat{G}$, x_{gen}, to the discriminator, leading to a vector of predicted class labels $\hat{D}(x_{\text{gen}}; \theta_d)$. This is the adversarial component of the GAN, it is only the feedback of $\hat{D}$ that influences the training of $\hat{G}$. In this stacked model all training parameters of the discriminator, θ_d, are frozen. The trainable parameters, θ_g, of the generator are updated based on the loss function, L_g, defined as

$$L_g = \frac{1}{n} \sum_{i=1}^{n} -\log(\hat{D}(x^i_{\text{gen}}; \theta_d)). \tag{4.3}$$

Low values of $\hat{D}(x_{\text{gen}}; \theta_d)$ indicate that the discriminator is confident that the sample x_{gen} originated from the generator, leading to a large value of L_g. Generated samples that closely resemble training samples will return higher values of $\hat{D}(x_{\text{gen}}; \theta_d)$ and consequently lower values of L_g as the discriminator was successfully tricked. Again, see Fig. 4.4 for a visualisation of the form of this loss.

The training of the GAN is completed when generated samples $\hat{G}(z; \theta_g)$ are either indistinguishable from training samples or their quality is no longer improving. This point is hard to define and there exists no universal definition of when to stop training a GAN. Section 4.3.2 introduces a metric to monitor the quality of the generator throughout the training process tailored for this application, this gives an indication of when to stop.

4.3 GANs for the SHiP experiment

The training sample discussed in Sec. 4.1 contains the following initial muon kinematics: the position $\mathbf{r}$, the momentum $\mathbf{p}$ and the charge of 3.4×10^8 muons that passed through the target and un-magnetised hadron absorber and reached the scoring plane. This sample is split into four sub-samples. From each of these sub-samples a separate GAN is trained, separated by muon charge and prompt or displaced origin. As mentioned, this training sample is artificially enriched with muons from rare processes. Therefore, in order to obtain a GAN that can produce a physical admixture of muons from the various sources, batches of muons are extracted from the training sample according to a probability that cancels out this enhancement.

The GAN will generate position and momentum vectors of muons within the target. After generation, muons can then be propagated through the active muon shield and the Decay Spectrometer, relying on `GEANT4` to simulate muon interactions with matter. This approach allows for a fast production of large muon samples, while maintaining the flexibility to optimise the muon shield and downstream detector elements of SHiP. Additionally, this approach leaves in the ability to change the spatial smearing of the proton beam impinging on the target.

The x- and y-coordinates of muons originating from prompt decays of mesons produced in the initial proton collision such as the ρ^0, ϕ and J/ψ are always the same. This is just a consequence of using a training sample that relies on `Pythia` with no smearing of the proton-beam distribution. As mentioned, muons from prompt sources are treated separately from muons originating from other sources. Therefore, the GANs trained on prompt muons generate just four features (z, $\mathbf{p}$), and the GANs for non-prompt muons generate six ($\mathbf{r}$, $\mathbf{p}$).

In this approach correlations between muons produced in pairs from, for example vector-meson decays, are ignored. Muons are generated individually. This is an unimportant effect as while the probability of a single muon to reach the detectors is low the probability that a correlated pair do is vanishingly small. It is also likely that any correlation is lost via the multiple scattering of the muons through the hadron absorber and muon shield.

4.3.1 Pre-processing

The distribution of the x- and y- coordinates of muons from non-prompt sources is extremely peaked around the interaction point. Therefore, each value of the x- (y-) distribution x^i (y^i) is transformed as

$$x^i_{\text{transformed}} = \begin{cases} -\sqrt{|x^i - \overline{x}|} & \text{if } x^i < \overline{x}, \\ \sqrt{|x^i - \overline{x}|} & \text{if } x^i > \overline{x}, \end{cases} \tag{4.4}$$

before training the GANs. This transformation widens the distributions, which proves easier for the GANs to model. The distributions of all the input features are then normalised to values between -1 and 1. These transformations will later be reversed to obtain physical values from the generated output.

4.3.2 Figure of merit

An important requirement of the full simulation of the SHiP detector is to accurately model the flux of muons reaching the DS. This flux crucially depends on the momentum distribution of the muons entering the active muon shield. Therefore, the kinematic distribution of the muons generated through the GAN approach must closely match that of the muons produced in the target using the full simulation.

In order to optimise the architecture of the networks and to quantify the quality of the training procedure a figure of merit, FoM, is developed with the following requirements. The FoM must account for how well the GAN is able to model individual features and the correlations between them. Furthermore, it is important that the FoM offers an independent metric for the quality of the training of the GAN. Using a metric derived from the loss function of the discriminator or the generator will not help as both networks improve in tandem during the iterative training procedure. Finally, the computation of the FoM must be fast enough such that it does not significantly impact the training speed.

During the training process, at regular intervals test samples are generated to query the FoM for a measure of the training progress. The muon features span a six dimensional space so a small generated sample of muons results in a very sparsely populated feature space. Therefore,

traditional binned goodness of fit methods, such as χ^2-tests, break down as almost all bins in the space have a low occupancy. Boosted decision trees (BDT) can overcome this issue [121] and satisfy the aforementioned requirements of the FoM.

A gradient boosted decision tree is fit to distinguish between generated and fully simulated muon samples. The BDT uses 100,000 muons generated from the latest GAN configuration and 100,000 randomly selected fully simulated muons. Half of the muons in each sample are used for fitting, the other half for querying the BDT. The resulting performance of the BDT is quantified by computing the area under the receiver operating characteristic curve (ROC AUC). A generated sample that is indistinguishable from a fully simulated sample would return an ROC AUC value of ~ 0.5, and a value of ~ 1 for a fully distinguishable sample. The sample size of 100,000 is confirmed as large enough by fitting multiple BDTs with bootstrapped training data. Accounting for the fact that when fitting on two samples from the same source the resulting ROC AUC value could fall either side of 0.5, a mean FoM value of 0.5011 ± 0.0003 is observed. A sample size of 100,000 is therefore satisfactory unless the GAN reaches scores close to this.

4.3.3 Network optimisation and GAN architecture

All networks are trained with a mini-batch gradient descent approach [122], where at each training step the networks use a small sub-sample of training data. A batch size of 50 is chosen. This choice attempts to keep the benefits of both small and large batches. Large batch sizes result in more accurate gradients and are more computationally efficient to compute. However, smaller batches provide a stochastic component, this can enable the gradient descent procedure to kick itself out of any local minima.

The architecture of the generators and discriminators of the GANs are kept simple, they are built using only fully connected layers. The fully connected architecture is agnostic to any prior belief that some correlations between parameters will be more or less important. This keeps the result independent of the original ordering of the muon parameters in the vectors of features that make up the training sample. The number of nodes, the batch size, the number of layers and the learning rate of the networks are coarsely optimised through a manual grid search over these hyperparameters. The four GANs are trained until the ROC AUC of the BDT based FoM

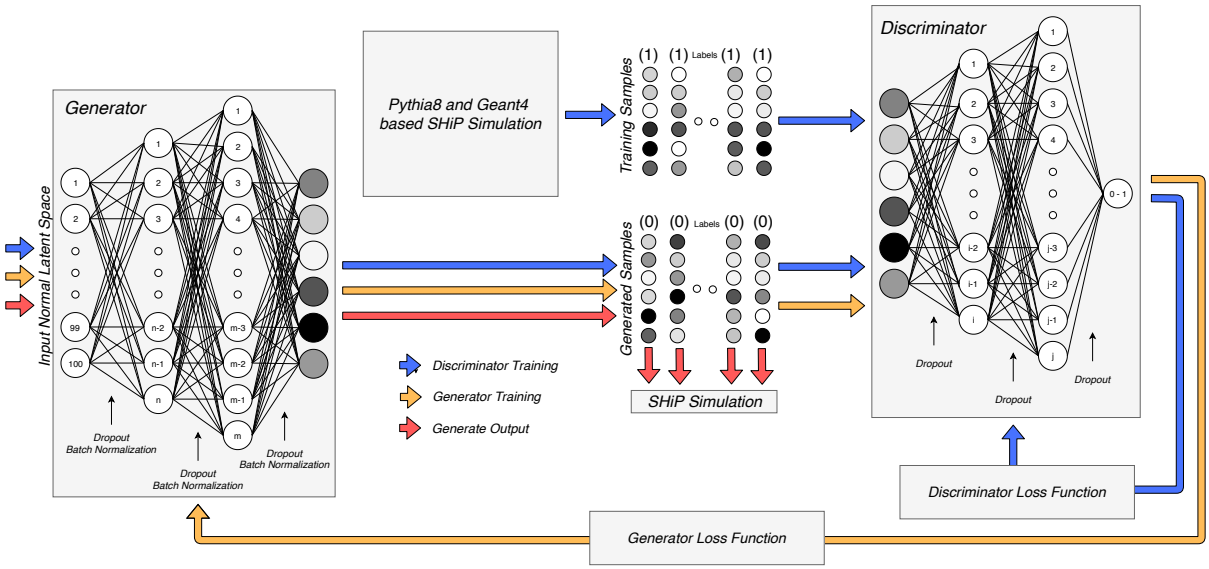


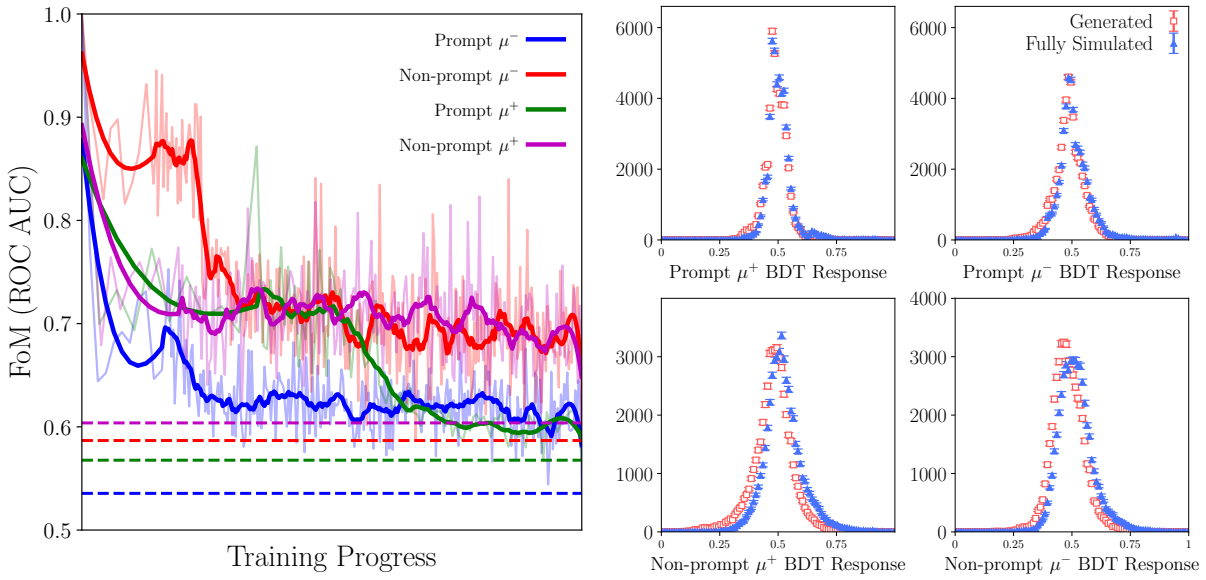
Figure 4.5: Schematic of the GAN architecture employed. Arrows indicated the flow of samples and loss information during each stage of training.

described in Sec. 4.3.2 flattens out, and the architecture selected is that which minimised the FoM.

As a result of this optimisation procedure, the GANs for both prompt and non-prompt muons follow the architecture depicted in Fig. 4.5. Leaky rectified linear unit activation functions are used at every hidden layer. The $\hat{G}$ and $\hat{D}$ networks each have two hidden layers in an inverted pyramidal structure. For the prompt muon GANs, the number of nodes in each hidden layer of $\hat{G}$ are 512 and 1024 and for $\hat{D}$ are 768 and 1536. For the non-prompt GANs, the number of corresponding nodes are 1536 and 3072 for $\hat{G}$ and 768 and 1536 for $\hat{D}$. The latent input to the generators sample from an 100 dimensional unit Gaussian distribution. The last layer of $\hat{G}$ has a tanh activation function in accordance with the transformed range of the input features described in Sec. 4.3.1. The last layer of $\hat{D}$ has a sigmoid activation function providing an output between 0 and 1 that represents the judgement of $\hat{D}$ on the origin of each input sample. Dropout layers with a dropout probability of 0.25 are added between each layer of $\hat{G}$ and $\hat{D}$ to help prevent overfitting [123], and batch normalisation layers are also added between layers of $\hat{G}$ [124]. This network structure is summarised in Table 4.1.

$\hat{G}_{prompt}$	$\hat{D}_{prompt}$	$\hat{G}_{non-prompt}$	$\hat{D}_{non-prompt}$
(100,) (Gaussian noise)	**(6,)**	**(100,)** (Gaussian noise)	**(6,)**
(512,) (Fully connected)	(768,) (Fully connected)	(1536,) (Fully connected)	(768,) (Fully connected)
Dropout(0.25)	Dropout(0.25)	Dropout(0.25)	Dropout(0.25)
Leaky ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU
Batch normalisation	(1536,) (Fully connected)	Batch normalisation	(1536,) (Fully connected)
(1024,) (Fully connected)	Dropout(0.25)	(3072,) (Fully connected)	Dropout(0.25)
Dropout(0.25)	Leaky ReLU	Dropout(0.25)	Leaky ReLU
Leaky ReLU	**(1,)** (Fully connected)	Leaky ReLU	**(1,)** (Fully connected)
Batch normalisation		Batch normalisation	
(6,) (Fully connected)		**(6,)** (Fully connected)	

Table 4.1: Summary of network architecture.



(a) (b)

Figure 4.6: (a) Progress of the FoM ROC AUC values throughout the training, raw and smoothed data is displayed. Dashed lines indicate the FoM values of final models, (b) Distributions of the BDT response for both fully simulated and GAN-based muon samples.

For this study the `Adam` optimisation algorithm [125] was used during training. Employing the `AMSgrad` algorithm with the `Adam` optimiser improved the stability of L_g and the FoM progress with training [126]. A momentum parameter of `Adam`, β_l, is used with a value of 0.5 to help control the progress of the gradient descent during the training of the network.

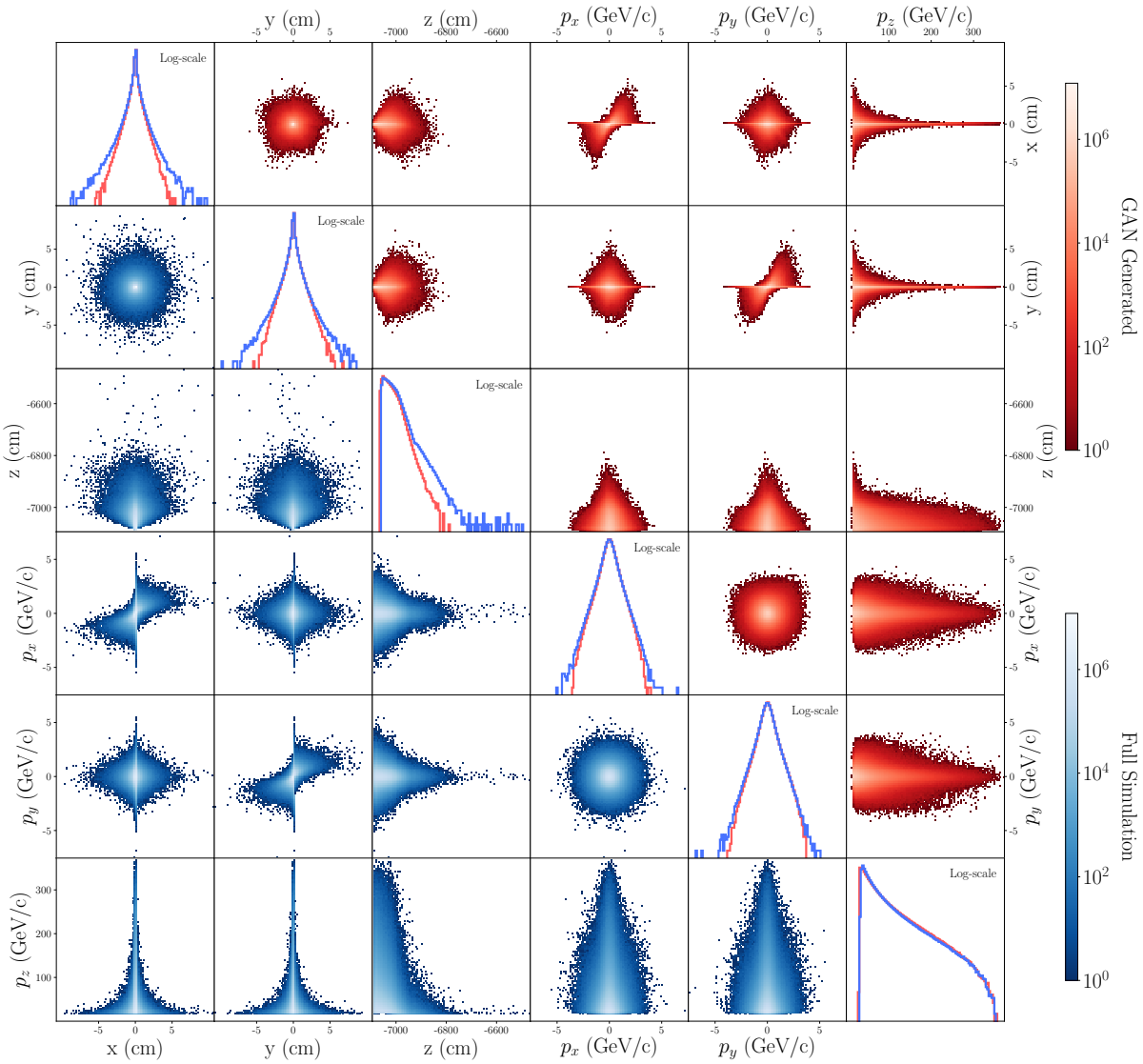


Figure 4.7: Two-dimensional distributions of all unique combinations of muon features for GAN based (upper-half) and fully simulated (lower-half) muons produced in the SHiP target. One-dimensional log scale comparisons of each feature are presented along the diagonal.

4.4 GAN performance

The progress of the FoM throughout the training of each GAN, as well as the distributions of BDT output of the optimal GAN models are shown in Fig. 4.6. The final FoM values for the prompt μ^+ and μ^- GAN models are 0.57 and 0.54 respectively. Whereas, the non-prompt μ^+ and μ^- GAN models return FoM values of 0.60 and 0.59 respectively. These results were obtained by taking the best results from multiple training runs for each GAN.

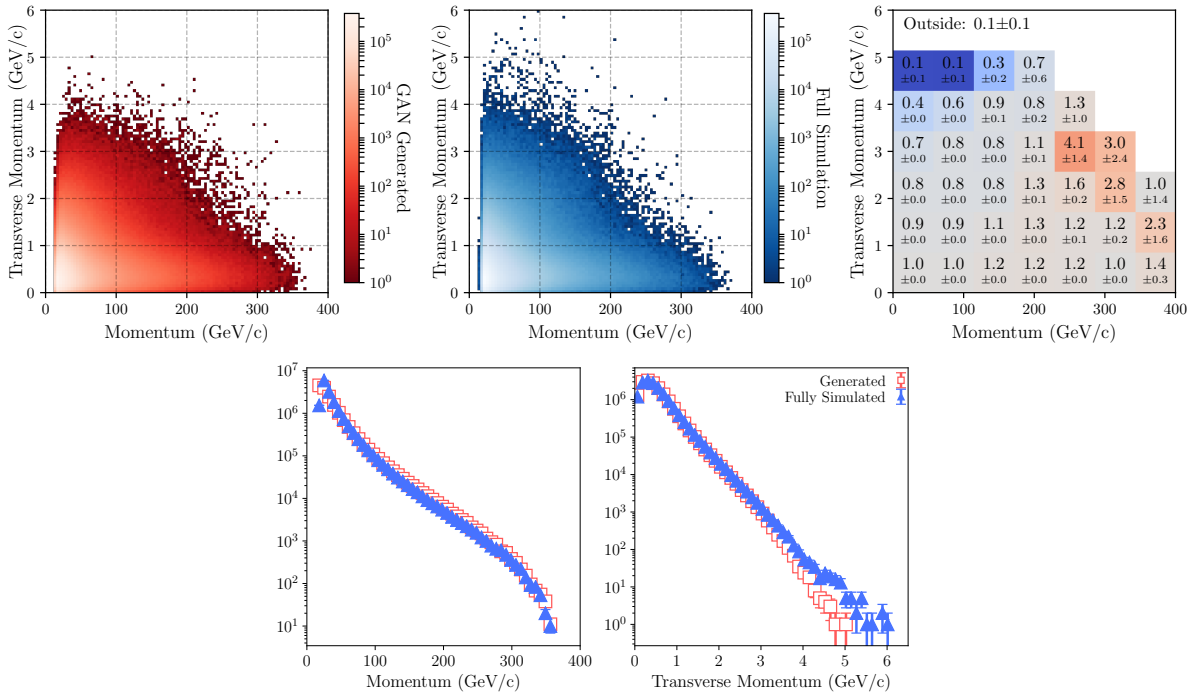


Figure 4.8: Two-dimensional p vs p_{T} distributions for GAN based (top-left), fully simulated (top-middle) and the ratio (top-right) of muons produced in the SHiP target. The comparisons of the one dimensional projections for p (bottom-left) and p_{T} (bottom-right) are also shown.

In order to visualise the level of agreement between the generated and fully simulated samples, a physical sample of GAN based muons is produced by combining the output from each of the four generators. This combination is made according to the expected production fractions of prompt and non-prompt muons in the simulation. Figure 4.7 compares the one- and two-dimensional distributions of each unique pair of features between fully simulated and generated muons.

Overall, it appears the GANs can reproduce the correct correlations between features, although the tails of the (x, y, z) position distributions are clearly underestimated. However, there are examples of asymmetries and over populated regions of the tails in the GAN sample in all 2D projections shown.

Modelling of the momentum (p) and transverse momentum (p_{T}) plane accurately is crucial in order to obtain the correct flux of muons reaching the SHiP Decay Spectrometer. Figure 4.8 compares the (p, p_{T}) plane between the fully simulated and generated samples. The GANs can largely reproduce the correlations between these features, however in particular they

underestimate the number of muons with $p_{\mathrm{T}} > 3$ GeV/c. To attempt to correct for this effect on the momentum distribution of the generated muons, the three-dimensional (p_x, p_y, p_z) distributions of the fully simulated and generated muons are each fit using a three-dimensional Kernel Density Estimator, for example see Ref. [127]. For each generated muon, an individual corrective weight is derived by taking the ratio between fully simulated over the fully generated KDEs at the corresponding (p_x, p_y, p_z) muon coordinate. Note that calculating weights like this has minimal effect on generation speed. Once it is obtained, querying the KDE needs only be done for the very small fraction of muons that reach the SHiP detectors.

4.5 Reconstructing GAN generated muons

Generated muon samples are introduced into the simulation at the required position within the SHiP target and with the specified momenta, as shown in the schematic of Fig. 4.1. As described in Chapter 3.1.1 the SPS proton beam will be moved over the target in a circular motion in order to reduce the peak heating of the target. To reproduce this effect the origin position of each muon added to the simulation is smeared in the x- and y- plane by Δx and Δy, calculated as

$$\begin{aligned} r &= r_s + r_b \times \mathcal{N}(0,1), & \phi &= \mathcal{U}(0,1) \times 2 \times \pi, \\ \Delta x &= r \times cos(\phi), & \Delta y &= r \times sin(\phi), \end{aligned} \tag{4.5}$$

where $\mathcal{N}(0,1)$ and $\mathcal{U}(0,1)$ are values generated from a normal and a uniform distribution respectively. The application of Δx and Δy accounts for a Gaussian-like blurring of the beam of scale r_b, and the circular sweeping, of radius r_s, over the target.

Following the smearing of the muon production vertex the `FairShip` software suite employs `GEANT4` to simulate passage through the magnetic shield and any interactions with the downstream SHiP detector. Digitization algorithms are run on the simulation truth level hits to emulate real detector effects, and then finally reconstruction algorithms create track candidates.

Figure 4.9 shows the momentum distribution of reconstructed muon tracks in the Decay Spectrometer of SHiP resulting from both the GAN based muon sample and the full simulation sample. Both a uniform and logged binning scheme are provided. Highly weighted events originating from the enhancement of some muon production channels in the fully simulated

sample as discussed in Section 4.1 are visible in these plots. The effect of the residual correction to the kinematics of the GAN based muon sample discussed in Sec. 4.4 is found to have a small effect.

Figure 4.10 shows the initial momentum distributions at the production point of the muons in the target, for muons that go on to be reconstructed in the DS. The GAN based and fully simulated muons display similar features in the p vs p_{T} plane. Note, the fully simulated sample exhibits extremely localised hot-spots. These are due to the event weights that account for enhancement factors of particular processes as discussed in Sec. 4.1.

The rate of muons that survive the magnetic shield and are reconstructed in the Decay Spectrometer is given in Table 4.2. Both the full rate and the rate of muons with initial momenta corresponding to the upper region of Fig. 4.10 agree when comparing the GAN based and fully simulated muon samples. The correction to the kinematic distributions of the GAN based muons discussed in Sec. 4.4 changes the rate of generated muons entering the Decay Spectrometer by just $\sim 4\%$.

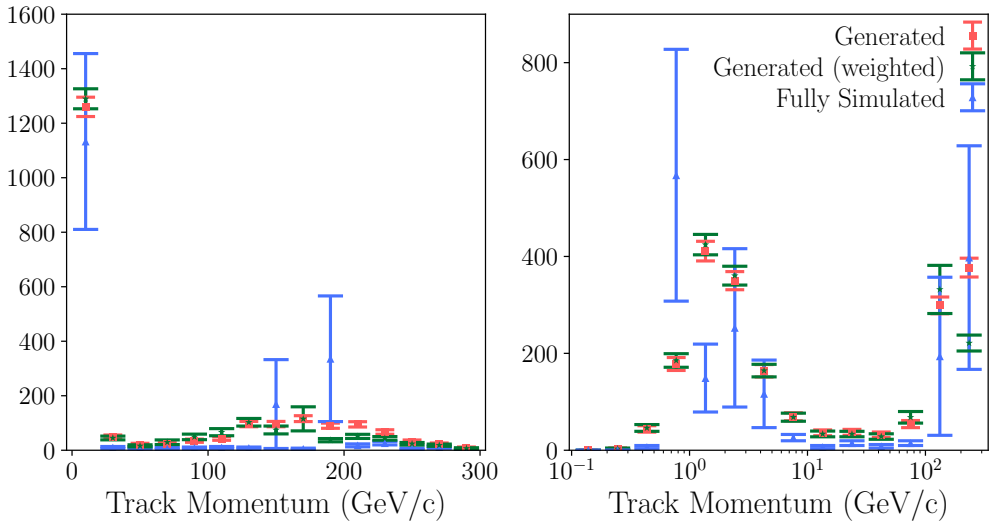


Figure 4.9: Distribution in linear (left) and log-scale (right) of the reconstructed track momentum of muons in the Decay Spectrometer. The distributions are normalised such that they correspond to the same POT.

Approach	Full Rate (kHz)	Upper Region Rate (kHz)
Full simulation	13.9 ± 3.4	4.7 ± 2.2
GAN	15.8 ± 0.3	5.5 ± 0.2
GAN (weighted)	15.2 ± 0.5	4.7 ± 0.4

Table 4.2: Rates of reconstructed muons in the Decay Spectrometer.

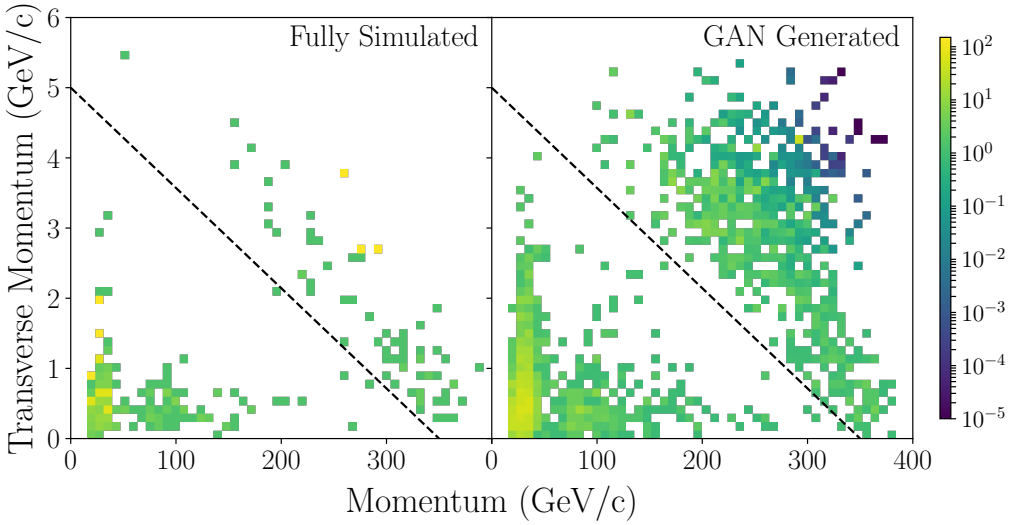


Figure 4.10: Initial momentum of muons with well reconstructed tracks in the Decay Spectrometer. Full simulation data is presented on the left and generated data on the right. The dashed line indicates the upper region referred to in Table 4.2.

4.6 Benchmarking

With a small expense in the fidelity between the generated and fully simulated sample, mostly manifesting in the tails of the distribution, this generative approach can produce samples of muons at greater speed. Generating samples of muons from GANs on GPUs provides a speed-up of $\mathcal{O}(10^6)$ relative to the full `Pythia8` and `GEANT4` proton-on-target simulation[1]. This test was performed using `Keras`(v2.1.5) [128] with a `TensorFlow` backend (v1.8.0) [129] on a single `Nvidia Tesla P100` GPU card [130]. This speed-up factor includes all the post-processing computations required to transform the output features of the generator into physical values.

Table 4.3 summarises the results of this performance test. The gain in speed using the

[1]Note, the generation code was optimised in Chapter 7 which resulted in approximately 5 times faster inference.

generative approach is partly due to the small production cross section of muons with $p >$ 10 GeV/c, requiring $\mathcal{O}(10^3)$ proton-on-target interactions to be simulated through `Pythia8` in order to generate a single muon that reaches the scoring plane. This effect is accounted for in the "Time to simulate a single muon (s)" column of Table 4.3.

Although it depends on specific hardware choices, for a rough idea, generating muons using the GAN approach on a CPU is generally an order of magnitude slower than on a GPU.

Target simulation method	Muons produced in 5 minutes	Time to simulate single muon (s)
`Pythia8` and `GEANT4`	~ 1	1.1×10^{-1}
GAN (CPU)	7.5×10^5	4.0×10^{-4}
GAN (GPU)	3.5×10^6	8.6×10^{-5}

Table 4.3: Summary of benchmarking results.

4.7 Caveats to the GAN approach

The understanding of some important backgrounds at SHiP is currently statistically limited, better understanding requires a larger muon sample. There is however no new muon background production campaign planned. Besides, even if there were it would be unlikely to be much larger than the first. Therefore some manner of data augmentation is required in order to extract as much information as possible from the current background sample.

This could partly be achieved with simple symmetry based transformations exploiting domain knowledge. For example, one could enhance the sample by rotating muon kinematics in the p_x-p_y- plane. While this would certainly work to some extent, nothing is learned about the true distribution and it would not take long to saturate the same regions of the 6D kinematical phase space. This is especially problematic as the SHiP muon shield simulation is essentially a very high dimensional filter whose behaviour may be far from smooth over the muon kinematical phase space. These transformations could potentially be combined with a noise component to smear the synthetic muons out, but this is unsatisfactory as such a process must be based on

arbitrary choices. An existing over-sampling technique could be used, SMOTE [131] or ADASYN [132] for example. However, these approaches rely on interpolating in straight lines between existing examples in the data set, this is particularly problematic in low density regions. The results have the same drawbacks as that of exploiting symmetries in the sample with rotations.

The GAN approach directly produces a model that approximates the intractable multidimensional true probability density function behind the training data, if a GAN was trained perfectly it could essentially be sampled indefinitely. Of course for the GAN to perfectly model the true distribution it must be trained on an infinite sample, and if this existed the GAN would be redundant. The approach is limited by the size of the training sample, as are the other oversampling options outlined above. Enhancing a sample with events generated from a GAN trained on a finite training sample cannot directly increase the statistical power the original sample, however it can provide additional information if some assumptions made are true.

We can imagine splitting the uncertainty of the full simulation and of the GAN model into statistical and systematic components, where the systematic uncertainty of the fully simulated sample originates from differences between the full simulation and reality, and its statistical uncertainty is limited by the computational complexity of its simulation. The systematic uncertainty of the trained GAN clearly must be as large or larger than that of the training sample, even for a perfect GAN [133]. The statistical uncertainty in the GAN sample is however no longer limited, as the GAN is orders of magnitude faster to sample from. Note though, to some extent the statistical uncertainty present in the training sample will be translated into a systematic uncertainty in the trained GAN model, this is certainly true if the GAN is overfitting to the training sample. It has been shown for various simple toy distributions that a large sample generated from a very simple GAN trained on a much smaller toy sample can better match the true distribution than the toy data set itself, and hence can provide an enhancement [134]. The additional information comes from the GAN assuming the training sample originated from a smooth and continuous distribution, where this is true it is possible for a GAN to accurately interpolate and provide an enhancement. The enhancement observed in Ref. [134] was larger in situations where higher resolution details of the distribution were important, here the interpolation from the GAN filling the gaps of the training sample is the

most beneficial. The GAN approach was also shown to be increasingly useful when enhancing distributions of higher dimensions, as the mean regional occupancy of a training sample is further reduced. In reality we have limited knowledge of the true probability density function behind the training sample, and so there is no way to check agreement other than to compare to the training sample. It is impossible to verify that any agreement holds as you sample more events from the GAN without producing more events with the slow simulation. Though, we do know that any miss modelling that does exist will be especially prevalent in the tail regions that are poorly described by the training sample. In these regions the validity of the assumption of accurate modelling of the true probability density function will diminish. Further extrapolation is possible but similarly assumes that the true probability density function continues smoothly, does not change behaviour in the extrapolated region, and furthermore that the GAN is stable enough in the region in order to correctly model this assumption. None of the alternative oversampling techniques can extrapolate without similar assumptions.

For the SHiP example, the muon shield effectively acts as a very high dimensional function. There is clearly potential for anisotropies and discontinuities to exist in the response of this function to a smooth input sample of muons. In fact this is observed even when the surviving muon properties are collapsed into 2D in Fig. 4.10. If the true muon distribution is indeed smooth and continuous, as it appears, the GAN should be able to accurately model it and provide an enhancement. This enhancement will be especially relevant if the discontinuities in the response of the muon shield are localised and the response is sensitive to higher resolution details of the input muon distribution [135]. In the SHiP set up, after GAN generation the complexity of the muon shield response to the smooth muon distribution can be then be modeled with the full `GEANT` simulation. The muon shield has a very high efficiency to successfully sweep muons out of the detector acceptance, therefore the statistical power of a single muon sample to describe what is happening per POT is reduced after the sample has passed through the shield. The GAN then also benefits from training on the more powerful sample before it passes through the shield, compared to the alternative of modelling only the muons that beat the shield. The more effective the shield is, the more valuable a reliable enhancement will be. Notably, a significant fraction of the muons reaching the detectors come from the core of the distribution, where

the assumption that the GAN is modelling the underlying true muon distribution well is more robust. In the tail regions, granted, the GAN will at least in part be miss modelling the true distribution. This is certainly true of the GANs presented in this chapter, see Fig. 4.7. However, we can still simulate the response of the muon shield and the detectors *given* the distribution of muons generated by the GAN and its potential miss modellings. For SHiP background studies the GAN effectively attempts to gets as much information as possible out of the small fully simulated sample. Generated muons can be used in parallel to complement ongoing background studies and sub-system optimisation studies, for example that of the muon shield.

Finally, something to note. The generative approach presented here could be trained directly on real data. A recent muon-flux beam-test campaign of the SHiP collaboration [136] was carried out to both validate the simulated distributions of muon kinematics and assess any requirement to tune the simulation. Employing a GAN approach directly on test-beam data would circumvent the challenge of tuning the multitude of parameters that control the simulation in order to match the data.

Chapter 5

Muon combinatorial background studies

The study of backgrounds at SHiP is of paramount importance. The goal is to develop a selection to reduce backgrounds to < 0.1 events over the full 5 years of data taking [16]. The collaboration have demonstrated this ultra-low background environment is achievable [18] and this result has been used in various sensitivity studies produced by the collaboration.

Muons are produced in both the initial collision of protons on the SHiP target and in the subsequent decays of hadrons produced in showers from material interactions. For the SHiP experiment muons have a dangerous mix of high penetrating power and a long lifetime. Muons are more capable than any other charged particle of traversing the full length of facility and creating tracks in the detectors. The hadron absorber acts as the first background defence softening the K/π momentum spectrum before such mesons decay to produce a large flux of muons, the vast majority of these muons with initial momentum > 10 GeV/c reach the end of the hadron absorber. This results in a muon flux of $\mathcal{O}(10^{11})$ Hz leaving the hadron absorber. The final section of the hadron absorber is magnetised and provides a head start on sweeping background muons out of the detector acceptance. In its current state[1] the active muon shield

[1] Shield performance may improve with a new optimisation campaign but any improvement is likely to be small.

achieves a suppression of $\mathcal{O}(10^6)$. This leaves $\mathcal{O}(10^4)$ muons entering the fiducial volume of the experiment every 1 s spill of the SPS.

The rate and characteristics of muons entering the decay volume and ultimately forming tracks in the HS spectrometer drift tubes must be understood. The occurrence of a pair of background tracks within the same timing window can mimic the signatures of signal hidden sector final states. This chapter outlines investigations into this background and the effectiveness of suppression.

5.1 Background samples

Two samples are employed in this chapter, these samples are kept separate throughout. The first corresponds to the fully simulated background sample introduced in Chapter 4.1. This sample is of a size equivalent to 6.5×10^{10} POT. Various channels were enhanced in the production, for these channels the sample is equivalent to $\sim 10\%$ of a spill. A sample of muons produced in the cascade are folded into the full sample and correctly weighted to account for the charm and beauty production cross-sections. This cascade sample is equivalent to 5×10^{10} POT.

The second sample used is generated using the GAN approach of Chapter 4, this sample corresponds to 6×10^{13} protons on target, corresponding to ~ 1 spill of the SPS. As previously described, the weights in the fully simulated sample coming from the enhancement present and the folding in of the cascade sample are accounted for in the training of the GAN, and so the GAN produces a physical sample. The weights associated with the GAN muons are derived from the corrective KDE introduced in Chapter 4.4.

5.2 Background estimation

Within the `FairShip` simulation suite the path of muons are simulated from their individual point of production within the target to the downstream detectors using `GEANT4`. Pattern recognition algorithms are used to reconstruct particle tracks in the SST from digitized hits in the simulation. These algorithms were developed by the collaboration and are described in Ref. [80]. Initial cuts on the properties of a track candidate including the track-fit quality are

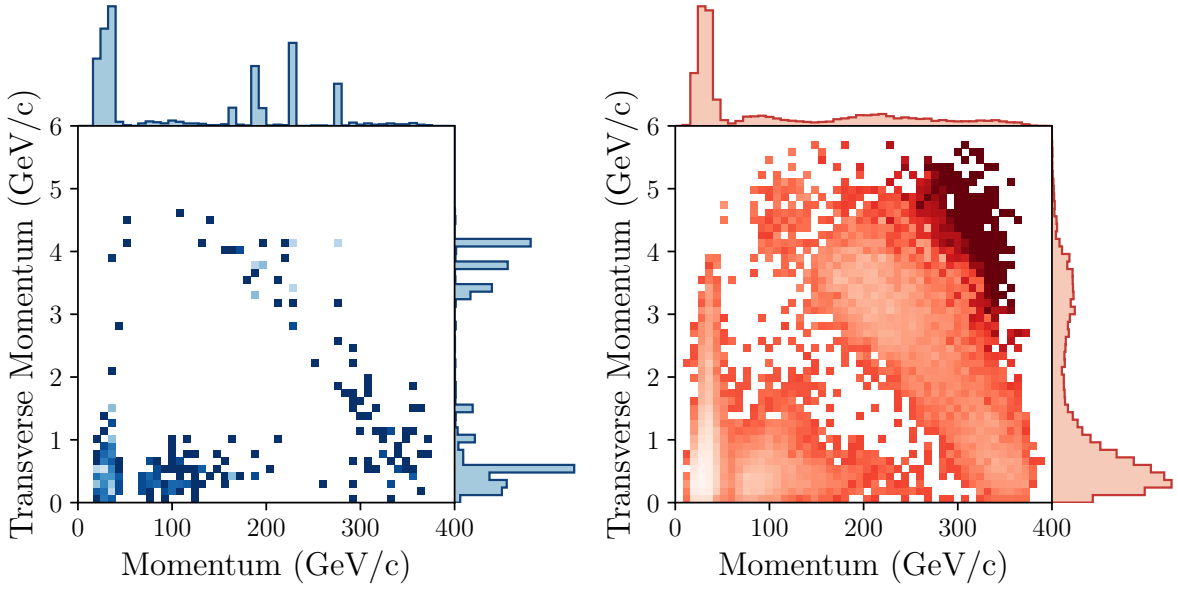


Figure 5.1: Initial momentum of well reconstructed background muons at production in the target. Fully simulated sample is in blue, the larger GAN based sample in red.

made, these are presented in Table 5.1. Muons passing these tracking cuts are labelled as *well reconstructed*.

Cut description	Value
Number of measurements	> 25
$\chi^2/ndof$	< 5
Hits before and after spectrometer magnet	True

Table 5.1: Cuts applied to track candidates.

Figure 5.1 displays the kinematics of muons at their production vertex within the target for the muons that form well reconstructed tracks in the HS spectrometer. The distribution has two distinct islands. The low momentum island contains muons that undergo large scattering in the shield. These events result in large deflections of muons into the wrong side of the magnet [18, 137]. Both the full rate and the well reconstructed rate of muons for the fully simulated and GAN samples are presented in Table 5.2. The GAN rates are higher than the fully simulated rates. The characteristics of this excess will become clearer in Section 5.2.3. This excess is left in as it only provides a tougher sample to test the SHiP background rejection methods against.

Simulation Flavour	Rate (kHz)	Well Reconstructed Rate (kHz)
Fully Simulated	18.13 ± 3.05	11.45 ± 2.42
GAN	24.96 ± 0.14	14.81 ± 0.13

Table 5.2: Muon track rate and well reconstructed track rate for each background sample.

5.2.1 Timing resolution

Each 1 s spill of the SPS will deliver 4×10^{13} 400 GeV protons to the SHiP experiment every 7.2 s. Spills will be delivered via a slow extraction, with the goal of providing an as flat as possible beam profile. This, as opposed to a single sharp peak in proton delivery, lowers the probability for the coincidence of two background muon events. The final details of the beam profile the SPS will provide are not known, so for this analysis a perfectly flat beam is assumed for simplicity. A profile close to this appears to be achievable, see Refs. [18, 69].

A high resolution timing detector (TD) placed immediately downstream of the decay vessel will provide the largest suppression of combinatorial background events. The design of the TD is not finalised however viable plastic scintillator and MRPC options have been presented. For both designs a timing resolution of 100 ps appears to be very feasible. For large or full scale prototypes resolutions of 85 ps and < 60 ps have been measured for the plastic scintillator and MRPC options respectively [19]. For this study fully distinguishable timing windows were conservatively defined as $3\sigma = 340$ ps.

From the well reconstructed rates in Table. 4.2 the probability of two background muon tracks occurring in a single timing window of 340 ps can be calculated with Poisson statistics. The mean number of events expected in a single timing window, λ, is,

$$\lambda = R_\mu \times \tau \equiv R_\mu \times 340\,\text{ps} \tag{5.1}$$

where R_μ is the well reconstructed muon rate provided in the right-most column of Table 5.2, and τ is the size of each timing window. The probability for a combinatorial event in a single window is then,

$$P_{combi}^{340\,\text{ps}} = 1 - \frac{\lambda^0 \times e^{-\lambda}}{0!} - \frac{\lambda^1 \times e^{-\lambda}}{1!}. \tag{5.2}$$

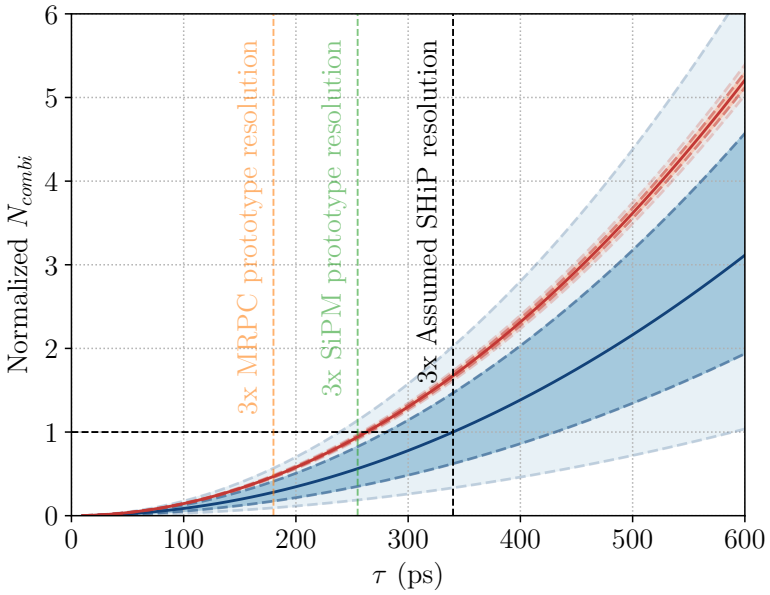


Figure 5.2: Relationship between timing window, τ and the expected number of combinatorial events over 5 years, N_{combi}. N_{combi} is normalized to the value expected from the fully simulated muon sample and $\tau = 340$ ps. In blue is the relationship calculated using the rate observed in the fully simulated sample, and in red using the rate in the GAN based sample.

The total number of expected combinatorial background events over 5 years is,

$$N_{combi}^{340\,\mathrm{ps}} = P_{combi}^{340\,\mathrm{ps}} \times N_{windows} \times N_{spills} \tag{5.3}$$

where the number of timing windows per spill $N_{windows} = 1\,\mathrm{s} \times \tau^{-1}$ and the number of spills $N_{spills} = 4 \times 10^6$. The background suppression efficiency of the timing detector is calculated by comparing $N_{combi}^{340\,ps}$ to the number of combinatorial events expected with no timing resolution. In this case spills would still be separable, so the effective timing resolution would be 1 s. The number of expected muons per timing window would then just be equal to the rate, R_μ, and $N_{combi}^{1\,s}$ would be the number of unique pairs of these muons,

$$N_{combi}^{1\,\mathrm{s}} = \frac{R_\mu \times (R_\mu - 1)}{2} \times N_{spills}. \tag{5.4}$$

The background reduction efficiency ϵ_τ is then,

$$\epsilon_\tau = N_{combi}^{340\,\mathrm{ps}} \Big/ N_{combi}^{1\,\mathrm{s}}. \tag{5.5}$$

The relationship between N_{combi}^{τ} and τ is presented in Fig. 5.2 for the values of R_μ obtained from the simulation of the fully simulated muon sample and the GAN based muon sample. As

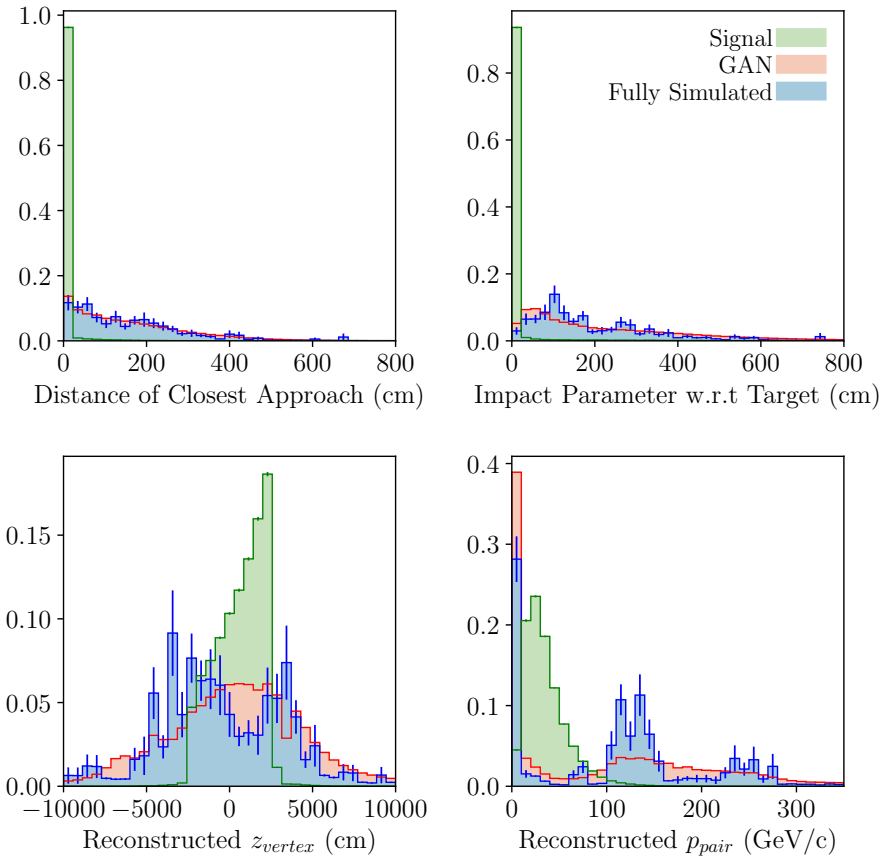


Figure 5.3: Distributions of the reconstructed properties of pairs of tracks. Fully simulated background muon pairs are presented in blue and GAN based background muon pairs in red. In green are the distributions for a benchmark signal sample.

mentioned, the GAN based sample has a higher well reconstructed rate, this results in a higher expected number of combinatorial events in 5 years for all values of τ.

5.2.2 Selection cuts

SHiP is designed to observe the decay of a HS particle within the vacuum decay volume. This particle can decay to a final state producing two or more tracks in the SHiP spectrometer. The second approach to background reduction is to exploit the differing characteristics of signal and combinatorial background events at SHiP.

From all N_μ well reconstructed muons we can produce N_{pairs} unique pairs of tracks where,

$$N_{pairs} = \frac{N_\mu \times (N_\mu - 1)}{2}. \tag{5.6}$$

Each pair is then given a weight w, calculated as $w_i \times w_j$. From each pair of tracks a variety of signal candidate properties can be calculated. These are qualitatively summarised as follows:

- the distance of closest approach (DOCA) is the shortest distance between the extrapolated paths of the two charged tracks. A signal pair of charged tracks will point back to a real decay vertex and DOCA will be a very small value.

- the signal candidate reconstructed decay vertex (r_{vertex}) is the $\{x, y, z\}$ coordinates of the DOCA position. A signal vertex is required to be inside the fiducial volume and not within a safety margin of 5 cm to a vessel wall.

- the signal candidate momentum (p_{pair}) is the vector sum of the momenta of each track at the reconstructed vertex.

- the impact parameter with respect to the target (IP) is the distance between the centre of the target and the backwards extrapolated origin of the signal candidate. The momentum of the reconstructed signal candidate particle is extrapolated in a straight line back to the z- coordinate of the SHiP target. A simple straight extrapolation is all that is necessary as it is imagined the reconstructed signal candidate particle is a HS particle.

The distributions of these derived properties are presented in Fig. 5.3 for the fully simulated and GAN based background samples. The structure in the fully simulated sample is exaggerated by the large weights present in the sample, this especially true of any pair made up of two highly weighted muons. As each highly weighted muon appears in multiple pairs the error bars presented are likely underestimating the uncertainty in each bin. Additionally a νMSM signal model with a HNL mass of 1 GeV/c^2 is provided, in this model the HNL decays to $\mu^- + \pi^+$ and is fully reconstructed. The differing characteristics of these signal and background distributions can be exploited by simple single valued selection cuts. The cuts developed by the collaboration are presented in Table. 5.3.

The efficiency of a particular cut or a combination of cuts, $\epsilon_{selection}$, can be defined as,

$$\epsilon_{selection} = W_{post-cuts}/W_{pre-cuts}, \tag{5.7}$$

Cut description	Value
Individual track momentum	$>$ 1.0 GeV/c
DOCA	$<$ 1 cm
Signal candidate decay vertex	in fiducial volume, distance to closest wall $>$ 5 cm
Impact parameter w.r.t target	$<$ 2.5 m

Table 5.3: Single value background selection cuts.

Cut description	ϵ - Full Simulation	ϵ - GAN	ϵ - Signal
Individual track momentum	$7.55\pm0.72 \times 10^{-1}$	$7.49\pm0.01 \times 10^{-1}$	$9.81\pm0.04 \times 10^{-1}$
Distance of closest approach	$1.41\pm1.09 \times 10^{-2}$	$7.15\pm0.05 \times 10^{-3}$	$8.64\pm0.03 \times 10^{-1}$
Vertex position (fiducial)	$2.38\pm0.33 \times 10^{-1}$	$3.07\pm0.03 \times 10^{-1}$	$8.88\pm0.03 \times 10^{-1}$
Impact parameter w.r.t target	$6.84\pm0.68 \times 10^{-1}$	$6.24\pm0.06 \times 10^{-1}$	$9.67\pm0.04 \times 10^{-1}$
Combined	$3.68\pm1.51 \times 10^{-4}$	$8.43\pm0.17 \times 10^{-4}$	$8.01\pm0.03 \times 10^{-1}$

Table 5.4: Efficiency of background reducing cuts. Values are presented for each cut individually and then combined.

where $W_{pre-cuts}$ and $W_{post-cuts}$ are the sum of the weights of signal candidate pairs of tracks before and after cuts respectively. Values for the efficiency of each cut applied individually and when combined together are presented in Table. 5.4. Due to the extremely large number of unique pairs that can be constructed from the GAN based sample, only a randomly selected sub-sample of track pairs is used. Also provided in Table. 5.4 are efficiencies for the HNL signal sample. Inevitably some signal will be lost to these cuts, the extent of this loss is dependent on the characteristics of the signal. To accommodate for signal particles decaying to final states that include neutralinos, which carry some momentum information out of the reconstructed system, the IP cut is very wide at 2.5 m. Fully reconstructed final states point back to within $\mathcal{O}(10)$ cm. A stricter 10 cm cut can later be applied to any candidate that is fully reconstructed [18].

Muons that emanate from the target have their kinematics perturbed by the active muon shield. This is especially true of low momentum muons. Low momentum background muons that do form tracks enter the vacuum decay volume from a wide range of angles and so are not likely to point back to the target. There is a population of high momentum background muons that

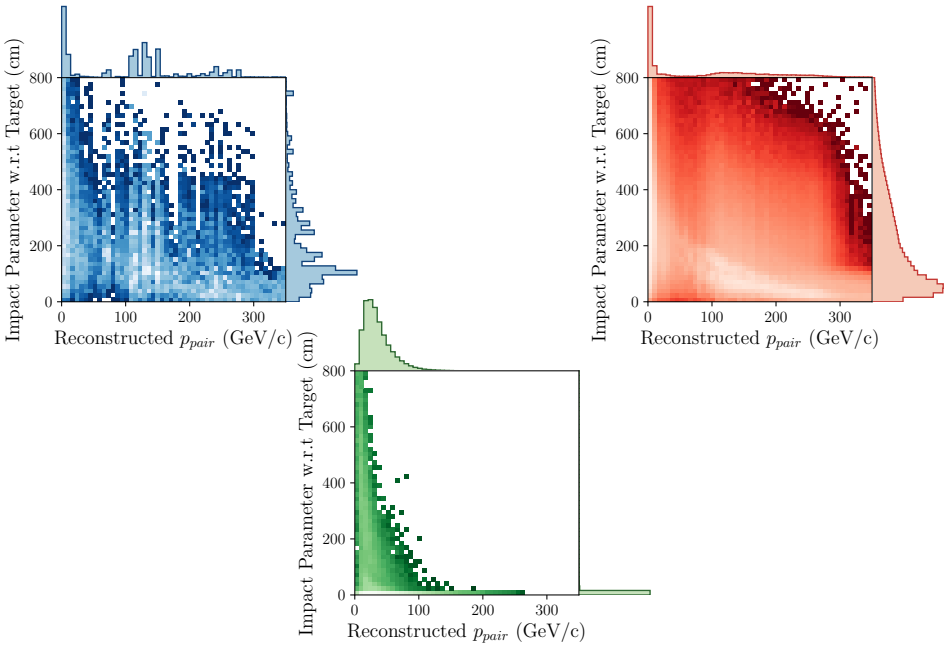


Figure 5.4: Relationship between the IP feature and the reconstructed momentum of the signal candidate particle. In blue is the fully simulated sample, in red the GAN based and in green is a benchmark signal sample.

pass the shield without much deflection, these enter the decay volume from the front. Pairs of these muons are the most likely to be mistaken as signal as they could be reconstructed to signal candidates with a low IP values. Figure. 5.4 shows a large population of background pairs that form signal candidate mother particles of high momentum which pass the wider IP cut of 2.5 m.

To visualise the effects of the IP cut Fig. 5.5 displays the reconstructed origin position after extrapolation in the x- and y- plane at the z- position of the target.

5.2.3 Upstream vetoing systems

The final defense employed to suppress combinatorial background are the vetoing sub-systems that surround the decay volume. At the front of the vessel are multiple resistive plate chamber detectors making up the upstream background tagger. Placed around the outside of the decay volume, covering almost 360 degrees, is the liquid scintillating surrounding background tagger. To enter into the acceptance of the HS spectrometer a background muon must pass through at least one of these detectors. Therefore information from these detectors can be used to veto any

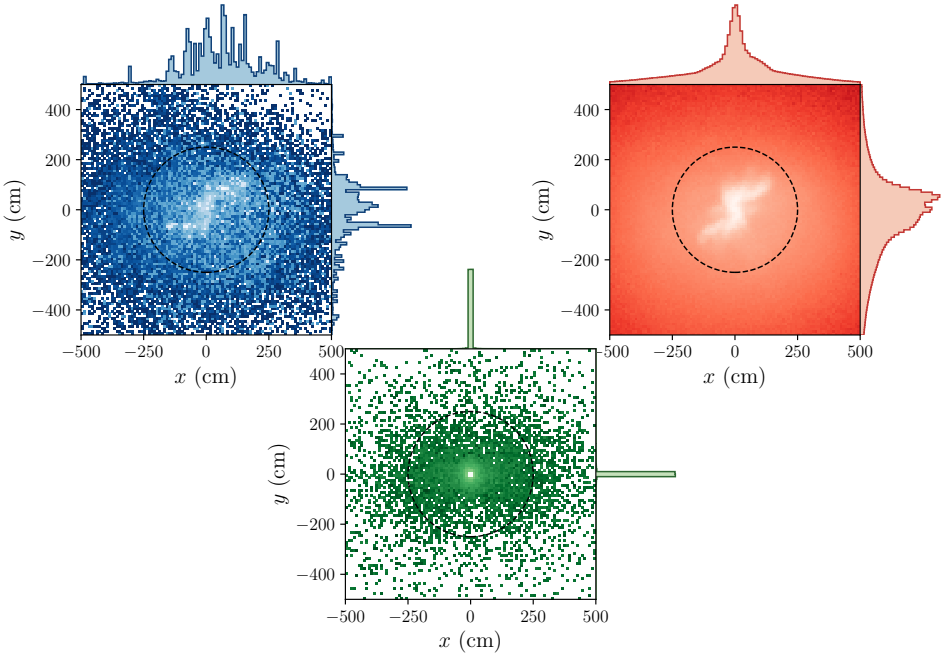


Figure 5.5: Distributions of the extrapolated x-, y- coordinates of the signal candidate particle back to the z- position of the SHiP target. The fully simulated background muon sample is in blue, the GAN based muon background in red and a benchmark signal sample in green.

timing window that its known to be contaminated. Unnecessarily vetoed timing windows can arise from background particles making hits in the veto systems but not in the straw tracker or noise in the veto systems themselves. This effect is suppressed using timing information, the UBT and SBT systems will have timing resolutions of 300 ps and 1 ns respectively [18, 19]. To further avoid signal loss, tracks in the spectrometer are extrapolated back to the veto systems. The distance between the extrapolated point and recorded position in the simulation will be referred to as Δr. A threshold is defined on Δr for a veto to be enforced. The choice of this distance threshold will affect the background reduction efficiency of the systems. Of course a larger threshold will remove more background, however if the threshold is too large signal losses from noise in upstream systems may start to impact discovery potential. This extrapolation and requirements such as requiring hits in successive RPCs of the UBT will help reduce any signal loss effects.

For each background sample the distribution of muon hits in each subsystem is displayed in Fig. 5.6, and the rates of well reconstructed muons making hits in each upstream sub-system is

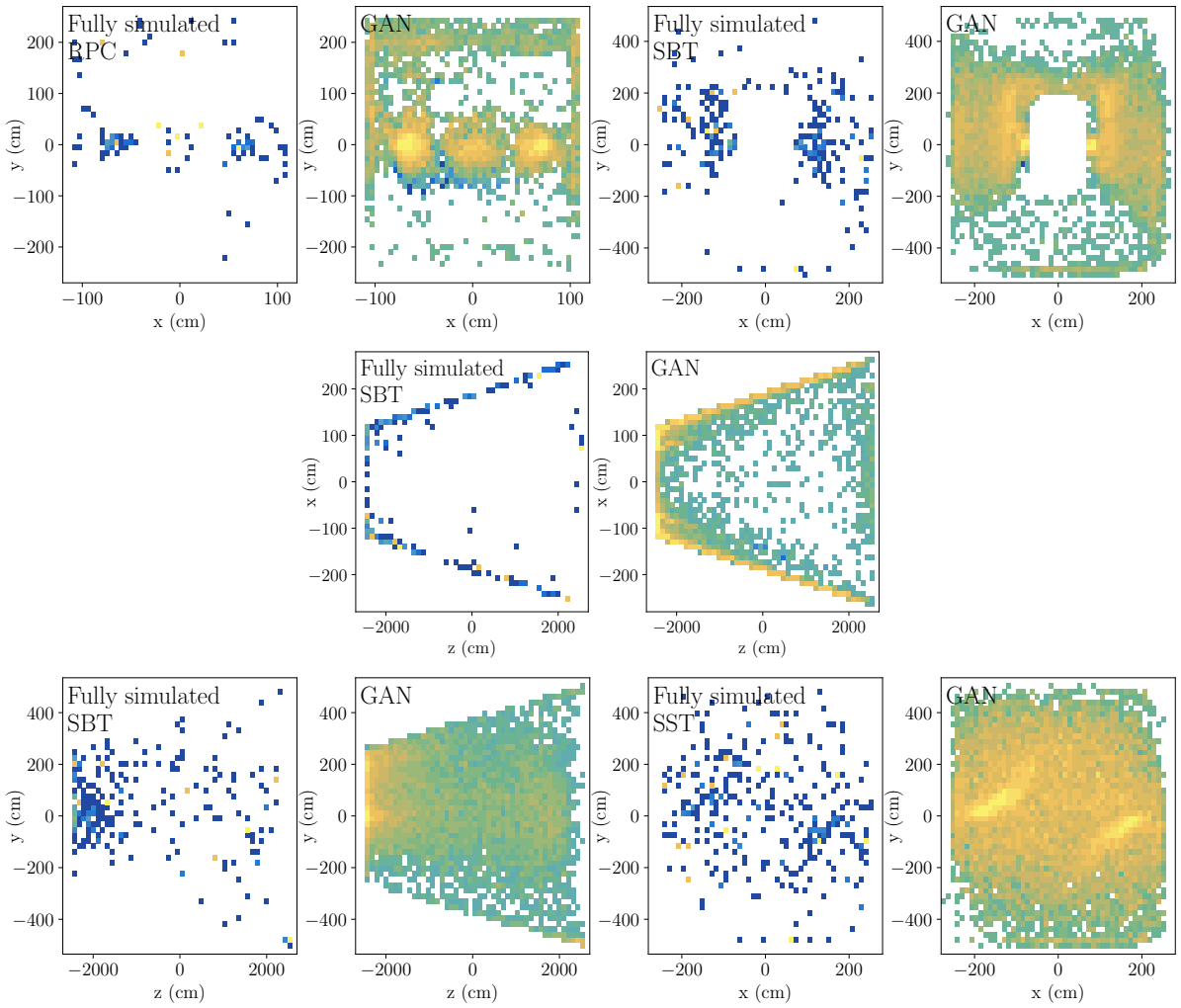


Figure 5.6: Hits made by well reconstructed muons in the downstream detectors of SHiP, only the first hit in each detector is shown.

Sub-system	Rate - Full simulation (Hz)	Rate - GAN (Hz)
μ hits both SBT and RPC	853 ± 355	2872 ± 55
μ hits RPC only	4770 ± 1756	3591 ± 79
μ hits SBT only	5825 ± 1630	8215 ± 84

Table 5.5: Rates of background muons, that form well reconstructed tracks in the HS spectrometer and that pass through each vetoing system.

presented in Table 5.5. The fully simulated rates are dominated by a small number of events with large weights. Even considering this, the rates appear significantly different between the fully simulated and GAN samples. This is especially true of muons inducing hits in both the SBT and the RPC. Figure 5.7 presents the initial kinematics of each population of muons at their point

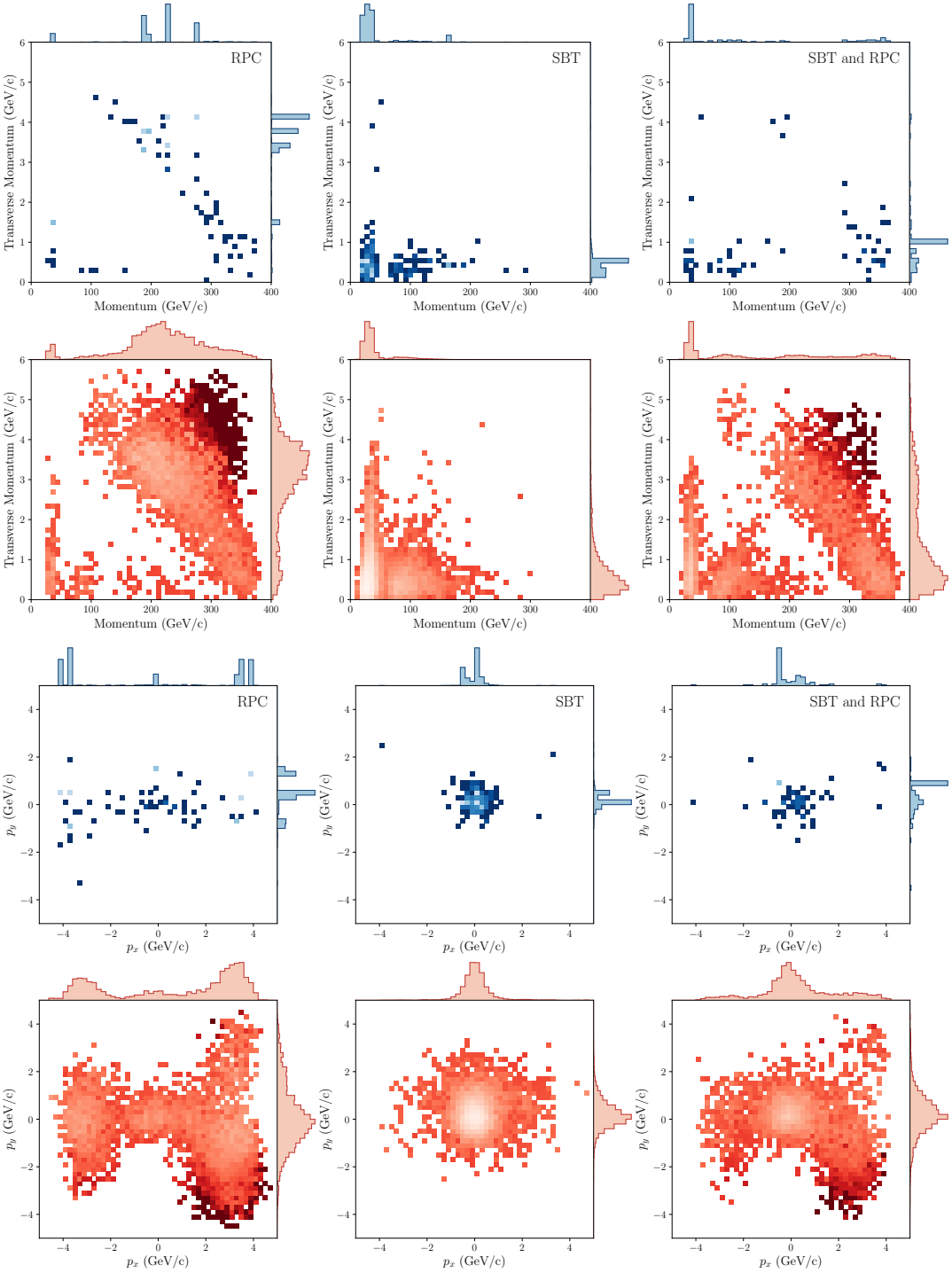


Figure 5.7: Initial momentum of muons at point of production in the SHiP target, muons that hit the SBT only, RPC both and both sub-systems are presented separately. Fully simulated samples are presented in blue and GAN samples in red.

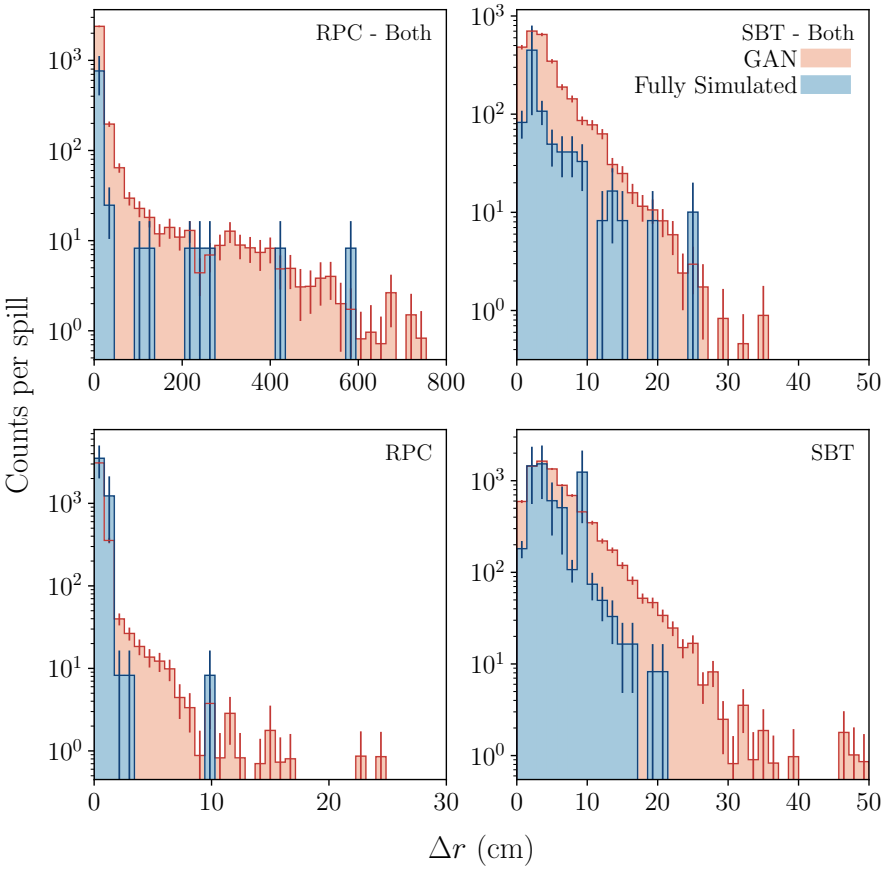


Figure 5.8: Distributions of Δr values for muons presented separately for each group of muons that hit only the SBT system, only the RPC and those that hit both systems. The fully simulated muon sample is presented in blue and the GAN based sample in red.

of production in the target. This reveals significant miss modelling in the GAN that manifests as a region of overpopulation most clearly shown in the p_x- p_y- plane. The KDE re-weighting of GAN muons is only somewhat successful at controlling this excess. The muon shield and the SHiP detectors are symmetrical in the plane around $p_x = 0$, so equal amounts of reconstructed muons are expected either side of this line. Looking just at muons with $|p_x| > 1$ the fraction with positive p_x is 75% before any re-weighting. Although this drops to 64% when events are weighted this excess is still significant. This is unlikely to explain the large discrepancy observed in the SBT and RPC rate which appears to be a mix of the SBT only and RPC only populations, where individually there is agreement between GAN and fully simulated sample rates. It is not clear how to decipher the origin of this discrepancy with the small sample sizes available.

Each well reconstructed muon is extrapolated back to the RPC and SBT systems and Δr is

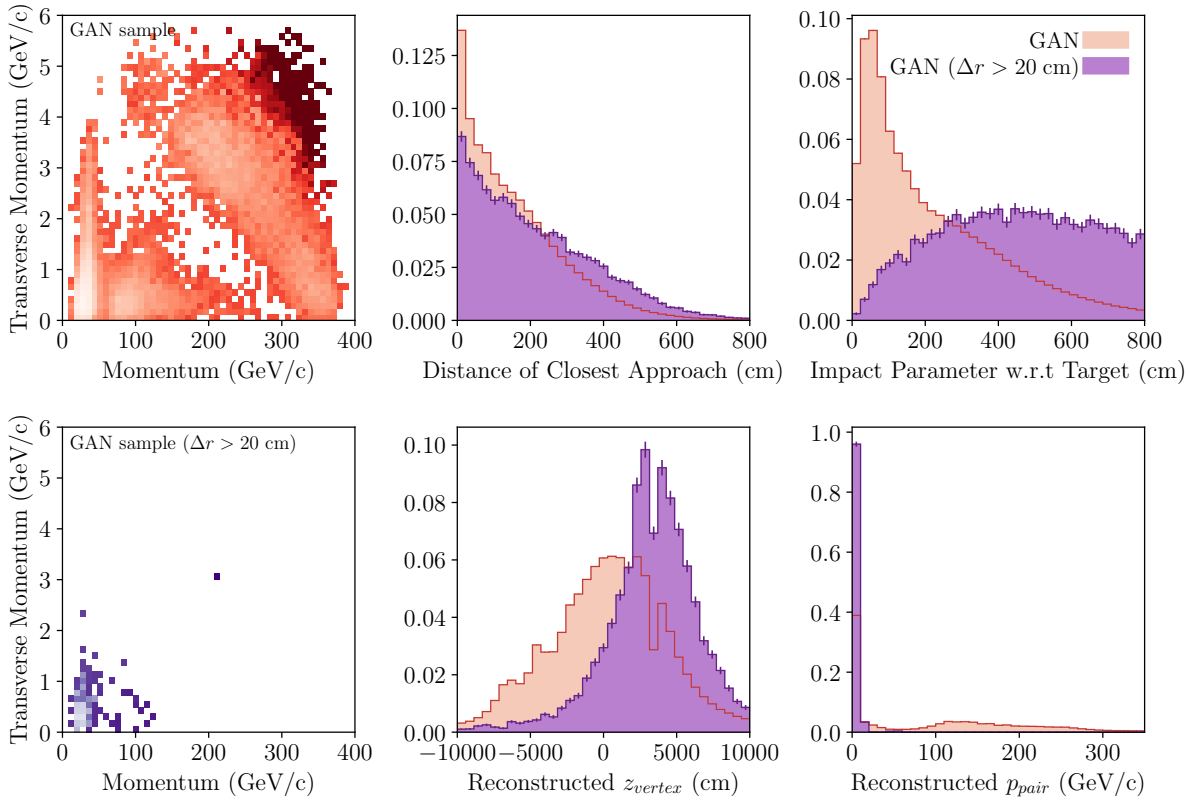


Figure 5.9: Normalised distributions of signal candidate properties discussed in Section. 5.2.2. In purple are distributions for all unique pairs tracks of GAN based muons who have values of Δr greater than 20 cm. In red are the distributions for whole of the GAN based sample, as previously presented in Fig. 5.3.

Cut description	ϵ - GAN	ϵ - GAN (muons $\Delta r > 20$ cm)	Ratio
Individual track momentum	$7.49 \pm 0.01 \times 10^{-1}$	$3.58 \pm 0.06 \times 10^{-1}$	0.478 ± 0.008
Distance of closest approach	$7.15 \pm 0.05 \times 10^{-3}$	$4.15 \pm 0.55 \times 10^{-3}$	0.580 ± 0.08
Vertex position (fiducial)	$3.07 \pm 0.03 \times 10^{-1}$	$8.47 \pm 0.26 \times 10^{-2}$	0.276 ± 0.009
Impact parameter w.r.t target	$6.24 \pm 0.06 \times 10^{-1}$	$1.49 \pm 0.04 \times 10^{-1}$	0.239 ± 0.007
Combined $\epsilon_{selection}$	$8.43 \pm 0.17 \times 10^{-4}$	N/A	-

Table 5.6: Efficiency of background selection cuts. Values are presented for each cut individually and then combined. Here "N/A" indicates that all samples were removed by the selection.

calculated for each track. The distributions of Δr are presented in Fig. 5.8, again separately based on which vetoing system they make hits in. A range of values for the threshold on Δr are selected for investigation, 20 cm and 40 cm are chosen as examples. The efficiency, ϵ_{veto}, can be

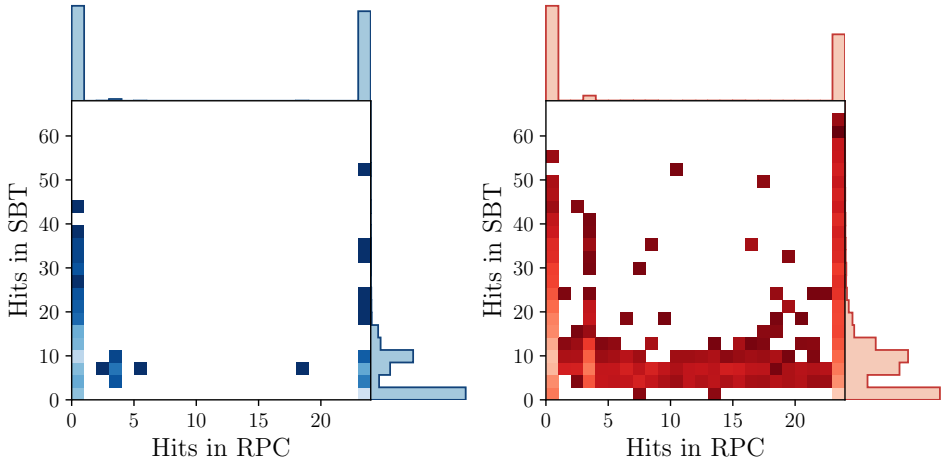


Figure 5.10: Distributions of the number of hits in each sub-system for each well reconstructed muon. In blue is the fully simulated sample and in red the GAN based sample.

calculated using the sum of the weights of the pairs of tracks that pass each threshold value. In the fully simulated sample only one muon track has a value of Δr greater than 20 cm. Due to the small sample size, for the final estimation of the combinatorial background rate using the fully simulated muons it must be assumed that ϵ_{veto} factorises with $\epsilon_{selection}$. While this assumption will also be used for the larger GAN based muon sample, there are now enough events to investigate its validity.

From the GAN based sample all muon tracks with $\Delta r > 20$ cm are collected, this sample is presented in purple in Fig. 5.9. The vetoing cut softens the momentum spectra of the background sample. Note, that the surviving muons are all from the core of the original muon kinematical distribution used to train the GAN, this is the region where the GAN approach is expected to be most accurately modelling the true distribution of muons. From only this sub-sample all unique pairs of muons are created and signal candidate properties are again calculated. The normalised distributions of these properties are presented alongside those of the full GAN muon sample in Fig. 5.9. Clearly these distributions are different and the assumption previously used that these two efficiencies factorise is certainly not valid. Table 5.6 presents a comparison between the efficiencies of each individual selection cut on the unique pairs of the full GAN based sample and the unique pairs of the sub-sample ($\Delta r > 20$ cm). For each cut the efficiency is significantly improved on the sub-sample. The ratio of selection efficiencies is also presented in Table 5.6. The

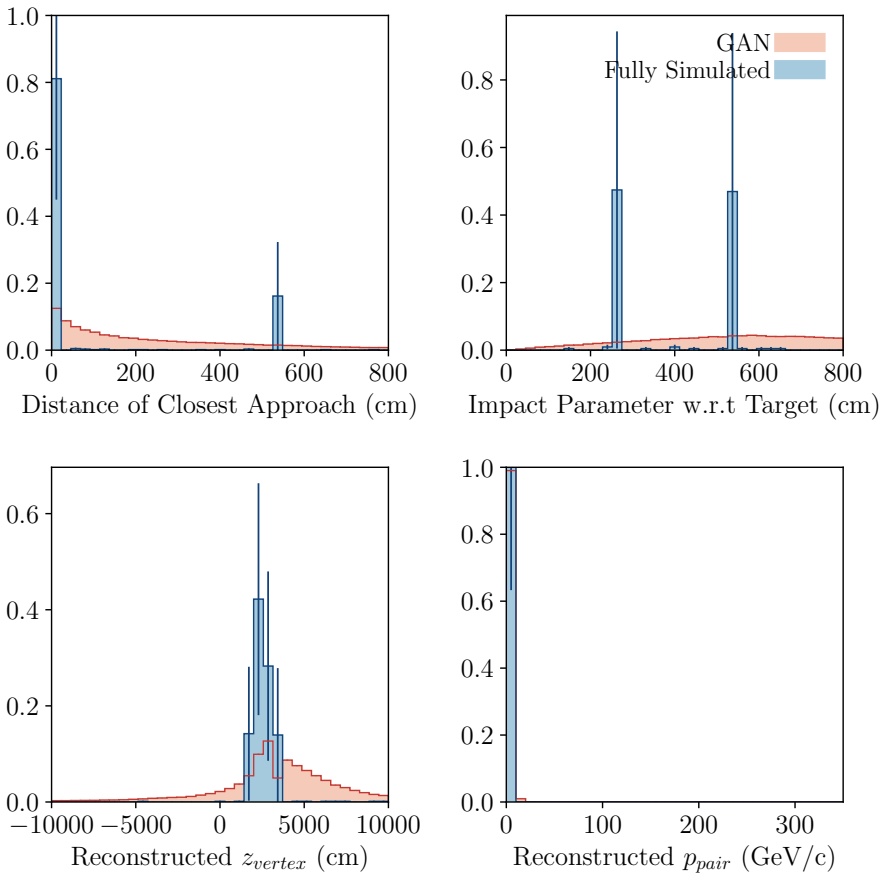


Figure 5.11: Distributions of signal candidate properties discuss in Section. 5.2.2 for all unique pairs of muons tracks that have less than 4 hits in either upstream vetoing sub-system. The fully simulated sample is presented in blue, the GAN based sample in red.

largest gain arises in the IP cut. This demonstrates that any final estimate of the combinatorial background rate that uses the factorisation assumption is likely a robust upper limit. For the GAN based sample, combining selection cuts removes all pairs from the $\Delta r > 20$ cm sub-sample. A value for the full efficiency of the cuts is therefore not deduced.

The efficiencies of the upstream veto detectors and the probability for muons to be missed must also be assessed. As mentioned in Sec. 5.2.1, the final design of these detectors has not been settled but if a conservative efficiency of 95% is assumed for both the RPC and SBT systems [19], a muon with 4 truth level hits in the simulation then has a probability of 6×10^{-6} of being missed. Figure 5.10 displays the number of hits of each muon in each sub-system. Only a small fraction have < 4 hits in either system. Figure 5.11 presents the properties of all unique

pairs constructed from a sub-sample of muons that have less than 4 hits in either system. If we enact the selection cuts from Table 5.3 on these properties, all signal candidates are easily removed. It is therefore assumed that muons passing through upstream vetoing sub-systems undetected have negligible impact on the combinatorial background rate.

5.3 Background rate estimation

The expected number of combinatorial events over 5 years can then be calculated as,

$$N_{combi} = N^{1\,\mathrm{s}}_{combi} \times \epsilon_{\tau} \times \epsilon_{selection} \times \epsilon_{veto}, \tag{5.8}$$

assuming each suppression efficiency factorises. A summary of results obtained from the fully simulated sample is presented in Table 5.7 and for the GAN sample in Table 5.8. For the fully simulated sample it is shown that the combinatorial background rate can be controlled to < 0.1 events over the full 5 years data taking at SHiP. This estimation can be viewed as an upper bound, based on the following conservative approximations made in the calculation. Firstly, a cautious value is taken for the timing resolution of the TD and secondly, as shown for the larger GAN based sample, the factorisation assumption reduces the effectiveness of the selection cuts.

	Full Simulation		
	20 cm upstream veto 250 cm IP cut	20 cm upstream veto 10 cm IP cut	40 cm upstream veto 250 cm IP cut
$N^{1\,\mathrm{s}}_{combi}$	$2.62\pm0.78 \times 10^{14}$	$2.62\pm0.78 \times 10^{14}$	$2.62\pm0.78 \times 10^{14}$
Timing Resolution ϵ_{τ}	$3.40\pm1.44 \times 10^{-10}$	$3.40\pm1.44 \times 10^{-10}$	$3.40\pm1.44 \times 10^{-10}$
Selection Cuts $\epsilon_{selection}$	$3.68\pm1.51 \times 10^{-4}$	N/A	$3.68\pm1.51 \times 10^{-4}$
Upstream Veto ϵ_{veto}	$7.19\pm7.19 \times 10^{-4}$	$7.19\pm7.19 \times 10^{-4}$	N/A
N_{combi} events in 5 years	0.024 ± 0.028	N/A	N/A

Table 5.7: Summary of the total combinatorial background rate using the fully simulated sample.

	GAN		
	20 cm upstream veto 250 cm IP cut	20 cm upstream veto 10 cm IP cut	40 cm upstream veto 250 cm IP cut
$N^{1\,s}_{combi}$	$4.38\pm0.05 \times 10^{14}$	$4.38\pm0.05 \times 10^{14}$	$4.38\pm0.05 \times 10^{14}$
Timing Resolution ϵ_τ	$3.40\pm0.06 \times 10^{-10}$	$3.40\pm0.06 \times 10^{-10}$	$3.40\pm0.06 \times 10^{-10}$
Selection Cuts $\epsilon_{selection}$	$8.43\pm0.17 \times 10^{-4}$	$4.51\pm1.08 \times 10^{-6}$	$8.43\pm0.17 \times 10^{-4}$
Upstream Veto ϵ_{veto}	$9.44\pm0.74 \times 10^{-3}$	$9.44\pm0.74 \times 10^{-3}$	$6.49\pm1.97 \times 10^{-4}$
N_{combi} events in 5 years	1.19 ± 0.10	0.0063 ± 0.0016	0.081 ± 0.025

	GAN
	No factorisation
N_{combi} events in 5 years	N/A

Table 5.8: Summary of the total combinatorial background rate using the GAN sample.

The GAN based estimation using the 20 cm vetoing cut has the expected number of background events in excess of 0.1 events. This is not viewed as an issue as these results also include the factorisation assumption. Not including this assumption removed all signal candidates within the GAN sample. The GAN sample also includes some production excesses in dangerous regions which are not fully accounted for in the corrective KDE weighting, this only makes the GAN sample a more conservative test of these background defenses.

The middle columns of Tables 5.7 and 5.8 present background rates with an IP cut of 10 cm. The estimate for the background obtained is well below the goal of < 0.1 events. For a signal candidate that is fully reconstructed and that points back to the target to within 10 cm the collaboration can therefore be extremely confident in confirming this is not a combinatorial background event.

In this study $\epsilon_{selection}$ is calculated based on simple single value selection cuts. A small improvement in background discrimination could be achieved by instead using a multi-variate method such as a BDT. This would consider relations between signal candidate features in a decision. This may also help improve the conservation of true signal events of which the single

value cuts can remove up to 20% of depending on the signal properties. To avoid overfitting on such a small sample of combinatorial events and to obtain good performance on a wide variety of unknown signal flavours, a BDT could be trained on pre-binned values of each signal candidate feature [138].

Chapter 6

R-parity violating neutralinos

As introduced in Section 2.3.5, the SHiP experiment will be sensitive to supersymmetic models that do not include the conservation of the R-parity quantum number. For sufficiently small couplings or small neutralino masses, the lightest supersymmetric particle, the RPV neutralino, can be long-lived and decay to SM particles inside the decay volume of the SHiP detector. Neutralino candidates can then be reconstructed through their decay products that can go on to form tracks in the spectrometer of the SHiP detector.

The sensitivity of the SHiP experiment to light and long-lived RPV neutralinos has previously been studied using toy simulations, this is described in Ref. [67]. However, this study did not account for the cascade enhancement of heavy-flavour production in the SHiP target which has now been estimated [70]. For simplicity the study also ignored the effect of backgrounds and reconstruction efficiencies. This chapter presents updated sensitivities of the SHiP experiment using a realistic simulation of the newly optimised detector configuration, reconstruction algorithms, selection efficiencies and the cascade enhancements of heavy flavour production in the SHiP target.

As the dominant production channel at SHiP, this study focuses on neutralino production from the decays of charm and bottom mesons via the λ'_{ijk} couplings of equation 2.30. The partial

width for such production channels are proportional to $(\lambda'_{ijk})^2$. Figure 2.4 presents diagrams of an example meson production and neutralino decay channel. The widths of various production and decay channels are obtained following the formalism described in Ref. [67].

For a single lepton flavour, there exists 36 unique pairs of production and decay λ'_{ijk} couplings, denoted as (λ'_P, λ'_D). From these a selection of benchmark scenarios were considered in Ref. [67], these were subsequently used in Refs. [62, 139] to assess the sensitivity of other proposed facilities including CODEX-b [140], FASER [141, 142] and MATHUSLA [143], again with toy simulations. These already established benchmarks are used again in this study, these are labelled Benchmarks 1–5. In addition, this study provides sensitivity estimates for unexplored scenarios involving neutralino production from B_s^0 and $B_c^\pm$ meson decays, denoted as Benchmarks 6, 7. Furthermore, new benchmarks involving fully leptonic neutralino decays, denoted as Benchmarks 8, and 9 are also presented. A summary of each benchmark is provided in Table 6.1, where the couplings involved along with the accessible production and decay modes of the neutralino are provided.

Coupling pairs that only provide sensitivity to a small region of $m_{\widetilde{\chi}_1^0}$ are not considered. For example, a scenario in which only the couplings λ'_{132} and λ'_{113} were non-zero could only be detected at SHiP if $\widetilde{\chi}_1^0$ production occurs in the following channel,

$$\text{via } \lambda'_{132} \ : \quad B_s^0 \ \rightarrow \ \widetilde{\chi}_1^0 \ + \ \nu_e \ / \ \bar{B_s}^0 \ \rightarrow \ \widetilde{\chi}_1^0 \ + \ \bar{\nu}_e, \tag{6.1}$$

where the mass of B_s^0 is ~ 5.37 GeV/c^2 and neutralino decay occurs in the following channel,

$$\text{via } \lambda'_{113} \ : \quad \widetilde{\chi}_1^0 \rightarrow \ \begin{cases} B^\pm + e^\mp, \\ B^0 + \nu_e \ / \ \bar{B}^0 + \bar{\nu}_e, \end{cases} \tag{6.2}$$

where the $B^\pm$ and B^0 are only slightly lighter with masses ~ 5.28 GeV/c^2. Benchmarks that include λ'_{i11} couplings allowing the neutralino to decay to pion final states are not studied due to large uncertainties in the simulation of light mesons with `Pythia` [67]. Additionally, neutralino production from pion decays via λ'_{i11} couplings is not studied as the measured width of the pion provides stringent constraints to such models which SHiP would not be sensitivity enough to overcome [144]. Although $\mathcal{O}(10^{20})$ kaons are expected to be produced at the BDF, most of these are stopped in the SHiP target or hadron absorber before decaying. Therefore,

neutralino production from K decays is also not considered. Direct production of neutralinos in pair production via t-channel squark exchange was shown to be small compared to production from meson decays, see Ref. [67, 145], so is not considered. The production of neutralinos in decays of vector mesons is also expected to be negligible [67]. For example, for $D^{*\pm}$ and $D^{\pm}$ mesons the width of decay channels to $\widetilde{\chi}_1^0$ are roughly the same however the total decay with of $D^{*\pm}$ is much larger than that of $D^{\pm}$.

This study focuses on couplings to first generation leptons (λ'_{1jk}). The sensitivity to second generation lepton couplings, λ'_{2jk}, is similar to that of λ'_{1jk} away from the edges of the phase space of $\widetilde{\chi}_1^0$ production and decay. A dedicated benchmark studying the sensitivity to third generation lepton couplings, λ'_{3jk}, is included.

The present bounds on the λ'_{ijk} couplings [68] explored in this chapter are shown below. These limits, and the decay width expressions used later in these studies, depend on the mass of the sfermions involved, $m_{\tilde{f}}$. For each benchmark scenario all sfermions are set to have degenerate mass, where $m_{\tilde{f}} = 1$ TeV/c^2. This allows for easy comparison with sensitivity studies of other facilities which make the same definition. The limits on $\lambda'_{ijk}/m^2_{\tilde{f}}$ for $m_{\tilde{f}} = 1$ TeV/c^2 are also given:

$$
\begin{aligned}
\lambda'_{112} &< 0.03\,\frac{m_{\tilde{s}_R}}{100\ \mathrm{GeV/c^2}}, & \lambda'_{112}/m^2_{\tilde{f}} &< 3\times10^{-7}\,(\mathrm{GeV/c^2})^{-2},\\
\lambda'_{122} &< 0.2\,\frac{m_{\tilde{s}_R}}{100\ \mathrm{GeV/c^2}}, & \lambda'_{122}/m^2_{\tilde{f}} &< 2\times10^{-6}\,(\mathrm{GeV/c^2})^{-2},\\
\lambda'_{312} &< 0.06\,\frac{m_{\tilde{s}_R}}{100\ \mathrm{GeV/c^2}}, & \lambda'_{312}/m^2_{\tilde{f}} &< 6\times10^{-7}\,(\mathrm{GeV/c^2})^{-2},\\
\lambda'_{132} &< 0.28\,\frac{m_{\tilde{t}_L}}{100\ \mathrm{GeV/c^2}}, & \lambda'_{132}/m^2_{\tilde{f}} &< 2.8\times10^{-6}\,(\mathrm{GeV/c^2})^{-2},\\
\lambda'_{121} &< 0.2\,\frac{m_{\tilde{d}_R}}{100\ \mathrm{GeV/c^2}}, & \lambda'_{121}/m^2_{\tilde{f}} &< 2\times10^{-6}\,(\mathrm{GeV/c^2})^{-2},\\
\lambda'_{131} &< 0.03\,\frac{m_{\tilde{t}_L}}{100\ \mathrm{GeV/c^2}}, & \lambda'_{131}/m^2_{\tilde{f}} &< 3\times10^{-7}\,(\mathrm{GeV/c^2})^{-2},\\
\lambda'_{313} &< 0.06\,\frac{m_{\tilde{b}_R}}{100\ \mathrm{GeV/c^2}}, & \lambda'_{313}/m^2_{\tilde{f}} &< 6\times10^{-7}\,(\mathrm{GeV/c^2})^{-2},
\end{aligned}
\tag{6.3}
$$

$$\lambda'_{123} \;<\; 0.043\,\frac{m_{\tilde{b}_R}}{100\ \mathrm{GeV/c^2}}, \qquad \lambda'_{123}/m^2_{\tilde{f}} \;<\; 4.3\times10^{-7}\,(\mathrm{GeV/c^2})^{-2}. \tag{6.3}$$

The limits have a mass dependence on the sfermion masses as sfermions are involved in the production and decay of the neutralino as shown in Fig. 2.4, the heavier the sfermion the weaker the bound on the coupling. The limits taken are derived in Ref. [68] using a sfermion mass of $100\ \mathrm{GeV/c^2}$ from experimental measurements of various relevant decay rates.

Couplings (λ'_P,λ'_D)	Production Channel(s)	Charged Decay Channel(s)	Neutral Decay Channel(s)	Notes
λ'_{121}, λ'_{112}	$D^+ \to \tilde{\chi}^0_1 + e^+$	$\tilde{\chi}^0_1 \to (K^+, K^{*+}) + e^-$	$\tilde{\chi}^0_1 \to (K^0_L, K^0_S, K^*) + \nu_e$	Benchmark 1 See Fig. 6.4(a)
λ'_{122}, λ'_{112}	$D_s^+ \to \tilde{\chi}^0_1 + e^+$	$\tilde{\chi}^0_1 \to (K^+, K^{*+}) + e^-$	$\tilde{\chi}^0_1 \to (\eta, \eta', \phi) + \nu_e$ $\tilde{\chi}^0_1 \to (K^0_L, K^0_S, K^*) + \nu_e$	Benchmark 2 See Fig. 6.4(c)
λ'_{131}, λ'_{112}	$B^0 \to \tilde{\chi}^0_1 + \nu_e$	$\tilde{\chi}^0_1 \to (K^+, K^{*+}) + e^-$	$\tilde{\chi}^0_1 \to (K^0_L, K^0_S, K^*) + \nu_e$	Benchmark 3 See Fig. 6.6(a)
λ'_{131}, λ'_{121}	$B^0 \to \tilde{\chi}^0_1 + \nu_e$	$\tilde{\chi}^0_1 \to (D^+, D^{*+}) + e^-$	$\tilde{\chi}^0_1 \to (K^0_L, K^0_S, K^*) + \nu_e$	Benchmark 4 See Fig. 6.6(c)
λ'_{313}, λ'_{312}	$B^0 \to \tilde{\chi}^0_1 + \nu_\tau$ $B^+ \to \tilde{\chi}^0_1 + \tau^+$	$\tilde{\chi}^0_1 \to (K^+, K^{*+}) + \tau^-$	$\tilde{\chi}^0_1 \to (K^0_L, K^0_S, K^*) + \nu_\tau$	Benchmark 5 See Fig. 6.8(a)
λ'_{132}, λ'_{112}	$B_s^0 \to \tilde{\chi}^0_1 + \nu_e$	$\tilde{\chi}^0_1 \to (K^+, K^{*+}) + e^-$	$\tilde{\chi}^0_1 \to (K^0_L, K^0_S, K^*) + \nu_e$	Benchmark 6 See Fig. 6.7(a)
λ'_{123}, λ'_{112}	$B_s^0 \to \tilde{\chi}^0_1 + \nu_e$ $B_c^+ \to \tilde{\chi}^0_1 + e^+$	$\tilde{\chi}^0_1 \to (K^+, K^{*+}) + e^-$	$\tilde{\chi}^0_1 \to (K^0_L, K^0_S, K^*) + \nu_e$	Benchmark 7 See Fig. 6.7(b)
λ'_{122}, λ_{111}	$D_s^+ \to \tilde{\chi}^0_1 + e^+$	$\tilde{\chi}^0_1 \to e^+ + \nu_e + e^-$		Benchmark 8 See Fig. 6.10(a)
λ'_{131}, λ_{111}	$B^0 \to \tilde{\chi}^0_1 + \nu_e$	$\tilde{\chi}^0_1 \to e^+ + \nu_e + e^-$		Benchmark 9 See Fig. 6.10(b)

Table 6.1: Details of benchmark scenarios, each defined by a pair of R-parity violating couplings. The first table block contains pairs of semi-leptonic couplings λ', the second block contains benchmarks with production via a semi-leptonic coupling followed by decay via fully leptonic couplings λ. Neutral and charged final states are separated for clarity.

6.1 Sensitivity calculations

The expected number of fully reconstructed $\widetilde{\chi}_1^0$ decays that would pass all selection criteria in 5 years of operation of the SHiP experiment, N_{events}, is estimated as

$$N_{\text{events}} = N_{\text{prod}} \times P_{\text{reco}}, \tag{6.4}$$

where N_{prod} is the number of neutralinos produced and P_{reco} is the probability for a single neutralino to be successfully reconstructed and to pass selection criteria. An estimate for N_{prod} can be calculated as

$$N_{\text{prod}} = 2 \times X_{\bar{q}q} \times f_{\text{cascade}} \times N_{\text{POT}} \times f(q \to h) \times \mathcal{B}(h \to \widetilde{\chi}_1^0 + X), \tag{6.5}$$

where $f(q \to h)$ is the h meson production fraction at SHiP, and $\mathcal{B}(h \to \widetilde{\chi}_1^0 + X)$ is the branching ratio of h mesons to accessible states that include a neutralino. Flavour production fractions and cascade enhancement factors are denoted as $X_{\bar{q}q}$ and f_{cascade} respectively and N_{POT} denotes the number of protons-on-target over the lifetime of the experiment. Details of the calculations of $\mathcal{B}(h \to \widetilde{\chi}_1^0 + X)$ and the subsequent decay widths of $\widetilde{\chi}_1^0$ to various allowed final states are described in Refs. [67] and [146]. The sfermion mass $m_{\tilde{f}}$ enters these width expressions at fourth order, $\Gamma \propto 1/m_{\tilde{f}}^4$, and as mentioned previously $m_{\tilde{f}} = 1$ TeV/c^2 is set for all benchmarks. Also provided in Ref. [67] are values for the decay constant of various mesons, this study uses the same values. Examples of how limits change with different assumptions for $m_{\tilde{f}}$ will be presented in Sec. 6.2.1.

The probability, P_{reco}, for a neutralino to decay to a particular final state, be contained within the detector acceptance and to pass the selection requirements is given by

$$P_{\text{reco}} = \mathcal{B}(\widetilde{\chi}_1^0 \to \text{final state}) \times P_{\text{decay}} \times \epsilon_{det}, \tag{6.6}$$

where $\mathcal{B}(\widetilde{\chi}_1^0 \to \text{final state})$ is the branching ratio of neutralinos to the accessible final state, and ϵ_{det} is the efficiency of the detector, which is the product of the geometrical acceptance, reconstruction efficiency and selection efficiency for $\widetilde{\chi}_1^0$ decays occurring within acceptance. Finally P_{decay} is the probability for the $\widetilde{\chi}_1^0$ to decay inside the vacuum vessel given it is within the geometrical acceptance. The kinematics of neutralinos produced both in primary and cascade

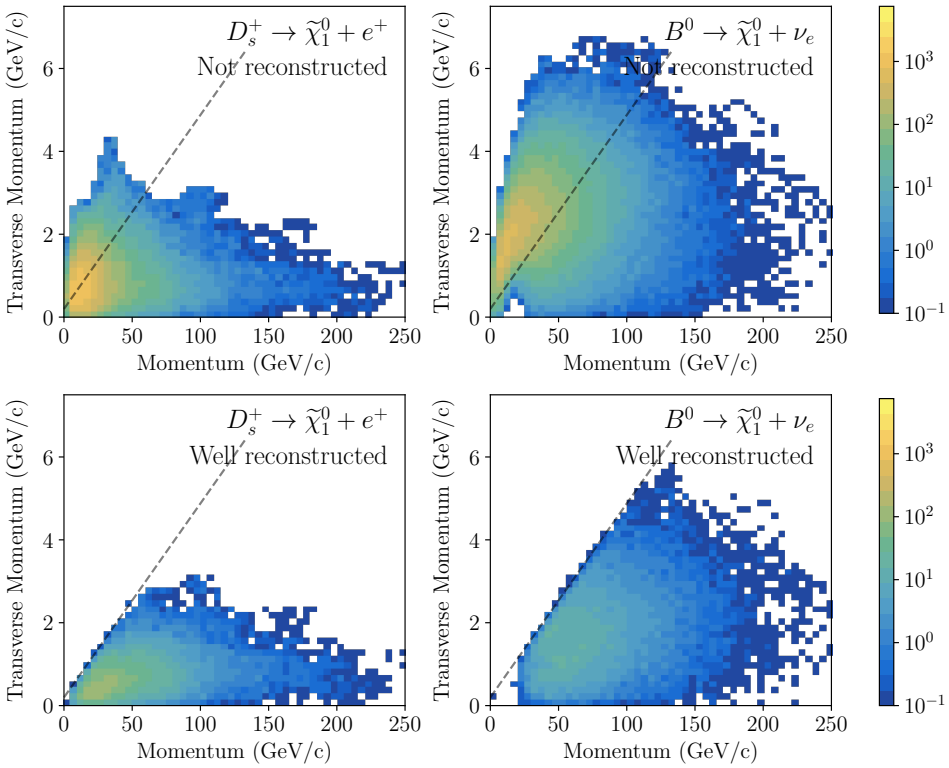


Figure 6.1: Kinematics of the $\widetilde{\chi}_1^0$ produced in decays of charm and beauty mesons. Both distributions are produced with $\lambda'/m_{\tilde{f}}^2 = 1 \times 10^{-6}$ $(\mathrm{GeV/c^2})^{-2}$. Presented in the bottom row are the kinematics of reconstructed $\widetilde{\chi}_1^0$ that fulfill selection, and in the top row are the kinematics of the $\widetilde{\chi}_1^0$ which do not fulfill selection.

heavy flavour decays are shown in Fig. 6.1. Neutralinos whose decay products are successfully reconstructed and pass selection cuts are plotted separately to those that do not.

Figure 6.2 presents the product of the reconstruction efficiency and the background selection cut efficiency given a $\widetilde{\chi}_1^0$ decays within the vacuum vessel. This quantity is presented across a plane of $m_{\widetilde{\chi}_1^0}$ and λ'_{ijk} values for $\widetilde{\chi}_1^0$ production from both charm and beauty mesons. Each case in Fig. 6.2 is split into charged and neutral final states, similar to the approach of Refs. [62, 67]. Neutral final states are detectable at SHiP through the partial reconstruction of the $\widetilde{\chi}_1^0$ from the products of the subsequent decays of the neutral meson such as $K^{*0} \to K^+\pi^-$, $K_L^0 \to \pi^+\pi^-\pi^0$ and $K_S^0 \to \pi^+\pi^-$. A component of momentum information will be lost in the neutrino leading to some inaccuracy, but this effect can be small. These signals can be distinguished from background V^0 decays produced through muon deep inelastic scattering in upstream material by requiring the momentum of the reconstructed system must point to the SHiP target. Figure 6.2 demonstrates

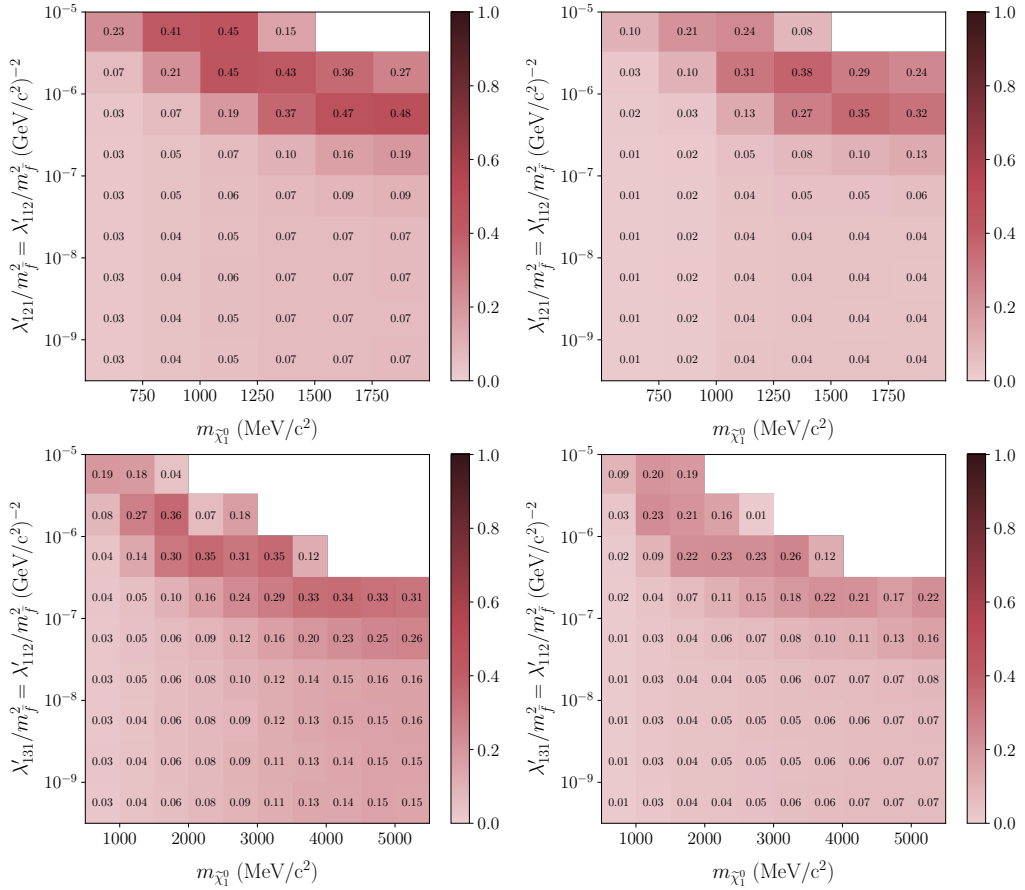


Figure 6.2: Efficiency to reconstruct and select a $\widetilde{\chi}_1^0$ candidate, given it decays within acceptance, as a function of coupling and mass. Efficiencies for (left) charged and (right) neutral decays are shown separately for (top) Benchmark 1 and (bottom) Benchmark 3.

that the sensitivity of SHiP to such neutral decays is $\sim 50\%$ lower than that of charged decays.

Finally, P_{decay} is calculated for each event as

$$P_{\text{decay}} = l_{\text{decay}} \times \left(\exp\left(-\frac{l_{\text{ini}}}{l_{\text{decay}}}\right) - \exp\left(-\frac{l_{\text{fin}}}{l_{\text{decay}}}\right) \right) \tag{6.7}$$

where l_{ini} is the distance travelled by the neutralino before entering the decay vessel, l_{fin} is the distance to the end of the decay vessel along the trajectory of the neutralino and $l_{\text{decay}} = c\gamma\tau$ is the neutralino decay length.

Within an accessible $m_{\widetilde{\chi}_1^0}$ region the shape of a sensitivity contour is dictated by both the coupling strengths and N_{prod}. Where the coupling is stronger, there are two effects. While N_{prod} is higher the lifetime of $\widetilde{\chi}_1^0$ is shorter and so a larger fraction decay of neutralinos before they reach the fiducial decay volume. Therefore SHiP loses sensitivity to larger couplings. Where the

couplings are weaker the lifetime of $\widetilde{\chi}_1^0$ is longer and a larger fraction of neutralinos decay after passing through the SHiP decay volume and furthermore the number of $\widetilde{\chi}_1^0$ produced is lower and so sensitivity drops off fast.

The simulation of events was carried out with the `FairShip` software suite, as outlined in Chapter 3.2. Simulations each of 2500 signal events were run at the points of a 50×50 grid across each plane. Events are weighted by P_{decay} and averaged. The selection criteria presented in Chapter 5 is applied to each reconstructed signal candidate. The weights of events that do not pass signal selection criteria not included in this average, this process approximates ϵ_{det}. This selection has been comprehensively shown to result in < 0.1 background events over the lifetime of the experiment. Therefore, this sensitivity study can assume zero background.

6.2 SHiP sensitivity in different benchmarks

This section presents the sensitivity of the SHiP experiment in the various benchmark scenarios outlined in Table 6.1. The first two benchmarks will be presented in the most detail as the same concepts will apply to all benchmarks.

The full sensitivity results, that include both neutral and charged final states, are the most comprehensive estimates of the SHiP sensitivity to RPV neutralinos to date. Making the assumption of zero background events the expected sensitivity at 90% confidence interval is determined by the location where the expected number of signal events over the 5 years of operation of the experiment is ≥ 2.3 [41]. Examples of picking difference limits are provided in Sec. 6.2.1.

Table 6.2 shows the fraction of proton-on-target collisions that are expected to produce $c\bar{c}$ and $b\bar{b}$ quark pairs. The production fraction of various charm and beauty mesons are provided in Table 6.3. Yields of various heavy mesons can be calculated from this information. For example, it is expected that 3.2×10^{17} $D^{\pm}$ mesons and 4.5×10^{13} B^0 mesons will be produced over the 5 year operation of the SHiP experiment. Figure 6.3 presents N_{prod} for 5 years of operation as a function of $m_{\widetilde{\chi}_1^0}$ for the various $\widetilde{\chi}_1^0$ production channels explored in this chapter. Values in Fig. 6.3 are calculated using Tables 6.2 and 6.3 and equation 6.5 where for this example the

Protons-on-target over 5 years N_{POT}	$\bar{c}c$ fraction $X_{\bar{c}c}$	$\bar{b}b$ fraction $X_{\bar{b}b}$	Cascade Enhancement $f_{cascade}$	
			charm	beatuy
2×10^{20}	1.7×10^{-3}	1.6×10^{-7}	2.3	1.7

Table 6.2: Charm and beauty production fractions and cascade enhancement factors for the SHiP experiment [41, 70].

meson	$f(q \to \text{meson})$
D^+	0.207
D^0	0.632
D_s	0.088
J/ψ	0.01

meson	$f(q \to \text{meson})$
B^+	0.417
B^0	0.418
B_s	0.113
B_c	$\leq 2.6 \times 10^{-3}$

Table 6.3: Production fraction of different mesons at SHiP taking into account cascade production [41, 70].

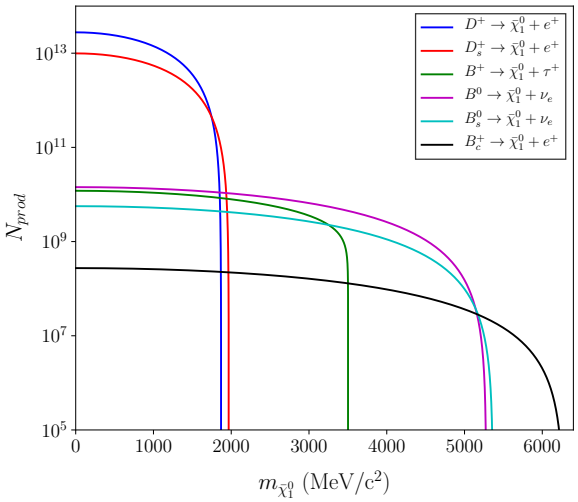


Figure 6.3: Expected numbers of $\tilde{\chi}_1^0$ from production channels allowed by some of the benchmark scenarios investigated in this paper.

production coupling is set to $\lambda'/m_{\tilde{f}}^2 = 1 \times 10^{-6}$ (GeV/c^2)$^{-2}$.

Sensitivity results for each benchmark are presented in two different planes following the format of Refs. [62, 67]. Firstly, the sensitivity is presented in the plane of λ' against $m_{\tilde{\chi}_1^0}$ with the production and decay couplings set to be equal, $\lambda'_P = \lambda'_D \equiv \lambda'$. Secondly, the sensitivity is presented in the plane of λ'_P against λ'_D for three fixed values of $m_{\tilde{\chi}_1^0}$.

6.2.1 Production from charm mesons

The first benchmarks investigated are based on $\widetilde{\chi}_1^0$ production from the decays of charm mesons. Benchmark 1 pairs the production coupling $\lambda'_P = \lambda'_{121}$ with the decay coupling $\lambda'_D = \lambda'_{112}$. Therefore, neutralino production proceeds via the channel

$$D^{\pm} \;\rightarrow\; \widetilde{\chi}_1^0 \;+\; e^{\pm}. \tag{6.8}$$

where the partial width of this process is proportional to $(\lambda'_{121})^2$. The decay of $\widetilde{\chi}_1^0$ can occur via both λ'_{112} and λ'_{121} in the following processes:

$$\text{via } \lambda'_{112} \;:\quad \widetilde{\chi}_1^0 \rightarrow \begin{cases} K^{(*)\pm} + e^{\mp} & \text{(charged)}, \\ (K_L^0, K_S^0, K^*) + \nu_e & \text{(neutral)}, \end{cases} \tag{6.9}$$

$$\text{via } \lambda'_{121} \;:\quad \widetilde{\chi}_1^0 \rightarrow \; (K_L^0, K_S^0, K^*) + \nu_e \qquad \text{(neutral)}. \tag{6.10}$$

The second benchmark investigated, Benchmark 2, has (λ'_{122}, λ'_{112}) as the pair of non-zero couplings. In this scenario neutralinos are produced from $D_s^{\pm}$ decays via λ'_{122}:

$$D_s^{\pm} \;\rightarrow\; \widetilde{\chi}_1^0 \;+\; e^{\pm}. \tag{6.11}$$

Subsequently, the decay of the $\widetilde{\chi}_1^0$ occurs via both λ'_{112} and λ'_{122} in the following decays:

$$\text{via } \lambda'_{112} \;:\quad \widetilde{\chi}_1^0 \rightarrow \begin{cases} K^{(*)\pm} + e^{\mp} & \text{(charged)}, \\ (K_L^0, K_S^0, K^*) + \nu_e & \text{(neutral)}, \end{cases} \tag{6.12}$$

$$\text{via } \lambda'_{122} \;:\quad \widetilde{\chi}_1^0 \rightarrow \; (\eta/\eta'/\phi) + \nu_e \qquad \text{(neutral)}. \tag{6.13}$$

For Benchmark 1, the sensitivity in the plane λ' against $m_{\widetilde{\chi}_1^0}$ is shown in Fig. 6.4(a), and for Benchmark 2 in Fig. 6.4(c). Two limits are provided in accordance with previous studies. Firstly, in dashed lines, sensitivities to all possible final states are presented. Secondly, in solid lines is the sensitivity if neutral final states are not included. In Fig. 6.4(a) and Fig. 6.4(c) the gain in sensitivity from including all final states is small. However, in other benchmarks this difference will be more substantial.

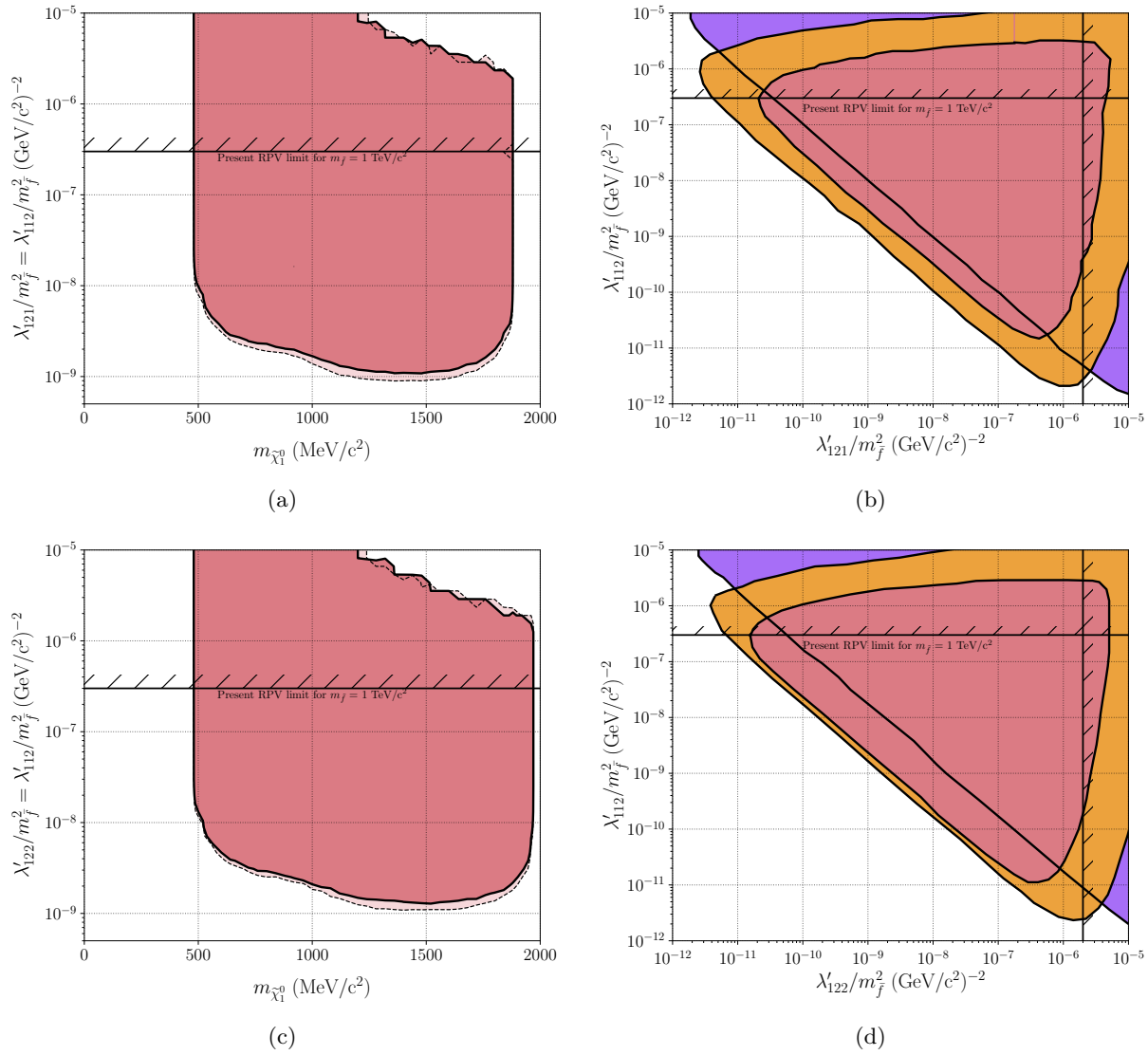
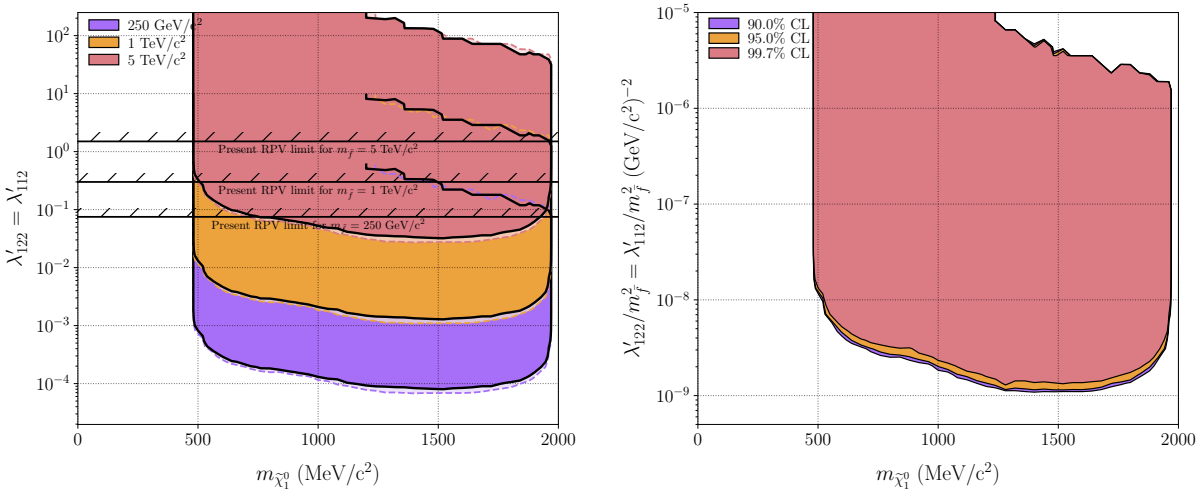


Figure 6.4: Sensitivity curves evaluated at 90% CL, corresponding to > 2.3 expected events, are shown for (top) Benchmark 1 and (bottom) Benchmark 2. In particular, the sensitivity in the plane of (left) $m_{\tilde{\chi}^0_1}$ against $\lambda'_P/m^2_{\tilde{f}} = \lambda'_D/m^2_{\tilde{f}}$, is shown. Including (excluding) decay channels with neutral final state particles results in the region denoted by the dashed (solid) contour. The sensitivity in the plane of (right) λ'_P against λ'_D is shown for three $m_{\tilde{\chi}^0_1}$ values: 600 MeV/c^2, 1200 MeV/c^2 and 1800 MeV/c^2 which are coloured in purple, orange and red respectively. Only charged decay channels are investigated here, to compare to previous studies. In all plots, current bounds for $m_{\tilde{f}} = 1$ TeV are indicated by hashed solid lines.

The uppermost accessible $m_{\tilde{\chi}^0_1}$ value in Benchmark 2 is higher than in Benchmark 1 owing to $m_{D_s} > m_{D^+}$. The lowermost accessible $m_{\tilde{\chi}^0_1}$ value depends on the masses of the lightest available final state. For charged final states in Benchmark 1, this is given by $m_{K^+} + m_{e^-}$.

At $m_{\widetilde{\chi}_1^0} \sim 900\,\mathrm{MeV/c^2}$ there is a kink in the sensitivity limit. This is due to the opening of decay channels to final states including K^* mesons. This reduces the lifetime of $\widetilde{\chi}_1^0$ and, for regions of low coupling, increases the probability of $\widetilde{\chi}_1^0$ to decay in the vacuum vessel.

Presented in Figs. 6.4(b) and 6.4(d) are sensitivities across a plane of λ'_P against λ'_D for three fixed values of $m_{\widetilde{\chi}_1^0}$. In accordance with previous studies, in this plane only charged final states are considered. Hence the shape of these contours are heavily affected by changes across the plane of the branching ratio to charged final states. For the particular examples of Benchmarks 1 and 2 neutralino decay is possible through both λ'_P and λ'_D. However, the final states accessible via λ'_P in these particular benchmarks are neutral. Therefore, although increasing λ'_P does increase neutralino production, it also reduces the neutralino lifetime and reduces the branching ratio to charged final states. These effects lead to the limit in this plane being bounded in both directions. Other benchmarks will show this is not always the case.



(a) (b)

Figure 6.5: (left) The effect of changing $m_{\tilde{f}}$ on the sensitivity of SHiP to λ' in Benchmark 2. As before the limits that only include charged final states are presented with solid lines, and the limit with all final states with dashed lines. (right) Variation of the sensitivity limit in Benchmark 2 including all final states when changing for different confidence limit requirements.

Current limits for $m_{\tilde{f}} = 1$ TeV, as provided in equations 6.3 and Ref. [68], are overlayed on the plots in Fig. 6.4. Some examples of the effect changing the assumption of $m_{\tilde{f}}$ has on the sensitivity of SHiP are presented in Fig. 6.5(a). To make the effect clear the sensitivity is

shown on an axis of λ', rather than $\lambda'/m_{\tilde{f}}^2$ as in other plots. Assuming lighter sfermions shifts the excluded region to lower values of λ', on an axis of $\lambda'/m_{\tilde{f}}^2$ the position of the region remains unchanged if $m_{\tilde{f}}$ is changed. The shape of the excluded region remains unchanged at SHiP as the changing $m_{\tilde{f}}$ affects both the production and decay widths together. Therefore, although lower values of λ' can be accessed with lighter sfermions, then sensitivity to higher values of λ' is lost as neutralinos are increasingly likely to decay before reaching the decay vessel. Figure 6.5(a) also provides the corresponding change to the current limits, these limits are not obtained from searches but from width measurements of mesons. These limits do have the same $m_{\tilde{f}}$ dependence, and it can be seen that for lower mass sfermions SHiP will extend the excluded region of phase space by multiple orders of magnitude.

Figure 6.5(b) provides some examples of how the sensitivity limit defined for Benchmark 2 moves for different choices of confidence limit. These are all computed with the assumption of zero background events. Therefore, as expressed before 90% CL corresponds to expecting > 2.3 events, then additional examples are provided for 95% CL and 99.7% CL (3σ discovery contour) which correspond to > 3.0 and > 5.8 events respectively.

6.2.2 Production from beauty mesons

The next benchmarks present sensitivities to $\widetilde{\chi}_1^0$ produced in the decays of beauty mesons. As the bottom quark is significantly heavier than the charm, these scenarios have a larger $m_{\widetilde{\chi}_1^0}$ reach.

Benchmark 3 is the first beauty production scenario explored. This benchmark has a non-zero coupling pair of (λ'_{131}, λ'_{112}) which opens neutralino production in the decays of B^0 mesons,

$$B^0 \;\rightarrow\; \widetilde{\chi}_1^0 \;+\; \nu_e. \tag{6.14}$$

The neutralino can then decay via the following channels,

$$\text{via } \lambda'_{112} \;:\quad \widetilde{\chi}_1^0 \rightarrow \begin{cases} K^{(*)\pm} + e^{\mp} & \text{(charged)}, \\ (K_L^0, K_S^0, K^*) + \nu_e & \text{(neutral)}, \end{cases} \tag{6.15}$$

and the sensitive mass region for charged final states is $(m_{K^\pm} + m_e) < m_{\widetilde{\chi}_1^0} < m_{B^0}$. The sensitivity contours in the plane of $m_{\widetilde{\chi}_1^0}$ and $\lambda'_P = \lambda'_D$ for this benchmark are presented in Fig. 6.6(a). The corresponding sensitivity in the plane λ'_P against λ'_D is shown in Fig. 6.6(b).

Benchmark 4 is an example of a benchmark with large regions of $m_{\widetilde{\chi}_1^0}$ only accessible via neutral final states. In this scenario the couplings λ'_{131} and λ'_{121} are non-zero and neutralinos are again produced in B^0 decays. The $\widetilde{\chi}_1^0$ can subsequently decay via λ'_{121} in the following channels,

$$\text{via } \lambda'_{121} \ : \quad \widetilde{\chi}_1^0 \rightarrow \begin{cases} D^{(*)\pm} + e^{\mp} & \text{(charged)}, \\ (K_L^0, K_S^0, K^*) + \nu_e & \text{(neutral)}. \end{cases} \tag{6.16}$$

For $m_{\widetilde{\chi}_1^0}$ values below $m_{D^\pm}$, sensitivity is only possible through neutral channels. This is clearly seen in the limits of this benchmark, presented in Figs. 6.6(c) and 6.6(d). A small kink in the charged sensitivity region is seen at ~ 2000 MeV/c^2, this is due to the difference in mass of $D^\pm$ and $D^{*\pm}$ and the neutralino decay channel $\widetilde{\chi}_1^0 \rightarrow D^{*\pm} + e^{\mp}$ opening.

Figures 6.7(a) and 6.7(b) present $m_{\widetilde{\chi}_1^0}$ vs λ' limits for two previously unexplored benchmarks. These include production from heavier beauty mesons. Benchmark 6 has (λ'_{132}, λ'_{112}) as a pair of non-zero coupling, which provides neutralino production from B_s^0 in

$$B_s^0 \ \rightarrow \ \widetilde{\chi}_1^0 \ + \ \nu_e, \tag{6.17}$$

and the same neutralino decays as in Benchmark 3. These are accessible via λ'_{112} as displayed in equation 6.15. Figure 6.7(a) displays the Benchmark 6 limit which can be compared to that of Benchmark 3 in Fig. 6.6(a). These benchmarks differ only by the meson involved in neutralino production. While the mass limits on the sensitivity are similar, the reach to lower couplings is reduced in Benchmark 6. This is predominately due to a smaller number of B_s mesons expected at SHiP compared to the number of B^0 mesons. This leads to a smaller number of neutralinos produced in Benchmark 6, see Fig. 6.3.

Benchmark 7, whose sensitivity limit is presented in Fig. 6.7(b), involves the coupling pair (λ'_{123}, λ'_{112}). The λ'_{123} coupling allows neutralino production from both B_s^0, as in equation 6.17, and $B_c^\pm$ mesons in

$$B_c^\pm \ \rightarrow \ \widetilde{\chi}_1^0 \ + \ e^\pm. \tag{6.18}$$

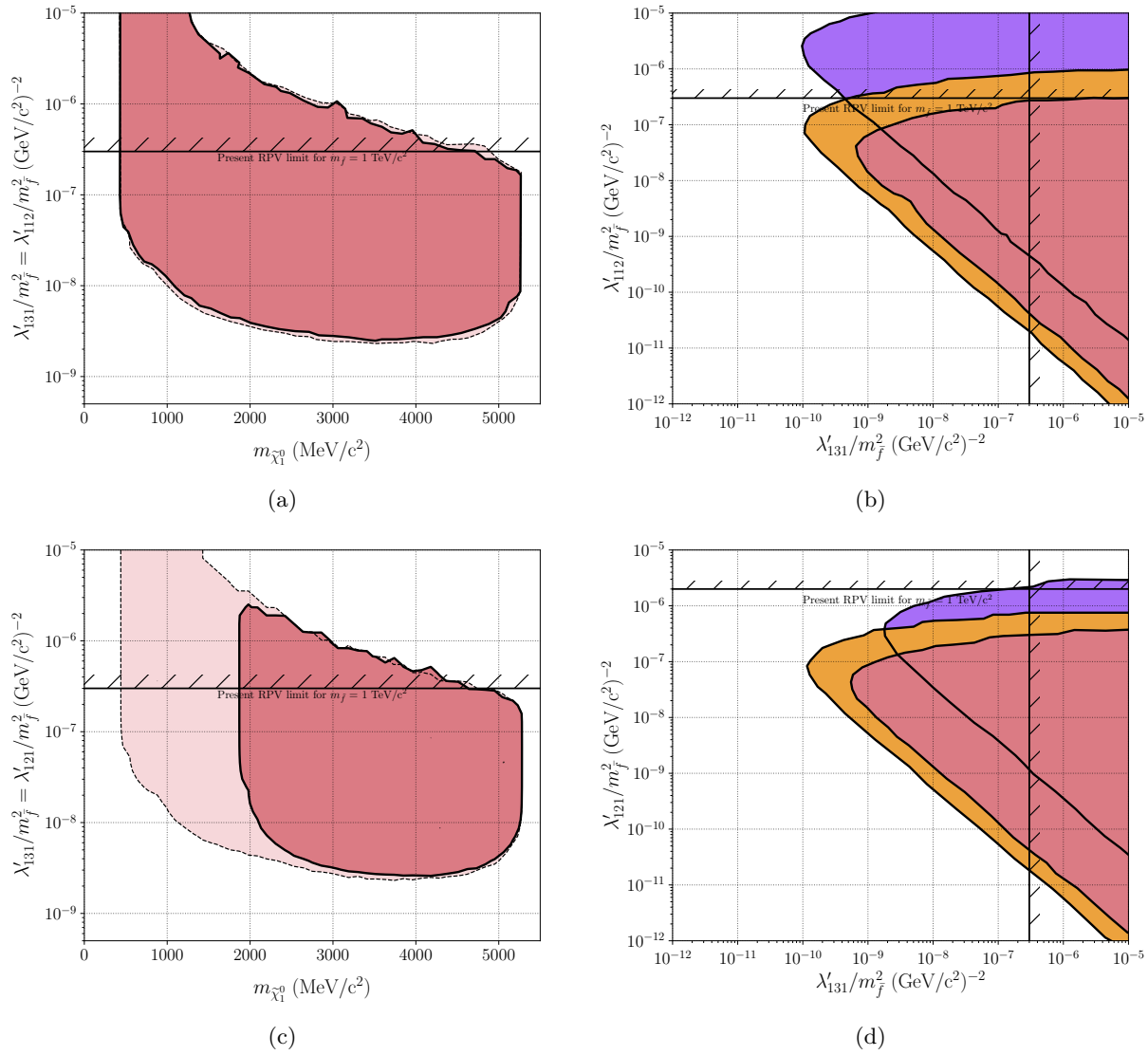


Figure 6.6: Sensitivity curves evaluated at 90% CL, corresponding to > 2.3 expected events, are shown for Benchmark 3 (top) and Benchmark 4 (bottom). In particular, the sensitivity in the plane of $m_{\tilde{\chi}^0_1}$ against $\lambda'_P/m^2_{\tilde{f}} = \lambda'_D/m^2_{\tilde{f}}$, is shown (left) including (excluding) decay channels with neutral final state particles results in the region denoted by the dashed (solid) contour. The sensitivity in the plane of λ'_P against λ'_D is shown (right) for three $m_{\tilde{\chi}^0_1}$ values: 1000 MeV/c^2, 3000 MeV/c^2 and 5000 MeV/c^2 for Benchmark 3 and 2000 MeV/c^2, 3500 MeV/c^2 and 5000 MeV/c^2 for Benchmark 4 which are each coloured in purple, orange and red respectively. Only charged decay channels are investigated here, to compare to previous studies. In all plots, current bounds for $m_{\tilde{f}} = 1$ TeV are indicated by hashed solid lines.

Neutralino decays proceed again via λ'_{112} as in equation 6.15. The production fraction of $B_c^{\pm}$ mesons is not known at the SHiP centre of mass energy. The value presented in Table 6.3 is

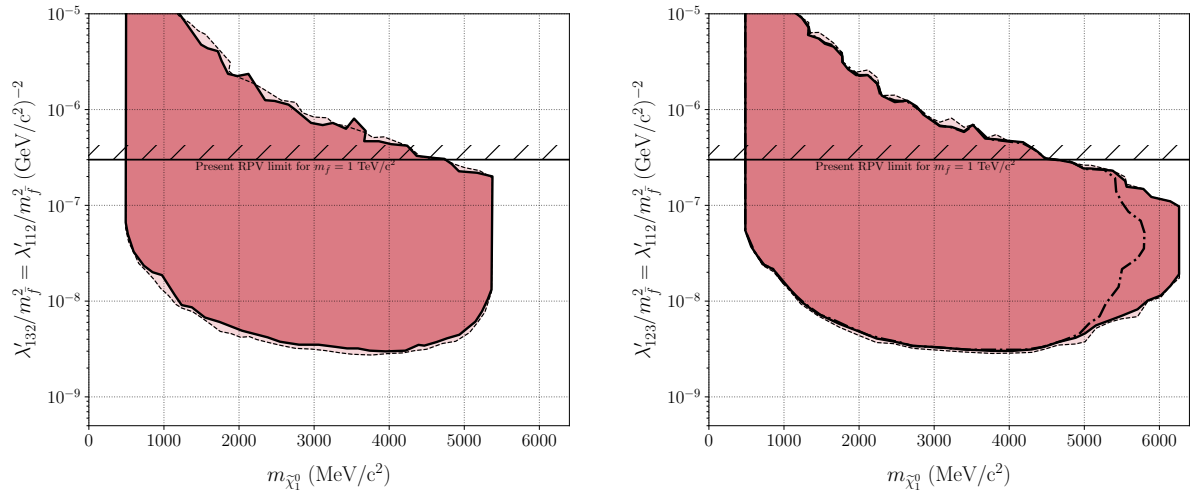
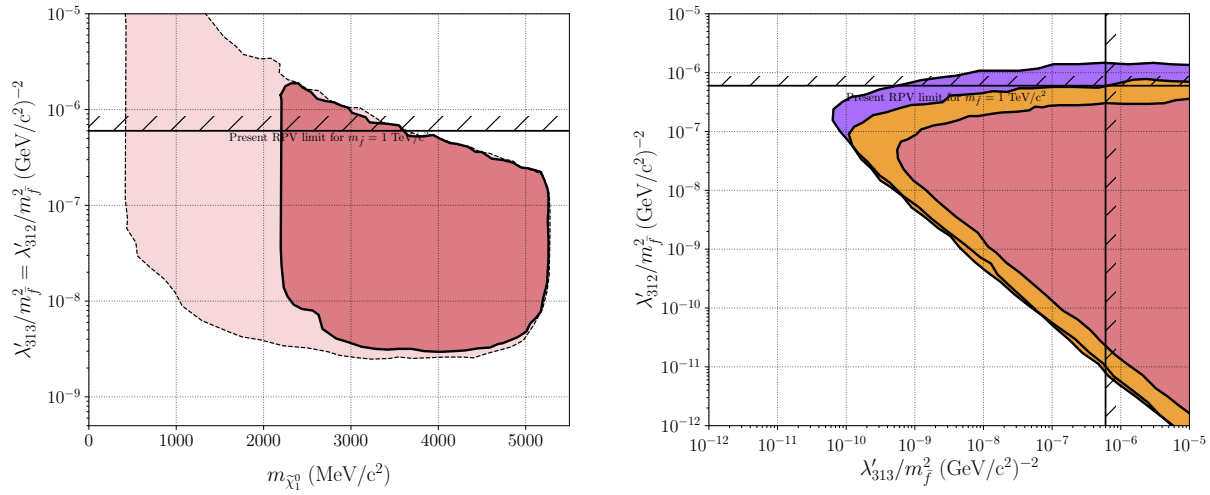


Figure 6.7: Sensitivity curves evaluated at 90% CL, corresponding to > 2.3 expected events, are shown for Benchmark 6 (left) and Benchmark 7 (right). In particular, the sensitivity in the plane of $m_{\tilde{\chi}^0_1}$ against $\lambda'_P/m^2_{\tilde{f}} = \lambda'_D/m^2_{\tilde{f}}$, is shown. Including (excluding) decay channels with neutral final state particles results in the region denoted by the dashed (solid) contour. For Benchmark 7 a second limit (dot-dashed line) is provided which uses a $B_c^{\pm}$ production fraction equal to 1% of the LHC value.

the value measured at the LHC [147, 148]. This study presents the limit produced using this value, however as the true value could be multiple orders of magnitude lower, a second limit is also provided in Fig. 6.7(b). This uses a production fraction equal to 1% of the LHC value, this was arbitrarily chosen as an example. The kinematics of produced $B_c^{\pm}$ mesons are also not known. Following the same procedure as Ref. [41], kinematics of $B_c^{\pm}$ mesons are generated based on those of $B^{\pm}$ mesons in the existing sample. The same angular distribution is used, and the energy distribution is re-scaled based on the heavier mass of the $B_c^{\pm}$ mesons as,

$$E_{B_c} = E_B \times \frac{m_{B_c}}{m_B}. \tag{6.19}$$

The decay constant used for the $B_c^{\pm}$ decay width calculation was taken from Ref. [149].



(a) (b)

Figure 6.8: Sensitivity curves evaluated at 90% CL, corresponding to > 2.3 expected events, are shown for Benchmark 5. In particular, the sensitivity in the plane of (left) $m_{\widetilde{\chi}^0_1}$ against $\lambda'_P/m^2_{\tilde{f}} = \lambda'_D/m^2_{\tilde{f}}$, is shown. Including (excluding) decay channels with neutral final state particles results in the region denoted by the dashed (solid) contour. The sensitivity in the plane of (right) λ'_P against λ'_D is shown for three $m_{\widetilde{\chi}^0_1}$ values: 2750 MeV/c^2, 3750 MeV/c^2 and 5000 MeV/c^2 which are each coloured in purple, orange and red respectively. Only charged decay channels are investigated here, to compare to previous studies. In all plots, current bounds for $m_{\tilde{f}} = 1$ TeV are indicated by hashed solid lines.

6.2.3 Coupling to 3$^{\text{rd}}$ generation leptons

Benchmark 5 investigates a scenario with non-zero couplings to third generation leptons. The production of $\widetilde{\chi}^0_1$ occurs via λ'_{313} in the following two processes:

$$B^0 \;\rightarrow\; \widetilde{\chi}^0_1 \;+\; \nu_\tau, \tag{6.20}$$

$$B^\pm \;\rightarrow\; \widetilde{\chi}^0_1 \;+\; \tau^\pm. \tag{6.21}$$

Neutralino decay then proceeds via λ'_{312} in the following processes,

$$\text{via } \lambda'_{312} \;:\quad \widetilde{\chi}^0_1 \rightarrow \begin{cases} K^{(*)\pm} + \tau^\mp & \text{(charged)}, \\ (K^0_L, K^0_S, K^*) + \nu_\tau & \text{(neutral)}. \end{cases} \tag{6.22}$$

Here, neutralino production from $B^\pm$ mesons is possible over a smaller region of $m_{\widetilde{\chi}^0_1}$ than that of B^0 mesons due to the requirement to also produce the high mass $\tau^\pm$. The branching ratio for

the production of $\widetilde{\chi}_1^0$ in B^0 decays is higher than from $B^\pm$ decays, and as a similar number of B^0 and $B^\pm$ mesons are expected at SHiP (see Table 6.3) the sensitivity curve is dominated by the contribution from B^0 mesons, see Fig. 6.3. The projected sensitivities of SHiP in Benchmark 5 are presented in Fig. 6.8.

There is an increase in the charged sensitivity limit at ~ 2600 MeV/c^2 with the opening of the channel to the second slightly heavier charged final state, $K^{*\pm} + \tau^\mp$. This opening makes less of an impact on the full sensitivity limit as more decay channels are already open at this point.

6.2.4 Fully leptonic neutralino decays

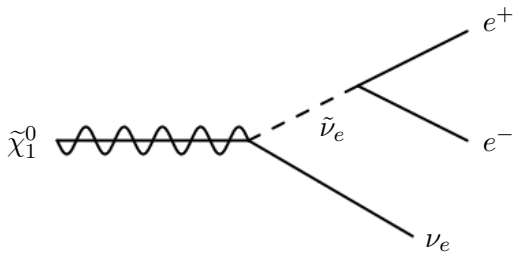


Figure 6.9: Diagram of $\widetilde{\chi}_1^0 \to e^+ + \nu_e + e^-$ [150].

Up to this point, each benchmark has only involved semileptonic λ' couplings. This section investigates two benchmarks in which the neutralino is allowed to decay to a fully leptonic final state via λ couplings, through $\lambda_{ijk} L_i L_j \bar{E}_k$ in equation 2.30. Specifically, these benchmarks involve the λ_{111} coupling, allowing neutralinos to decay via the following process

$$\widetilde{\chi}_1^0 \to e^+ + \nu_e + e^-. \tag{6.23}$$

A diagram of this process is provided in Fig. 6.9. The width of this process, assuming the final state leptons are massless, is presented in Ref. [146]. For this study this width is modified to account for the lepton masses using an adaptation applied to a similar HS decay in Ref. [151].

Benchmark 8 pairs the λ'_{122} production coupling with λ_{111}. Therefore, neutralino production can occur in decays of $D_s^\pm$ mesons, as in Benchmark 2. The sensitivity of the SHiP experiment in Benchmark 8 is presented in Fig. 6.10(a). The sensitivity limit does not extend to couplings as low as in Benchmark 2. However, the limit does reach much higher coupling values. This is

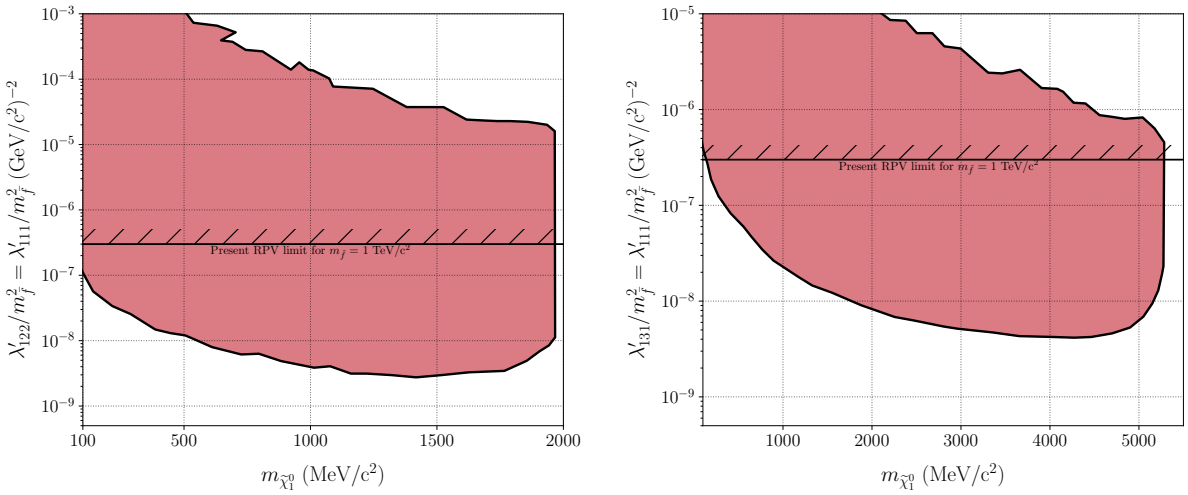


Figure 6.10: Sensitivity curves of SHiP to neutralinos produced via semi-leptonic couplings which and which decay via fully leptonic couplings evaluated at 90% CL, corresponding to > 2.3 expected events, are shown for (left) Benchmark 8 and (right) Benchmark 9. In particular, the sensitivity in the plane of $m_{\tilde{\chi}^0_1}$ against $\lambda'_P/m^2_{\tilde{f}} = \lambda'_D/m^2_{\tilde{f}}$, is shown.

especially clear at higher masses where sensitivity was lost in Benchmark 2 due to neutralinos decaying before the decay vessel. Larger couplings can be reached as the decay width of the neutralino into this fully leptonic final state is lower than the semi-leptonic one, for the same values of coupling and mass. The reach into lower $m_{\tilde{\chi}^0_1}$ is also extended as it is no longer limited by the necessity to produce a meson in the final state.

Similarly, Benchmark 9 which uses the λ'_{131} coupling paired with λ_{111}, is presented in Fig. 6.10(b). This benchmark has neutralino production from B^0 mesons, as in Benchmark 3, see equation 6.14. The same behaviour is observed as in Benchmark 8.

6.2.5 K^0_L mass cuts

The potential for V^0 particles to be produced, enter the decay vessel undetected and then go on to decay to signal topologies is a dangerous background at SHiP. The most dangerous component is from K^0_L particles, some relevant K^0_L decay channels are as follows,

$$K^0_L \to \pi^+ + e^- + \nu_e, \qquad K^0_L \to \pi^+ + \pi^- + \pi^0, \qquad K^0_L \to \pi^+ + \pi^-. \tag{6.24}$$

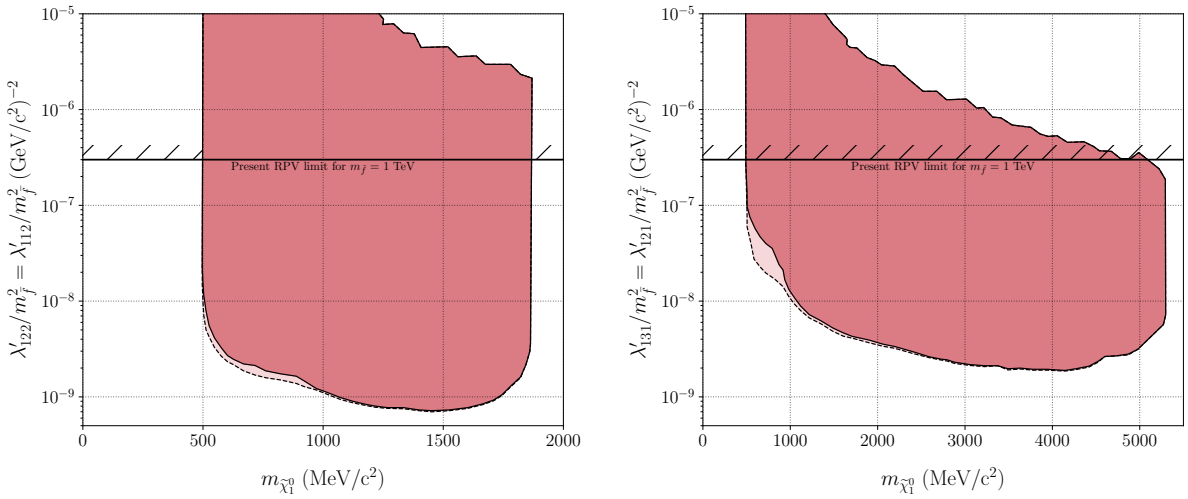
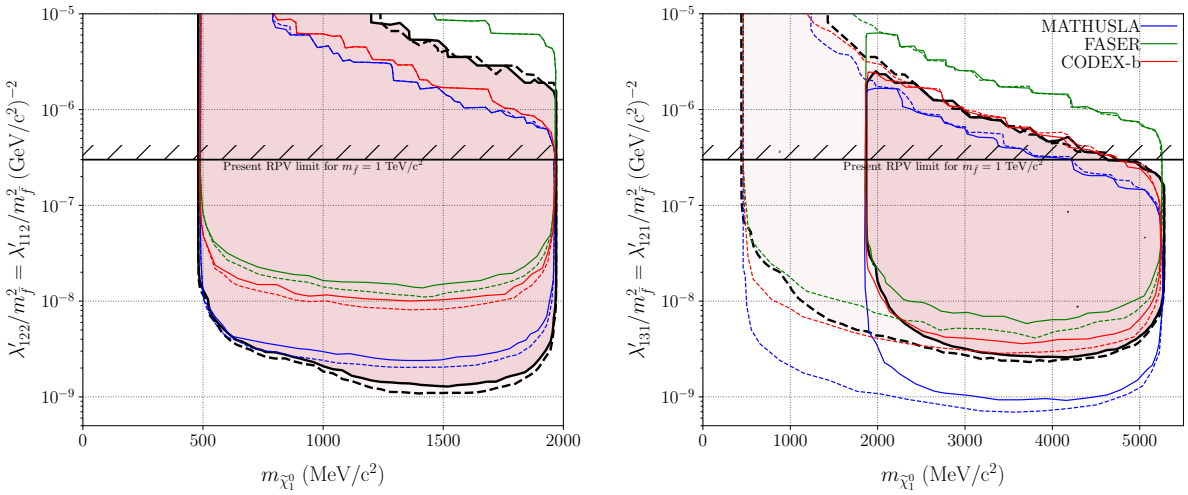


Figure 6.11: Impact of $m_{K_L^0}$ cuts on sensitivities to Benchmark 1 (left) and Benchmark 4 (right). The full sensitivity, including all channels is shown with the dashed line, then the sensitivity after the $m_{K_L^0}$ cuts is presented with a solid line.

Background induced from K_L^0 particles produced in muon or neutrino inelastic scattering events in the decay vessel, the cavern walls and the SND have been analysed and shown by the collaboration to be reducible with the cuts provided in Chapter 5.2.2. Studies have shown the background can be suppressed to $\mathcal{O}(10^{-4})$ events over the 5 year run period [18]. For V^0 particles produced in material away from the beamline, the cavern walls for example, reconstructed signal candidates do not point back to the target and are removed with the IP cut. Background events from DIS interactions producing V^0 particles closer to the beamline are vetoed by either detecting the incoming muon, for DIS events in the SBT itself, or by detecting the other products of the upstream DIS event in the SBT or UBT.

However, if a signal event does occur during the running of SHiP that passes selection and reconstructs to the K_L^0 mass there will still be suspicion. Some example sensitivity estimates have been produced with a cut introduced around the K_L^0 mass for Benchmark 1 and Benchmark 4. These are provided in Fig. 6.11. For this example any final state reconstructing back to a signal candidate with mass $(m_{K_L^0} - 25\text{MeV/c}^2) < m_{\tilde{\chi}_1^0} < (m_{K_L^0} + 25\text{MeV/c}^2)$ is cut. This window was chosen as an example, it is easily wide enough to remove the peak in a reconstructed mass distribution of a signal sample simulated with $m_{\tilde{\chi}_1^0} = m_{K_L^0}$. Here this cut maximally includes the



(a) (b)

Figure 6.12: Sensitivities of LLP experiments from Ref. [62], overlayed on SHiP estimates from Fig.6.4 and Fig.6.6. Presented in (a) is Benchmark 2, and (b) is Benchmark 4. As before solid and dashed lines refer to charged final states only, and all final states respectively.

effect of miss-identification of any meson of the final state. For each meson in a final state both π and K hypotheses are tested, if either results in a reconstructed mass close to $m_{K^0_L}$ the event is vetoed. The effect of this cut is minimal across the majority of the accessible $m_{\tilde{\chi}^0_1}$ region.

In Benchmark 1 the charged final state $K^{\pm} + e^{\mp}$ is accessible across the full $m_{\tilde{\chi}^0_1}$ region. In the region of $(m_{K^0_L} - 25\text{MeV/c}^2) < m_{\tilde{\chi}^0_1} < (m_{K^0_L} + 25\text{MeV/c}^2)$ around 99% of sensitivity is lost. The events passing the cuts in this region are those reconstructed to the incorrect mass. For example the decay chain $\tilde{\chi}^0_1 \rightarrow K^0_L + \nu_e$ and $K^0_L \rightarrow \pi^+ + \pi^0 + \pi^-$ where the π^0 is missed but the reconstructed signal candidate still points back to the target. Outside of this region and before K^* decay channels open the sensitivity lost is around 50%. After K^* channels open the effect diminishes.

For Benchmark 4, the lower region $m_{\tilde{\chi}^0_1}$ is only accessible to neutral final states. The lowest mass charged final state is $D^{\pm} + e^{\mp}$. The impact of the cut is larger. The fraction of sensitivity lost is $\sim 99\%$ at low $m_{\tilde{\chi}^0_1}$ until K^* decay channels open.

6.3 Comparison to other facilities

A number of other facilities are being designed in order to search for long lived particles (LLPs), similar new physics signals to that of SHiP. The most relevant rival efforts for the search for RPV neutralinos are the MATHUSLA, CODEX-b and FASER experiments[1]. These detectors are parasitic of the LHC (and the future HL-LHC), each placed close to an interaction point.

RPV neutralino signals at these facilities, as at SHiP, come from B and D meson decays. For a given RPV scenario the sensitivity of each facility is predominately determined by: the size of its fiducial volume, the distance between LLP production point and the facility, and the location relative to the beam axis. FASER for example lies close to the LHC beam axis taking advantage of D and B meson boosting in the forward direction. This makes the proposal very cost efficient, FASER is a small experiment but attains a relatively competitive sensitivity.

A comparison of facility sensitivities is provided for Benchmark 2 and 4 in Fig. 6.12 using estimates from Ref. [62]. MATHUSLA appears the most competitive with SHiP. The SHiP facility excels in sensitivity to $\widetilde{\chi}_1^0$ produced in charm decays. This is due to SHiP operating with an extremely intense beam, albeit at a lower centre of mass energy. The LHC, on the other hand, provides MATHUSLA with unparalleled sensitivity to $\widetilde{\chi}_1^0$ production from beauty mesons. As presented already, the SHiP sensitivity estimate in Fig. 6.12 includes the selection cuts verified by comprehensive background studies [18]. Limited studies exist for the MATHUSLA experiment, therefore this sensitivity estimate assumes no background and full signal efficiency.

While the discovery sensitivity of the SHiP facility is surpassed the HS detector of SHiP can provide additional information about each signal candidate. The spectrometer of the SHiP facility can provide a determination of whether a HS candidate originated from the target without any assumptions, very useful for background discrimination. Additionally, the spectrometer can provide a mass measurement of the signal candidate, and finally the calorimeter and muon systems of the SHiP experiment can provide some HS model discrimination.

[1] ANUBIS is a more recent and less developed proposal for a similar facility in the ATLAS/CMS cavern [152], it provides similar discovery potential to MATHUSLA [139].

6.4 Summary

This chapter has presented updated sensitivity estimates for the SHiP experiment to R-parity violating neutralinos produced in heavy-flavour meson-decays. This study has made use of detailed simulations of the newly optimised SHiP facility. It also accounts for cascade enhancements of heavy meson production in the SHiP target. In addition, the effects of the detector acceptance, reconstruction algorithms and selection criteria are taken into account. Therefore, these results supersede the previous sensitivity estimates of the SHiP experiment to R-parity violating neutralinos [67].

Multiple benchmark scenarios are presented, providing a comprehensive view of the sensitivity to various possible patterns of non-zero RPV couplings. Benchmarks are chosen to cover the full range of $\widetilde{\chi}_1^0$ masses that will be probed at the SHiP facility.

Finally, this study presents for the first time sensitivities for benchmark scenarios in which neutralinos are produced in decays of B_s and B_c mesons, as well as scenarios where neutralinos decay purely leptonically.

Chapter 7

Benchmarking Graphcore® IPUs for event generation

In the search for new physics, upgrades to existing experiments and new facilities are designed to operate at higher intensities and/or higher energies than previous searches. The HL-LHC will operate with a data rate exceeding $\mathcal{O}(10)$ TB/s [88]. As a comparison, the LHC has a typical data rate of $\mathcal{O}(1)$ TB/s. Data rates and simulation load will increase in parallel and the handling of both will be a huge computational challenge. These issues were introduced in Chapter 4 where a GAN based solution was presented for a particular simulation bottleneck at the SHiP experiment. GANs are an exotic solution which currently, at best, have an extremely limited scope. There exists no generalisation and each GAN based solution must be tailor-made. Optimisation efforts of traditional event generation software, including more efficient parallelisation, is ongoing. However, forward projections still forecast a significant short-fall over the next decade [89–92].

New hardware architectures are also being investigated to help cope with new computational challenges, these include using combinations of CPUs, GPUs, and occasionally field-programmable gate arrays (FPGAs) [153–157]. This chapter details the first particle physics focused investigation into the performance of Graphcore's Intelligence Processing Unit (IPU). The IPU is a chip developed specifically for machine learning applications. Benchmarking results published by

Graphcore themselves show some impressive capabilities of the IPU and IPUs were shown to outperform GPU for a range of applications [158]. The IPU is of specific interest here due to the rising prevalence of neural network focused machine learning within the particle physics community [159]. Catalysed by hardware improvements and the development of increasingly user-friendly programming environments such as `TensorFlow`, machine learning has become an accessible tool to particle physicists.

This was a preliminary study investigating the potential of this hardware, more detailed studies are expected to follow. The content of this chapter is focused solely on GANs and is based on Ref. [93], from which Sections 3.1 and 3.3 are my own work. Within Ref. [93] there are other studies detailing the performance of the IPUs for other particle physics related tasks.

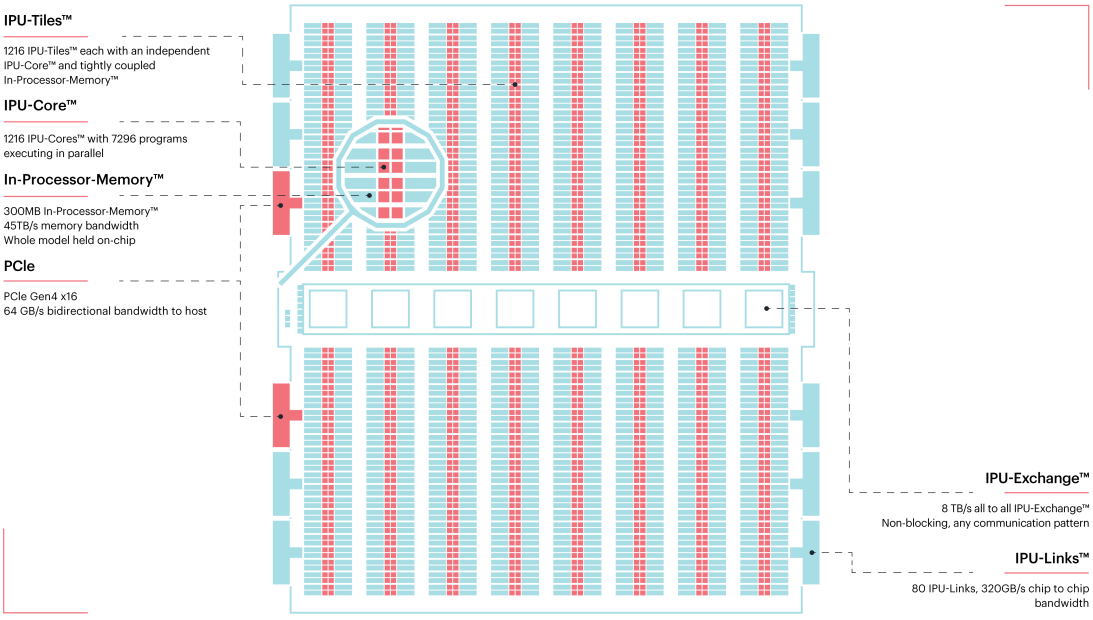


Figure 7.1: The Graphcore Colossus™ MK1 GC2 IPU [160].

7.1 IPUs

This study makes use of Graphcore's first generation Colossus™ MK1 GC2 IPU (see Fig. 7.1). The IPU has an architecture fundamentally different to that of a CPU or GPU. The IPU, like

the GPU, is highly-parallelised. The IPU comprises of 1,216 processing elements, called tiles. The IPU operates with a multiple instruction multiple data (MIMD) architecture allowing the independent running of each processing element. This is in contrast to the single instruction multiple data architecture of the GPU. Each tile consists of a computing core and 256 KiB of local memory. In total 7,296 threads can be executed in parallel in a single IPU. Tiles are connected together with an on-chip interconnect, the IPU exchange™, allowing for low-latency and high-bandwidth communication of up to 7.7 Tb/s. Each IPU card consists of two such IPUs. The IPUs are connected to each other via 80 IPU links™ reaching a total chip-to-chip bandwidth of 2.5 Tb/s, and are connected to the host via 16 PCIe Gen4 links (8 per IPU).

To accompany the IPU hardware, Graphcore provides some software frameworks. This investigation was completed with their Software Development Kit (SDK) version v1.2.0. This SDK includes Graphcore's C++ based Poplar software, which provides the most granular control of the IPU and its tiles. Also included are IPU compatible implementations of `PyTorch` [161] and `TensorFlow`, some of whose functions have been re-written and optimised specifically for the IPU architecture. It is the `TensorFlow` implementation that the results presented in this Chapter rely on.

7.2 Benchmarking

The majority of generative network approaches developed for particle physics applications are based on GANs. Moreover, the majority of these are GANs rely on convolutional layers. For many particle physics applications correlations between neighbouring data points are important, for example in a spatial distribution of energy deposited in a calorimeter. Convolutional layers can manage these problems more efficiently than simple dense layers, focusing on relationships between features close to each other in the feature space. This efficiency is especially relevant in high dimensional applications where network sizes would otherwise quickly grow leading to slow training and slow inference.

For this investigation two convolutions networks are taken from particle physics literature, the small convolutional DijetGAN from Ref. [102] and the larger locally-connected LAGAN

(location aware GAN) from Ref. [82]. Additionally investigated are both the prompt and non-prompt fully-connected GANs from Chapter 4. As discussed, both networks are of similar architecture, however the prompt network is narrower and has significantly less parameters. Similarly to the GANs developed for SHiP, the DijetGAN directly models a distribution of a handful of kinematical observables, in this case these are features of jets at the LHC. The LAGAN application models the distributions of energy deposits from jets in calorimeters. The LAGAN relies on locally-connected layers. These operate very similarly to convolutional layers however, where traditional convolutional layers each create a set of filters that are applied over the whole image the locally-connected layer has personalised filters for each region. Locally connected layers are not often applied to applications involving natural images due to their lack of flexibility to handle feature translation and their very low parameter efficiency [82]. For the LAGAN application the jet images are transformed such that in each training example the jet is aligned in the centre of the image. A specific filter for each region relative to the common centre is then an attractive option, providing more flexibility to model an energy deposition whose characteristics may change across each image.

For the following benchmark tests the networks are constructed in `TensorFlow` using information from within each publication or in code made available by the authors. The weights of the networks are initialised with random noise. This choice was made as the trained weights were not always made available. This will not affect the benchmarking results.

The scripts written to generate data on the IPU are different to those of the GPU and CPU by requirement, though to ensure these benchmarking comparisons were as fair as possible each approach was individually optimised. The IPU code must use Graphcore specific functions to compile the generator graph, loop over this graph and collate output samples. The GPU and CPU code made use of `tf.function` and `tf.while_loop` which allowed the model to be queried with `TensorFlow` tensors. This was observed to be significantly faster and more similar to the IPU approach than simply querying the `Keras` model with the `predict` function which relies on `Numpy` arrays. This is a significant speed up over the benchmarking results of Chapter 4 which were carried out with the inefficient `Keras predict` function. For all benchmarking tests, warm up runs were completed before any generation steps were timed. These warm up runs contain

	Name	Cores	Memory	Clock Speed	TDP
CPU	Intel Xeon Platinum 8168	24	732 GiB	2.7 – 3.7 GHz	205 W

	Name	Cores	Memory	32 bit FLOPS	TDP
GPU	Nvidia Tesla P100	3584	16000 MiB	9.3 TFLOPS	250 W
IPU	Graphcore Colossus™ GC2	1216	286 MiB	31.1 TFLOPS	120 W

Table 7.1: Key specifications of the processors used for this benchmarking as provided on manufacturer websites [130, 160, 162], and in [163, 164]. Performance in terms of floating point operations per second (FLOPS) is given for 32 bit single-precision operations. Thermal design power (TDP) is given for each processor, where for the IPU this is half of the total board TDP.

any compilation and graph creation overheads, which notably were much more significant for the IPU. On the IPU some models take $\mathcal{O}(5)$mins to compile. The `TensorFlow` profiler was used to check that the GPU code was indeed limited by arithmetic operation throughput and not by any other bottleneck. This confirms the results obtained are fair representations of the power of the GPU. Such sophisticated profiling tools are not as easily available on the IPU.

The hardware options used in this study and their key specifications are laid out in Table 7.1. Note that the CPU option is a very high end CPU and that the GPU option is chosen as it is of a comparable price to the IPU. Also notable is the significant power saving the IPU offers over the GPU[1].

Each hardware option was tested over a range of batch sizes, where the batch size is the number of samples generated by the network in a single tensor passing. The time taken to generate a large number of batches was then measured. These results are presented in Fig. 7.2. The relationship between inference speed and batch size is consistent for each hardware choice, where larger batch sizes show increased computational efficiency. There is a maximum accessible batch size to each hardware, this is smallest for the IPU which is the hardware choice with the smallest memory capacity. The efficiency gains obtained by increasing batch size begin to diminish towards the higher batch sizes, this could be due to any overheads existing at lower batch sizes becoming negligible or more likely the rise of some new overheads in dealing with large tensors. From these results the inference speeds at observed optimal batch size for each hardware are extracted and presented in Table 7.2. Ratios of the inference speed on the IPU vs

[1] An organisation's environmental impacts are increasingly in the public eye, energy efficiency is an evermore important issue. See Ref. [165].

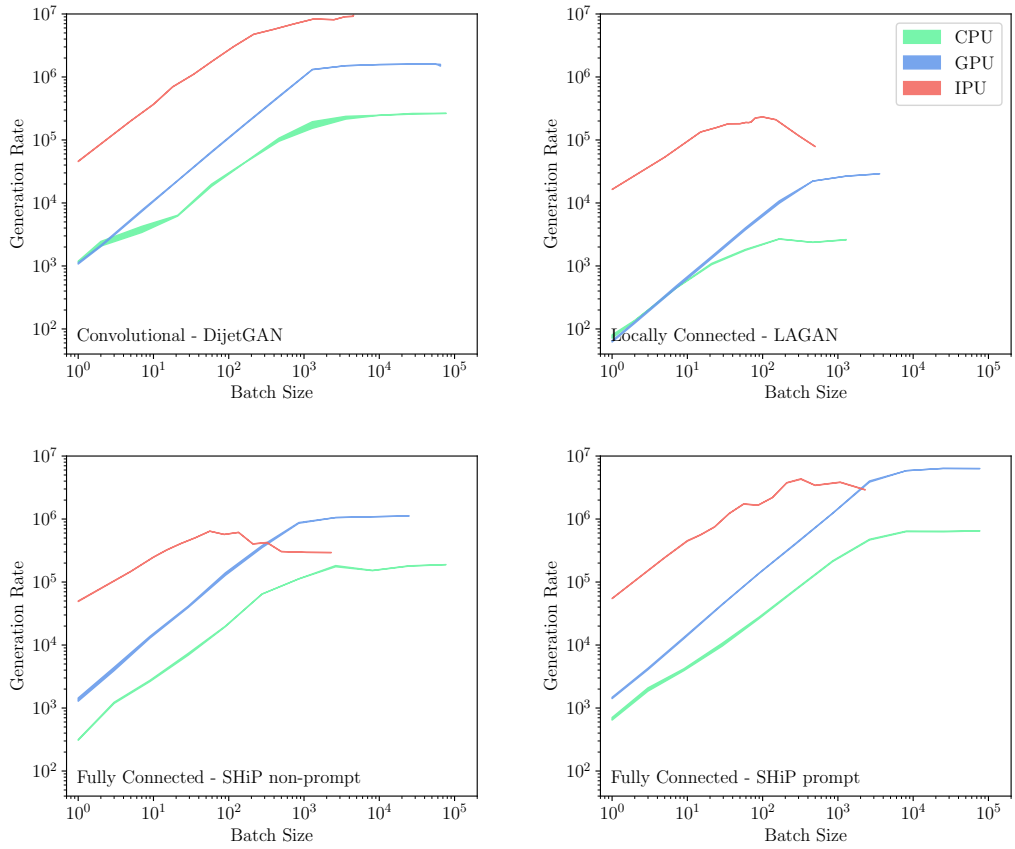


Figure 7.2: Benchmarking results of the event-generation rate as a function of the batch size of the network. Results are presented for IPU, GPU and CPU hardware options outlined in

Network Name	Number of Parameters	IPU/CPU rate	IPU/GPU rate
DijetGAN	3×10^4	36.3	6.0
LAGAN	4×10^6	86.5	8.0
SHiP non-prompt	5×10^6	3.4	0.6 ($\times$1.7 slower)
SHiP prompt	6×10^5	6.7	0.7 ($\times$1.4 slower)

Table 7.2: Benchmarking results calculated using optimal batch size for each hardware option.

the speed on the GPU and CPU are presented. The IPU is consistently faster than the CPU, as expected, however the IPU also outperformed the GPU in generating from the convolutional networks by almost an order of magnitude. The IPU was marginally slower than the GPU in generating from the fully-connected networks tested. This is due to the maximum accessible batch size of the IPU. Therefore the IPUs do not offer any improvement in generation speed from the SHiP muon background GANs presented in Chapter 4 due to memory limitations on batch size. The IPU has significant memory constraints. The 256 KiB of local memory per core equates

to only 318 MiB across the whole IPU. Even this is generous, as the full memory is only accessible to an operation that can perfectly spread out over all the tiles. Jobs that fail due to memory issues drop because a single tile was overloaded. This shortcoming is addressed in the design of the high-bandwidth and low-latency IPU links. Graphcore have software implementations that can manage the *sharding* of networks across multiple IPUs allowing for larger networks or larger batch sizes to be managed. This software includes data *pipelining* which organises queues of data to ensure each element of the sharded network has minimal downtime. For the purposes of these tests neither sharding or pipelining was employed as all generator networks tested fit on a single IPU. Splitting the network up and spreading over multiple IPUs may allow larger batch sizes to be used. This may increase inference speed but this increase would only be relevant if the gain was greater than that of two IPUs running the same network separately. This is not investigated here. It is worth noting in this context that the 2$^{\text{nd}}$ generation IPU recently released has triple the memory per tile compared to the 1$^{\text{st}}$ generation IPU used here[2].

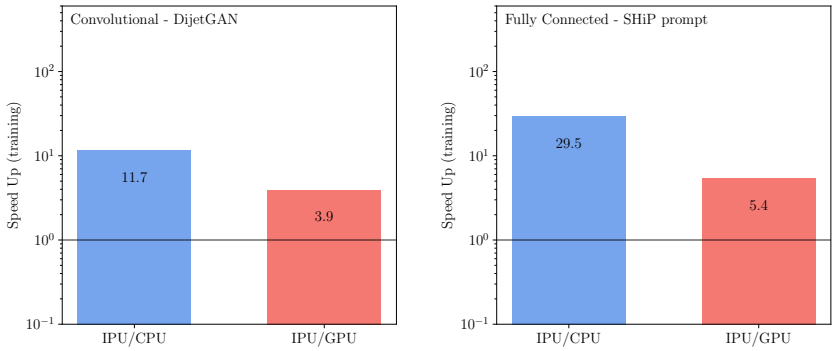


Figure 7.3: Comparison of the training speeds of the IPU relative to the CPU and the GPU of Table 7.1.

Notable from the inference benchmarking results is that the IPU, regardless of model, was significantly faster than the GPU at low batch sizes. As described in Chapter 4, lower batch sizes are often chosen for training GANs in an approach referred to as mini-batch gradient descent. Commonly the batch size chosen for training a GAN is $\mathcal{O}(50)$. This is a region that the IPU tends to dominate the GPU. Figure 7.3 presents the training speed up achieved whilst using the IPU for the applications of the DijetGAN and the SHiP prompt GAN using the batch size

[2]Initial benchmarking for this 2$^{\text{nd}}$ generation IPU were recently released showing significant improvements over the 1$^{\text{st}}$ generation IPU [166].

presented in their respective publications. The LAGAN and the SHiP non-prompt options could not be tested here as the full GAN models ($\hat{G}$ plus $\hat{D}$) were too big to fit on a single IPU. Whilst the sharding approach works well for a single network allowing for a large model to be split up across multiple IPUs, it is not yet possible for a GAN model. The GAN case is complicated by the continual interactions between models. This may be possible in the future. Figure 7.3 shows both networks training significantly faster on the IPU as expected from the inference results where for lower batch sizes the IPU consistently outperforms the GPU. Faster training means faster prototyping and which will lead to improved network performance. A wider range of architectures could have been tested for the application in Chapter 4 in the same amount of time. Then for tasks such a exploring the parameter space of a chosen architecture using the IPU would allow significantly faster testing and therefore a more detailed optimisation which would lead to improved performance of the algorithms in Chapter 4.

7.3 Outlook

This chapter has presented the results from some preliminary testing of the Graphcore IPUs. A small selection of models were tried, chosen to be roughly representative of the kind of GANs within the HEP literature. While IPUs may not be a golden bullet, for the application of GANs these tests conclude IPU performance ranges from keeping up with GPU competition to significantly outperforming based on application. The IPUs are competitively priced and their cost is similar to high end GPUs. Memory considerations currently limit the size of GAN models the IPU can train. For inference the network size is essential not limited, as the IPUs are designed with high-bandwidth and low latency connections allowing multiple IPUs to be linked. Large network can then be efficiently sharded across multiple IPUs. A more extensive analysis of IPU performance for a wider variety of architectures and network sizes may be useful to understand where the gains can really be made with the IPU, but for now preliminary results are promising.

Chapter 8

Enhanced generative networks

This chapter introduces developments which improve on the performance of the GAN approach to background muon generation presented in Chapter 4. This improvement comes from teaching the generator network to understand an input parameter that describes the local density of each sample within the full training sample. After training, the distribution of this parameter can be adapted to control the characteristics of generated samples. This is a first attempt at a modification of this style to the standard GAN approach. Potential applications for this improved GAN and this newfound flexibility within a new optimisation of the SHiP muon shield are outlined.

8.1 Motivation

The GAN architecture developed in Chapter 4 is extremely simple. The generator network, $\hat{G}$, simply maps latent noise to a muon kinematic vector, and in training the discriminator, $\hat{D}$, network maps a kinematic vector into a single decision on sample origin. From here this architecture will be referred to as a *vanilla* GAN. The vanilla architecture is inflexible. Once

trained, the output result cannot be altered and depends solely on the training sample used. As observed in Chapter 4 these GANs struggle to accurately model both the characteristics of and the number of events in the tails of distributions. For example, significant miss-modellings are visible in Fig 4.7. The tails are harder to model simply as these events are encountered less often during the training procedure, these regions are more poorly described by the training sample. Both miss-modelling and underestimating the tails is a significant drawback of these GANs, especially when modelling physical distributions the tails of which are often important regions.

An initial idea to address this shortcoming of the vanilla architecture was to *dope* the tails of the training data set. That is, to artificially enhance the tail regions. This could be achieved either by manually adding fake events to the training sample or by weighting up existing events. Clearly, this is an unsatisfactory approach as it requires arbitrary decisions to be made on how the tails are enhanced. For example, how many additional events should be added and where should they go? These decisions can only be optimised by comparing results from multiple lengthy training runs, which can each take days with the vanilla GANs of Chapter 4. An ideal solution would provide control of the generated output after training with the ability to generate more events in the tails if required.

Conditional GANs have additional label inputs provided to the generator which inform the network of certain characteristics the generated output should have [167–169]. These labels could be binary class labels for example. The discriminator of the conditional GAN makes each decision having seen both the sample vector along with its corresponding class labels. The discriminator is then able to decipher the meaning of these labels. This understanding is then subsequently transferred to the generator through the standard GAN training procedure. An auxiliary classifier GAN is a type of conditional GAN where instead of receiving the conditional label as an input, the discriminator is tasked with reconstructing the label as an auxiliary task [170–172].

For this application, muons in the training sample could be split into two classes, the tails and the core. While this would install flexibility in the trained generator model, where more events can be asked for from the tail class during generation if required, this approach would still depend on arbitrary decisions on how to define of tail and core classes. This can therefore be

improved by making the auxiliary task a regression one, where the label is a continuous variable.

8.2 Auxiliary GANs

The architecture of the auxiliary GAN is again made up of a generator $\hat{G}$ and a discriminator network $\hat{D}$. The inputs to the generator network are: a vector of latent noise, as before, and an auxiliary parameter. As before the output of $\hat{G}$ is a synthetic sample. In the auxiliary GAN the primary task of $\hat{D}$ is still to decide what the origin of each sample is, however, $\hat{D}$ is also tasked with predicting the auxiliary parameter associated with each sample. The total loss for $\hat{G}$ then is,

$$L_g = \lambda \times L_{vanilla} + L_{auxiliary}. \tag{8.1}$$

where $L_{vanilla}$ is the vanilla GAN loss, as used in Chapter 4:

$$L_{vanilla} = \frac{1}{n} \sum_{i=1}^{n} -\log(\hat{D}(x^i_{\text{gen}}; \theta_d)_0), \tag{8.2}$$

where $\hat{D}(x^i_{\text{gen}}; \theta_d)_0$ is the vanilla discriminator output. In equation 8.1, $L_{auxiliary}$ is the auxiliary loss, which is just the mean squared error (MSE),

$$L_{auxiliary} = \frac{1}{n} \sum_{i=1}^{n} (\hat{D}(x^i_{\text{gen}}; \theta_d)_1 - A_i)^2, \tag{8.3}$$

where A_i is the auxiliary value that was used to generate the sample, and $\hat{D}(x^i_{\text{gen}}; \theta_d)_1$ is the auxiliary value prediction of $\hat{D}$. The importance of the vanilla loss is weighted relative to the auxiliary by a factor λ. As before the discriminator is trained with a binary crossentropy loss, as in equation 4.2, and a mean squared error term, similar to that of equation 8.3. These are combined in a weighted sum similar to that of equation 8.1.

8.3 Training setup

For this investigation the training sample introduced in Chapter 4 is used. However, it is modified in the following ways: the sample size is reduced to just 1×10^7 muons for simplicity in this preliminary investigation, the weights associated with the enhancement of rare muon production channels and muon charges are discarded also for simplicity, and the x- and y- coordinates are

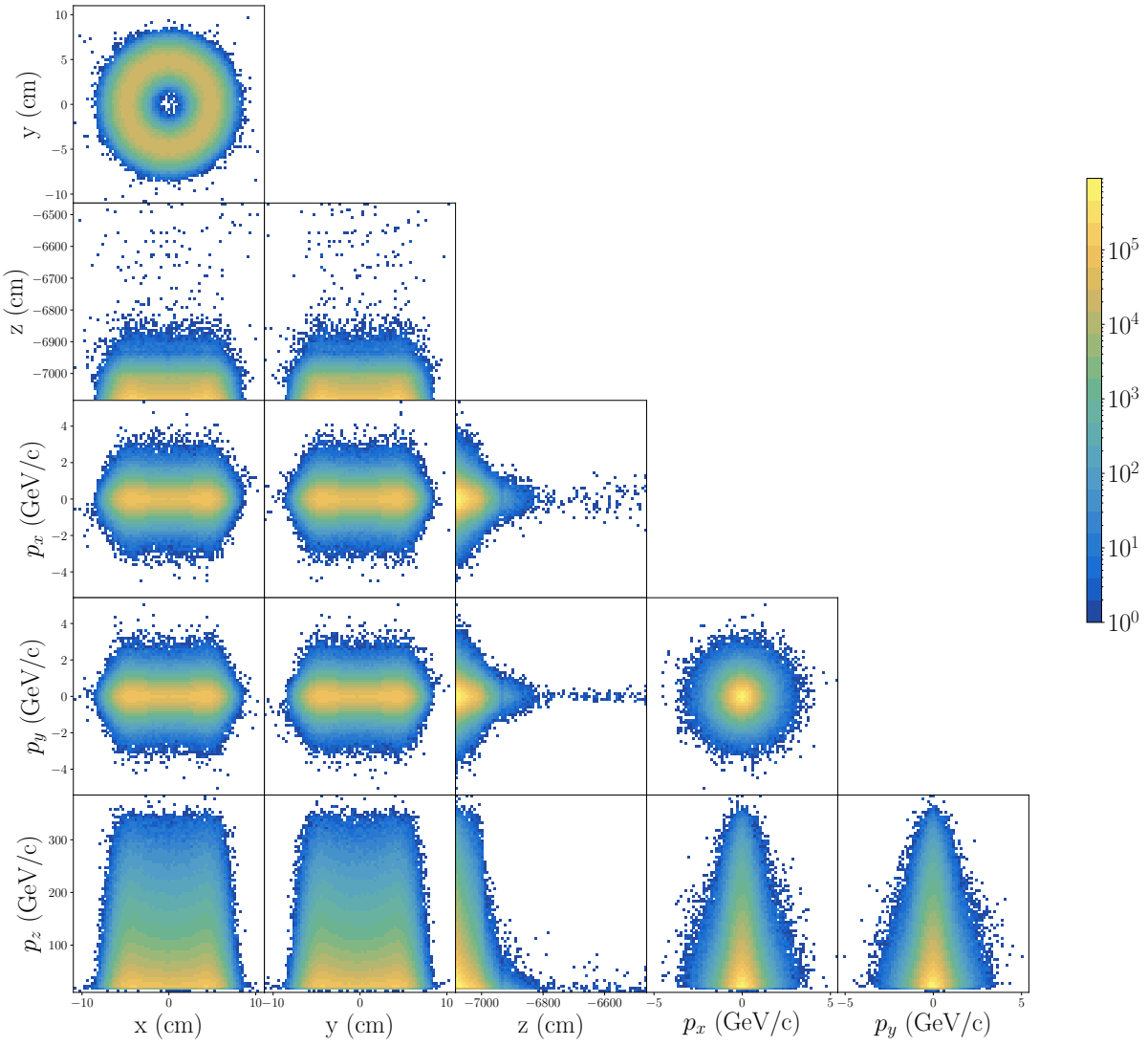


Figure 8.1: Training sample of muon kinematics, smeared in the x-, y- plane.

smeared in the fashion described in Chapter 4.5. The addition of this beam smearing is required for the applications described later in Section 8.4.2. The new training sample is presented in Fig. 8.1.

A simple improvement made over the approach of Chapter 4 is to use polar coordinates to train and generate from the GAN. A transformation is therefore made from $\{x, y, z, p_x, p_y, p_z\}$ to $\{r, \theta_r, z, p_\mathrm{T}, \theta_p, p_z\}$, where θ_r and θ_p are the angles the position $\mathbf{r}$ and the momentum $\mathbf{p}$ vectors make in the x- y- plane respectively. The shape of the transformed distribution is presented in Fig. 8.2. The distribution is symmetric around the beam axis and there is no

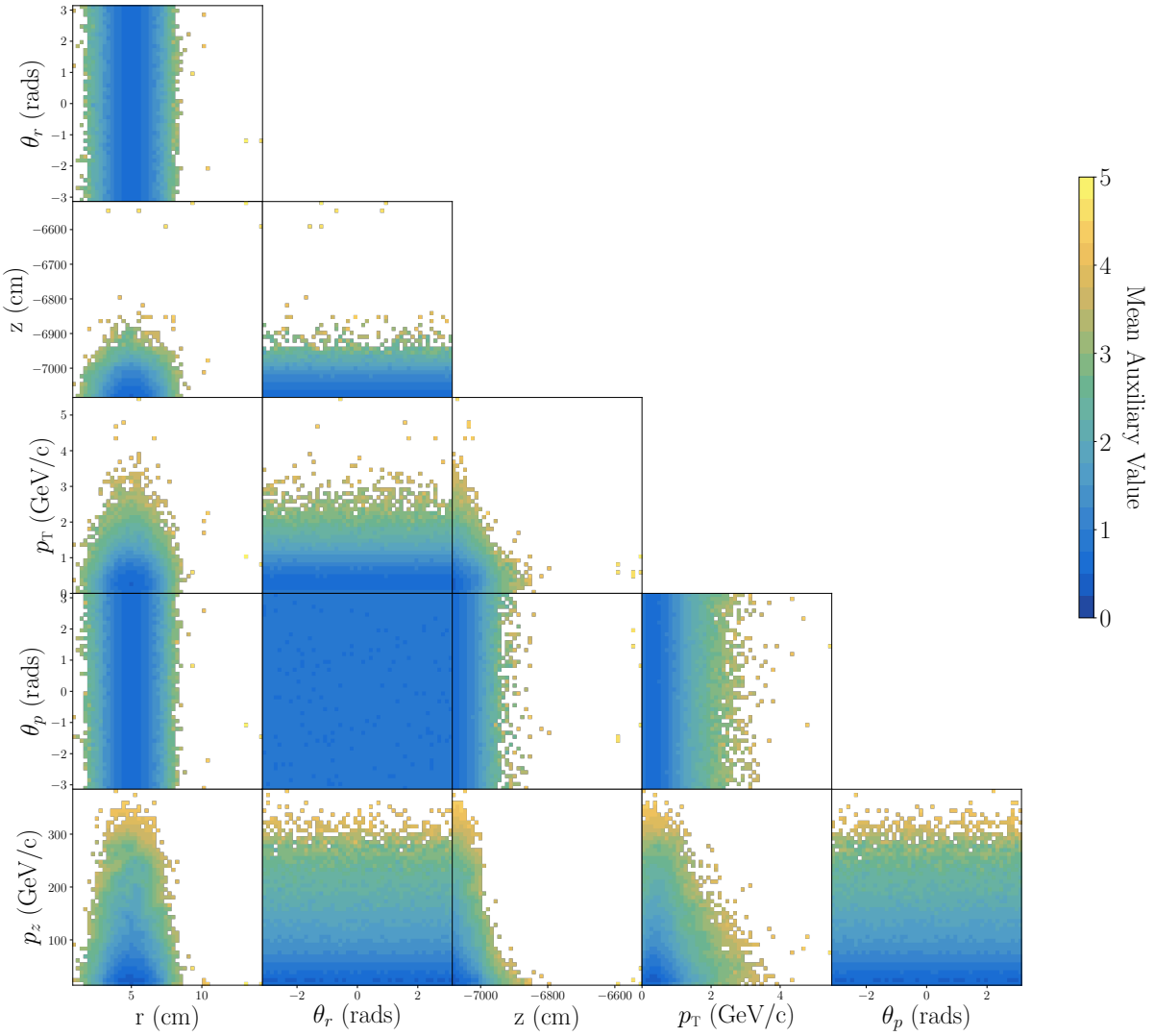


Figure 8.2: Mean A_1 auxiliary values in each bin. The sample is presented in the polar parameterisation.

dependence within the sample on either θ_r- or θ_p-. This allows us to assume complete uniformity and ignore these features in the generation. For each generated sample, values for these angles are sampled from uniform distributions and passed both into and around the generator network. That is, while the values for these values are inputs to the hidden layers their true values are also appended to the output, see Fig. 8.3 for a schematic of the GAN architecture. This approach is helpful as it guarantees symmetry in the generation. Any asymmetries observed in Chapter 5.2.3 in these dimensions will disappear. Passing values around the generator is not essential but it avoids any potential under-population at the edges of $\hat{G}$ output activation space.

In this case the auxiliary values will represent the *rareness* of each muon sample. Two sets of auxiliary parameters are derived in the training sample. The first encodes an estimation of the rareness of each muon in all dimensions into one value. This is calculated using the average distance to the nearest three neighbours in the space of $\{r,\ z,\ p_{\mathrm{T}},\ p_z\}$[1], this will be referred to as A_1. The number of neighbours used was chosen arbitrarily, the number was kept small to avoid lengthy computation of auxiliary values, though a larger number would produce a more accurate determination of local density. The mean value of this auxiliary parameter for each region of sample space is displayed in Fig. 8.2. The second set of auxiliary values is also calculated via a nearest neighbours approach, however, this time independently in each of the dimension resulting in four values for each muon sample, $\vec{A}_4$. A GAN will be trained on each set on auxiliary values separately. Note that these auxiliary values could correspond to any feature of a given problem. For higher dimensional problems these auxiliary values may be more convoluted. For example, for a GAN trained to generate distributions of jet energy deposition in calorimeters a useful auxiliary variable could be the width of each jet, or the total energy deposited.

During training, synthetic auxiliary values are used. Therefore, before training, the true auxiliary distributions of the training sample must be transformed into tractable distributions. These transformations can be non-linear and do not need to be reversible, as long as the information encoded in the auxiliary parameters is preserved. For this task, using single tail Gaussian distributions, $abs(\mathcal{N}(0,1))$, seemed the most natural. Here, low values will represent the core of the distribution and high values the less common muons, see Fig. 8.2. For a problem with one dimension of auxiliary values a transformation is easily achieved via a brute force mapping to a randomly generated vector from $abs(\mathcal{N}(0,1))$, and this mapping can be recomputed after each training epoch. For problems with multiple correlated auxiliary distributions the transformation is less obvious. Synthetically produced auxiliary values will be uncorrelated, so the training auxiliary distribution must also follow this shape. Separately transforming each auxiliary dimension via a brute force mapping separately will not provide the desired result unless the variables themselves are already uncorrelated. There exists no simple transformation

[1] Values are normalised to the range of $\{-1, 1\}$ in each dimension before distances are calculated.

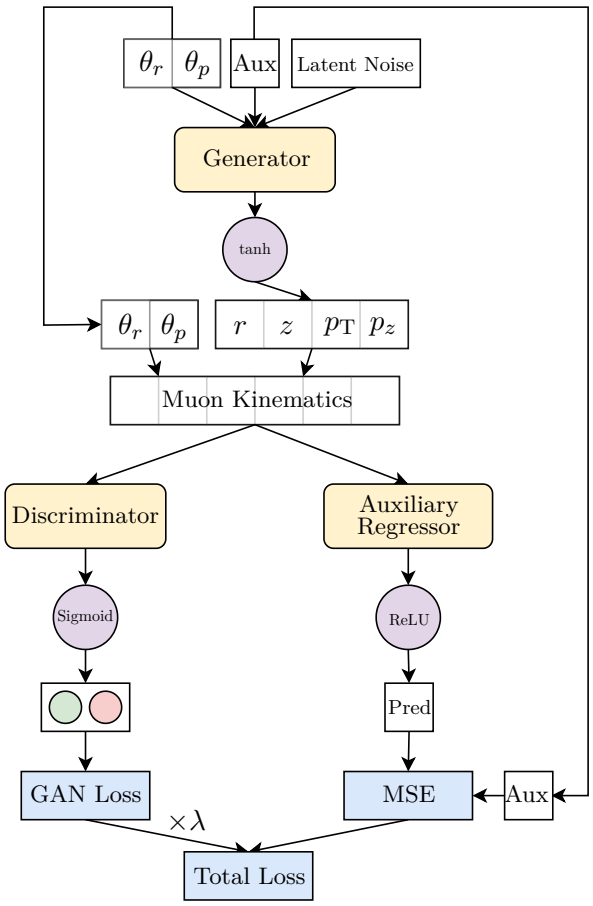


Figure 8.3: Schematic of the auxiliary GAN architecture.

in more than a few dimensions to map a general distribution into the shape of an uncorrelated multidimensional normal whilst also retaining as much information as possible [173, 174]. Such a mapping can be approximated via neural network approaches, for example the Adversarial Autoencoder [175] and the Variational Autoencoder, see Section 8.4.2 and Ref. [176]. For these preliminary studies however, the auxiliary distributions were sufficiently close to uncorrelated that no such complex approach was investigated.

To simplify training for this preliminary investigation, the auxiliary task (or tasks) of the discriminator is separated from the vanilla GAN discriminator. To obtain the results of this chapter, small auxiliary regressor networks are pre-trained to predict auxiliary parameters from information within each muon kinematic vector. These auxiliary networks are only shown information relevant to their task. For example the network regressing the p_T auxiliary parameter

of $\vec{A}_4$ is only shown p_{T}. They are trained using only fully simulated samples. The networks are queried separately to the vanilla discriminator during the GAN training process, see Fig. 8.3.

For λ, the parameter controlling relative weight of each term in the full generator loss of equation 8.1, a value of 10 was chosen to keep the vanilla loss dominant. The vanilla loss must dominate as the auxiliary labels are only obtained from crude estimates of local density, if the networks focus too much on these they will over-fit to any fluctuations. This will lead to asymmetries in the generated output. The choice of $\lambda = 10$ appeared to work well so no further optimisation was attempted. The vanilla loss dominates, maintaining overall sample quality, however the auxiliary loss is still powerful enough such that the generator network learns some understanding of the auxiliary input.

The auxiliary regressor networks are built from two dense layers of $[32, 64]$ nodes. The architecture of $\hat{G}$ and $\hat{D}$ was simplified for this initial testing. All GANs in this chapter have dense layers of $[1000, 1000, 250, 50]$ nodes in both $\hat{G}$ and $\hat{D}$, the sizes of these layers were not optimised. The muon kinematic values are again pre-processsed to values between -1 and 1 in order to employ the tanh activation function in the output layer of $\hat{G}$. Finally, the code used for this chapter is written in `TensorFlow 2` with the new training procedure being based on the code optimised in Chapter 7.

8.4 Network performance

Four different GANs are trained: a vanilla GAN training with the original p_x, p_y parameterisation, $\hat{V}_{xy}$, a vanilla GAN training with the p_{T}, θ_p parameterisation, $\hat{V}_{\theta r}$, an auxiliary GAN using A_1, $\hat{G}_1$, and finally an auxiliary GAN using $\vec{A}_4$, $\hat{G}_4$. The inclusion of the vanilla GAN training with the p_{T}, θ_p parameterisation will allow us to decouple any improvements from the change in parameterisation and the inclusion of the auxiliary term in the loss. The performance of each GAN is again monitored by assessing the figure of merit (FoM) developed in Chapter 4.3.2 at regular intervals during training. Figure 8.4 presents the progression of the FoM for each GAN. For each run the cumulative minima are calculated. It is these values that are then averaged over ten tests to obtain the solid lines.

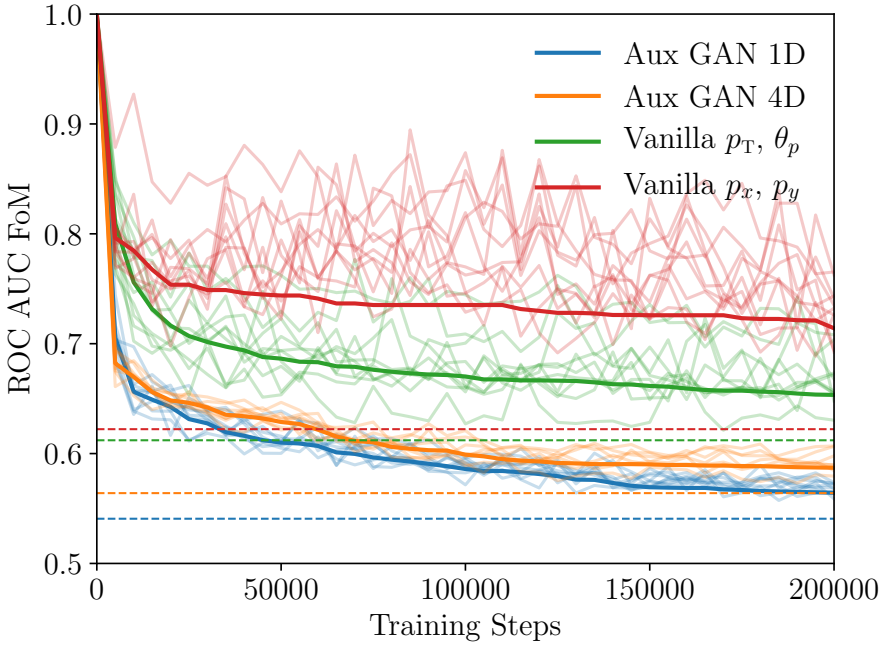


Figure 8.4: Mean of the cumulative minimum of FoM values with training progress.

The progression of the FoM for both vanilla options was unstable showing large variations over a small number of training steps, similar behaviour was observed in Fig. 4.6. These fluctuations in the FoM correspond to large spurious excesses appearing in the generated output. The impact of these effects is reduced by saving the model each time the FoM is evaluated and selecting as a final model that which minimized the FoM. The vanilla GAN did show improved performance when training in the p_T, θ_p parameterisation. Changing the parameterisation and exploiting the symmetry in the distributions and therefore avoiding the generation of θ_r and θ_p has simplified the task. Including the auxiliary term in the loss provided further improvement still. The auxiliary GANs converged faster, with more stability, to improved FoM values. Including the auxiliary information can more effectively punish the network for under- or overpopulating a specific region. The idea is that the auxiliary GAN learns more efficiently. The additional stability provided by the auxiliary loss removes the need cherry pick a model. The procedure is more consistent and large upticks in the FoM are uncommon in the auxiliary training. Note that, the FoM values presented here cannot directly be compared to those in Fig. 4.6, as a slightly different distribution is being modelled. Figure 4.6 also shows $\hat{G}_1$ consistently out performing $\hat{G}_4$. The reason for this could be either some overfitting to the more detailed information in the

4 dimensional auxiliary parameters, or some small differences between the transformed auxiliary distributions and the randomly generated uncorrelated auxiliary parameters used for training as mentioned in Section 8.3.

To obtain the final networks used to produce the results that follow in this chapter, single training runs of the GANs were left for ~ 3 times longer than shown in Fig. 8.4. Here the $\hat{V}_{xy}$ FoM values eventually approached that of $\hat{V}_{\theta r}$, but there was no other change in behaviour. The horizontal dashed lines in Fig. 8.4 indicate the final FoM values obtained.

The higher quality modelling of the distribution is clearly shown in the improvement in FoM values in Fig. 8.4. The FoM values achieved by the auxiliary GAN are far superior to those of either of the vanilla GANs. This is indicative of improved modelling overall but especially of the core of the distribution. It is muons from the core of the distribution which are most likely to contribute to the FoM, as the test only uses a randomly selected 1×10^5 muons. To quantitatively check for any improvements in the tails, 1×10^7 muons are generated from each of $\hat{V}_{xy}$, $\hat{V}_{\theta r}$ and $\hat{G}_1$. For each sample, in exactly the same fashion as before, auxiliary values are calculated from the local density in $\{r,\, z,\, p_{\mathrm{T}},\, p_z\}$ space. From each sample the events with the highest 1% of auxiliary values are extracted, these examples should have been the hardest to model. The FoM is then assessed using only these muons and the equivalent from the training sample. This returned values of 0.679 for $\hat{V}_{xy}$, 0.647 for $\hat{V}_{\theta r}$ and 0.601 for $\hat{G}_1$. The added auxiliary labels provided the superior result.

The auxiliary GANs were originally conceptualised to allow for some tuning of the auxiliary distributions after training. However, the change of parameterisation and the improved training efficiency achieved by adding the auxiliary term to the loss may have done enough to avoid any tuning of the auxiliary distribution for sample generation. This drastic of an improvement was slightly unexpected. If any tuning of the auxiliary distribution was required this could be quickly optimised. This would be easier than an optimisation campaign of a vanilla GAN which would require running multiple lengthy training runs, each with customized modifications to the training sample.

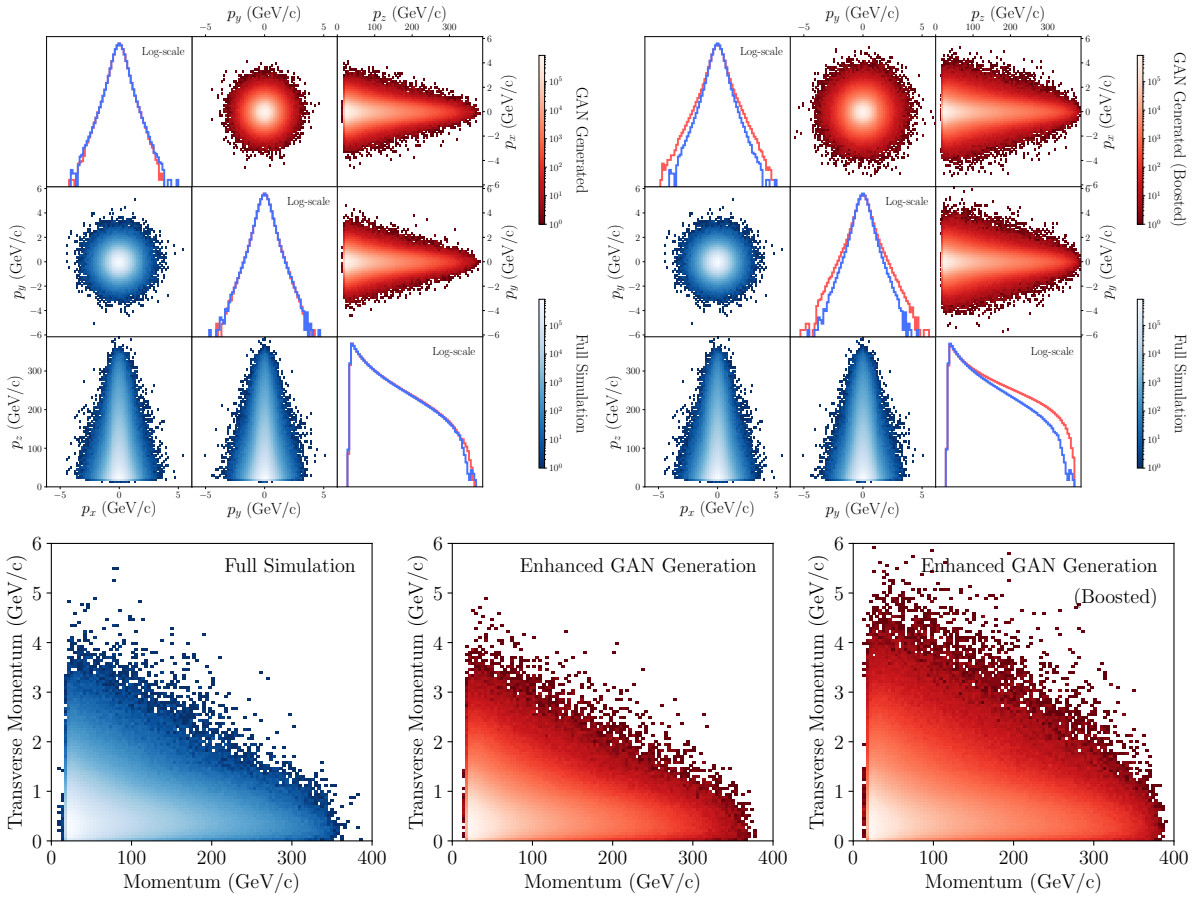


Figure 8.5: Generation examples for $\hat{G}_1$ compared to the training sample.

8.4.1 Generation examples

From the auxiliary GAN $\hat{G}_1$ we can first generate with *raw* auxiliary values, $abs(\mathcal{N}(0,1))$. This results in a generated output equivalent to the physical muon training distribution, see examples on the left of Fig 8.5. Any miss-modellings of the tail regions, such as previously observed in Chapter 4, are not obvious in this generated distribution. As the results of Chapter 4 showed, GANs have a tendency to underestimate the tails. Auxiliary GANs are more flexible and can respond to changes in auxiliary input distributions. The input auxiliary distribution can therefore be *boosted* to generate more rarer events in the tails. An over-exaggerated example is presented on the right of Fig 8.5, here $\hat{G}_1$ has been queried with an auxiliary distribution sampled from $1.2 \times abs(\mathcal{N}(0,1))$.

These results show a marked improvement over those in Chapter 4. These new GANs could

be trained on the physical muon sample and the studies of Chapter 5 could be repeated. The ability to boost the tails of the distribution as demonstrated here could be used as a stress test for the background defences. However, the conclusions of Chapter 5 would likely remain unchanged, the GANs of Chapter 4 already showed the breakdown of the factorisation assumption, removing this significantly boosts background rejection.

8.4.2 Muon shield optimisation

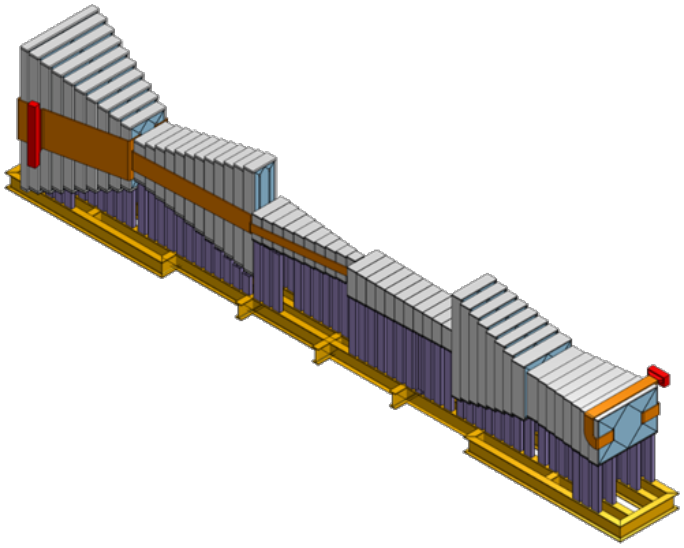


Figure 8.6: Schematic of the current muon shield optimum translated into a step-like configuration, taken from Ref. [18].

An initial optimisation of the muon shield has already been completed by the SHiP collaboration [74]. This provided a baseline configuration that has been shown capable of reducing muon flux within the acceptance of detectors to manageable levels. This optimisation involved selecting a starting point, parameterising the shape of the shield into $\mathcal{O}(50)$ parameters and constructing a loss function. This loss was based on the number of hits in downstream detectors, the length and the mass of the shield. The number of hits is included as reducing background rates is the primary aim of the shield. The length was also included as minimising the distance from target to decay volume is important to preserve the acceptance of the facility, due to large opening angles of HS particles in production within the target. Finally, the mass of the shield was used as a proxy for the overall cost of the shield.

The SHiP collaboration is about to undertake a new optimisation campaign. The main reasons for this are as follows:

- to employ a new loss function. This will include muon hit information from all sub-detectors including crucially the hits in the SND.
- the initial optimisation assumed idealised magnetic fields. Fully realistic fields still cannot be used as the calculations to obtain these are too time consuming to be performed at each iteration of the optimisation however a parameterised approximation is being worked on.
- engineering studies assessing shield construction techniques suggest the design must move to smaller modular magnet blocks. This creates a step-like shield, as depicted in Fig. 8.6. This new shape requires a new parameterisation.
- new materials will be considered for use in the shield. For example, a cobalt core, which although would be more expensive could achieve a stronger field for a small key region.
- the collaboration wish to employ a variety of optimisation techniques. These include the same Bayesian optimisation as the initial optimisation, evolutionary algorithms, and the use of local surrogates to approximate the gradients of the shield configuration [177].
- a new super conducting option will also be investigated that would reduce shield length further.

These new considerations will ensure that the next muon shield configuration is as high performing and as realistic as possible.

A single run of the original optimisation procedure took $\mathcal{O}(10^3)$ iterations, each of which required the simulation of a large sample of muons through a shield configuration. The number of muons used in each step is limited by the time it takes to simulate a muon sample. However, the sample size cannot be reduced too far as already only a small fraction of simulated muons reach the detector acceptance. Simulating the physical muon distribution would be inefficient. The tails of distribution, containing potentially the most dangerous muons, would not get enough attention. Therefore, a re-sampling procedure was employed to reduce sample size whilst both

boosting the representation from rarer events and maintaining coverage over the rest of the distribution [74]. To achieve this, the full sample was binned in the p- p_{T}- plane, each bin was then capped at a specified value. The number of muons in the tails of the distribution was augmented by creating new muons though rotating the momenta of muons in the original sample around the beam axis. This re-sampling procedure reduced the number of muons required to be simulated at each iteration of the optimisation. Note that with this augmented sample only $\mathcal{O}(10)$ muons enter the loss function, this leaves a large uncertainty ($\mathcal{O}(10\%)$) on its value. This problem gets worse for better performing shield configurations.

In the original optimisation each step of every optimisation run employed the same distribution of muons. This GAN approach can generate a new sample for each iteration. This would boost confidence that the optimisation was not overfitting to the background sample or getting stuck in a local minimum. These GAN samples can be generated fast, not affecting the optimisation time. The left plot of Fig. 8.7 presents an example re-sampled distribution of the fully simulated muon training sample that has been created in a very similar way to that of the original optimisation. The auxiliary values $\vec{A}_4$ from the muons in this distribution are fed into the auxiliary GAN $\hat{G}_4$ to produce a synthetic re-sampled distribution, the result of this procedure is presented on the right of Fig. 8.7. This distribution is uniquely generated at every step by synthesising new latent noise vectors. The GAN is trained on the full background muon sample, so does not throw away information about the huge fraction of the muons removed in the old re-sampling procedure. Therefore, with multiple unique generations it can provide more information about the full background sample than a single repeated sub-sample. For an optimisation run, the exact characteristics of the GAN distribution can be tuned as required by controlling the auxiliary distributions. The generated distribution can even be adapted during the optimisation. Using GAN generated samples for the optimisation leaves the fully simulated sample as a powerful unseen sample to test the performance of a final shield configuration.

Evolutionary algorithms have been proposed as new optimisation tools. The algorithms investigated start from a randomly generated population of shield configurations, these are all tests against some input spectrum of muons, features from most successful shield configurations are merged. A new generation of shield configurations is produced from these merged configurations

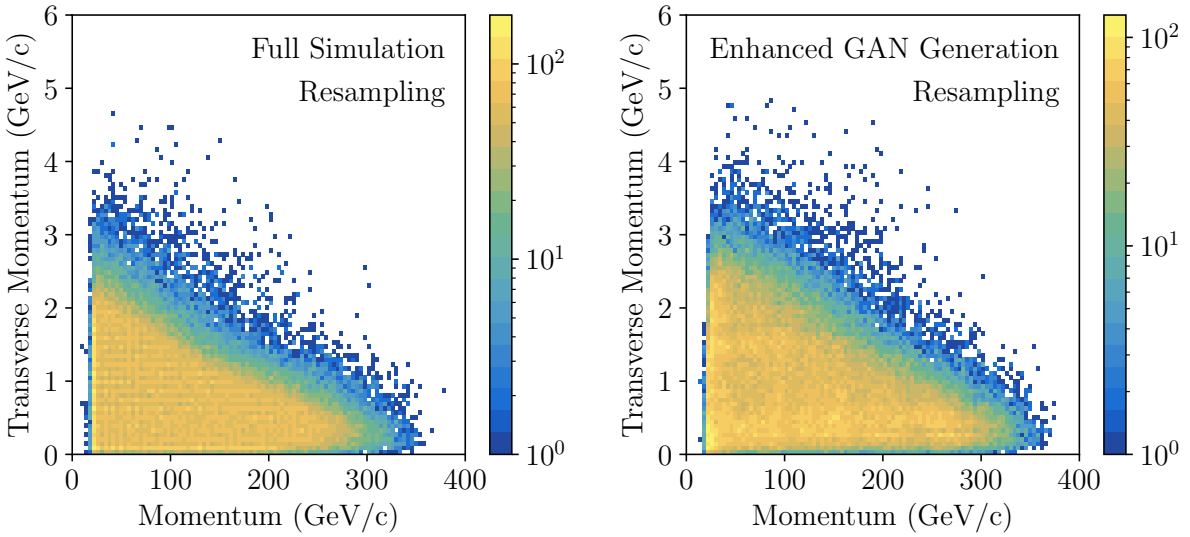


Figure 8.7: Example re-sampled muon distribution and a synthetic distribution from $\hat{G}_4$, there are 1×10^6 entries in each plot.

and the process is repeated. Some flavours of these algorithms require that a component of the input muon distribution at each iteration is based on the characteristics of the muons which beat the last shield configuration. The auxiliary GAN $\hat{G}_4$ can generate new muon samples of similar characteristics to the surviving muons of previous iterations. The kinematics of muons that beat the shield act as *seeds*, the kinematics are fed through the auxiliary regressor networks to derive new auxiliary parameters. These are then combined with latent noise vectors and passed to the generator to produce new, unseen, muons with similar characteristics. Figure 8.8 provides an example of this augmentation behaviour working for a representative input sample of 5 muons. There are a few things to note:

- in some distributions the generated muon kinematics are split into two islands, this is clearest in the x- y- projection. This is due to the choice of the auxiliary parameters being defined in terms of local density and so each generated island lies in a region of similar local density[2].

- the auxiliary regressor networks can miss-judge the correct auxiliary parameter, this is most clear in the purple example, where the z- coordinate is being over estimated.

[2]Note, the auxiliary parameters were defined in $\{r, \theta_r, z, p_T, \theta_p, p_z\}$, not in $\{x, y, z, p_x, p_y, p_z\}$ as muons are displayed in Fig. 8.8.

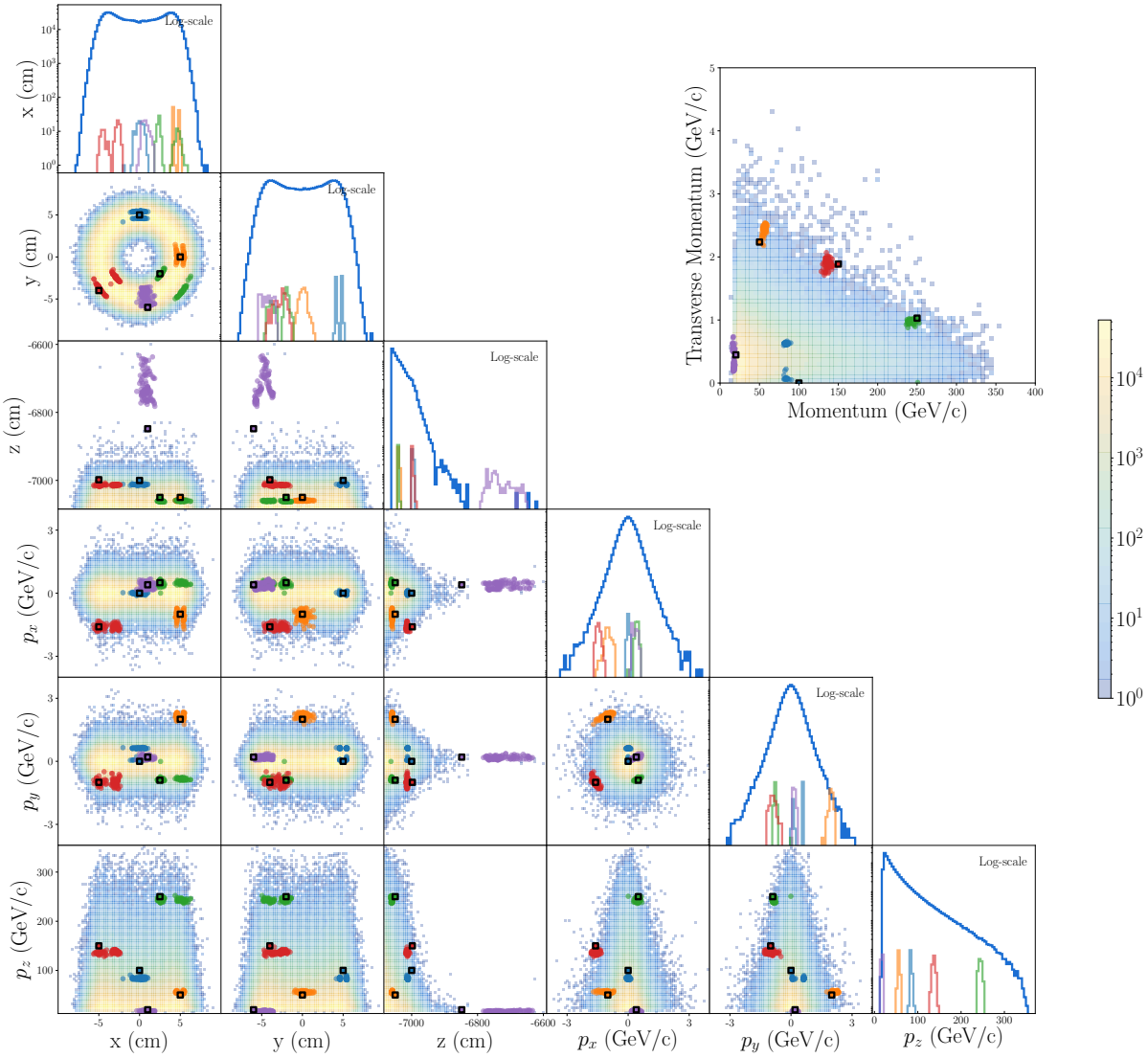


Figure 8.8: Five example seed kinematics are outlined with black squares and 100 generated muons from $\hat{G}_4$ are plotted with the corresponding colour. The points are overlayed on the full training sample.

- this auxiliary GAN architecture has θ_r and θ_p as $\hat{G}_4$ inputs which are passed around the network. Therefore, values for these angles in the seeds are first calculated exactly, and noise is then added in the generation of each sample.

8.4.3 Outlook and VAEs

Overall for a preliminary study, the performance shown in Fig. 8.8 is impressive, especially when considering that the auxiliary GANs were not originally conceived to solve this issue. The

correct tool for this job is probably the variational auto-encoder (VAE), though this has not yet been explored. A standard autoencoder is comprised of two networks, an encoder and a decoder. The encoder maps an input sample to a lower-dimensional latent space, a meaningful representation of the training sample. The decoder is then tasked at reconstructing the original sample. For the VAE, the encoder does not directly produce the latent variables. Instead the encoder outputs two values μ and $log(\sigma)$, these are then combined with a noise term $\epsilon = \mathcal{N}(0, 1)$ to sample a latent space z as,

$$z = \mu + e^{0.5 \times log(\sigma)} \times \epsilon. \tag{8.4}$$

The latent space is populated by the sampling of many of these small normal distributions. The latent space is then forced to form an n-dimensional normal distribution via a Kullback–Leibler divergence term in the loss,

$$L_{KL} = -0.5 \times \sum_i (1 + log(\sigma)_i - \mu_i^2 - e^{log(\sigma)}). \tag{8.5}$$

a reconstruction loss is combined with L_{KL}, this compares the initial sample with the reconstruction attempt. Synthetic samples can be generated from the VAE by passing synthetic $\mathcal{N}(0, 1)$ distributions through the decoder. For the task of generating muons with similar characteristics to those that broke the shield in the last iteration of the optimisation, the VAE provides a natural solution. The dangerous muon kinematics would be passed through the encoder to obtain seed values for μ and $log(\sigma)$. These can be combined with generated values of ϵ, which could be scaled depending on how similar one wanted the muons to be, to get a latent vector z'. Then z' would be passed through the decoder to obtain the augmented sample. VAEs are known to produce lower quality synthetic samples than GANs, due to the lack of an adversarial component in the training procedure. The reconstruction loss is distanced based, such losses can struggle for example at sharp features in a distribution. There have been attempts to address this with more exotic models combining GANs and VAEs [178–180] but no perfect model exists, and each application requires a bespoke solution.

Chapter 9

Angular analysis of $B^0 \to K^{*0}(\to K^+\pi^-)\,\mu^+\mu^-$ at LHCb

This chapter presents details of my contribution to the q^2-binned angular analysis of $B^0 \to K^{*0}(\to K^+\pi^-)\,\mu^+\mu^-$ at LHCb using the pp collision data from 2011, 2012, 2016, 2017 and 2018, corresponding to an integrated luminosity of 9fb^{-1}.

9.1 The LHCb detector

The LHCb detector at CERN is a single arm forward spectrometer designed to study the decays of particles containing b and c quarks. The acceptance of the detectors cover a pseudorapidity range of $2 < \eta < 5$. A schematic of the detector is presented in Fig. 9.1, the key characteristics are as follows. Extremely precise tracking is provided in the silicon-strip vertex locator (VELO) surrounding the pp interaction region. Information from the VELO is used to reconstruct and identify displaced secondary vertices, indicative of b- and c-hadron decays. The performance of the VELO in terms of the impact parameter (IP) resolution is summarised in Fig. 9.2. Efficient charged hadron discrimination is provided by two ring-imaging Cherenkov detectors, these offer $\sim 100\%$ K/π discrimination for all but the highest momentum particles, see Fig. 9.3. A large magnet of bending power 4 Tm is employed with tracking stations to measure charged

particle momentum, the momentum resolution achieved is displayed in Fig. 9.4. Downstream, the energy of photons, electrons and hadrons are measured in electromagnetic and hadron calorimeters. Finally, muons are identified in segmented multi-wire proportional chambers with $\sim 100\%$ efficiency, see Fig. 9.3. LHCb employs both an online hardware based trigger making decisions based on high-level information and a software based trigger which completes full event reconstruction.

A more detailed description of the detector can be found in Ref. [181].

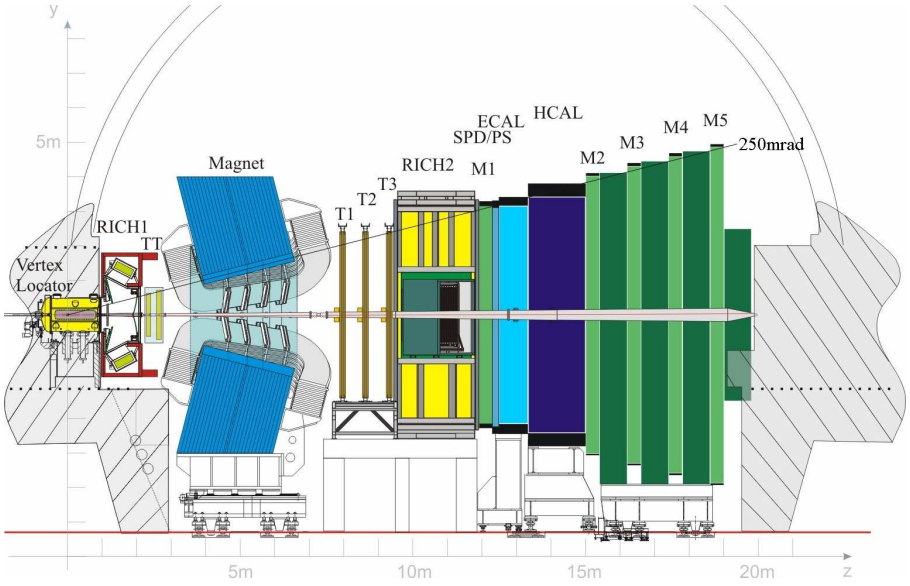


Figure 9.1: Schematic of the LHCb detector, from Ref. [181].

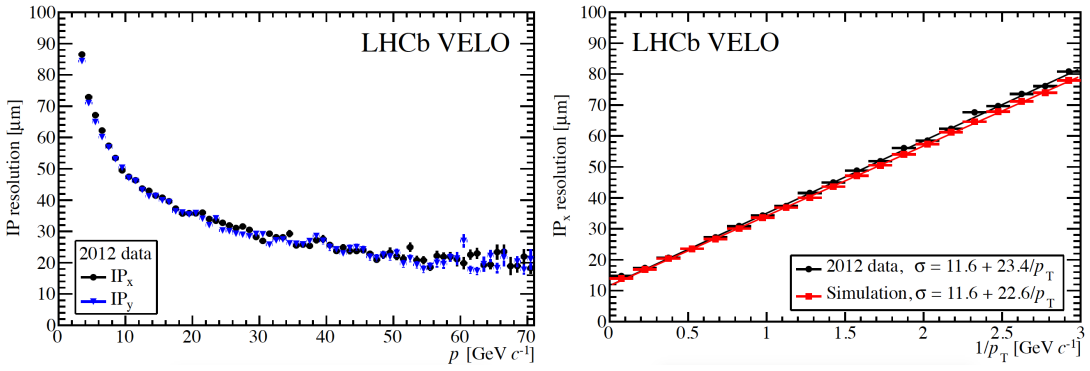


Figure 9.2: The resolution of impact parameter measurements as a function of p and $1/p_{\mathrm{T}}$ [182].

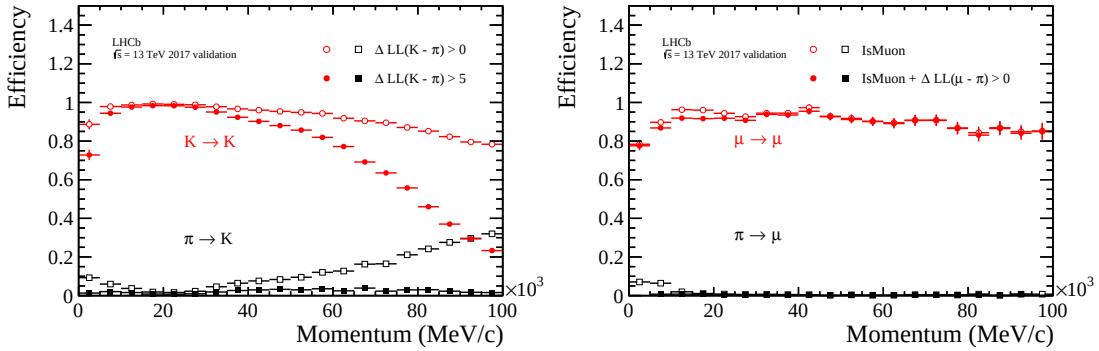


Figure 9.3: Particle identification efficiencies for kaons and muons as a function of momentum [183].

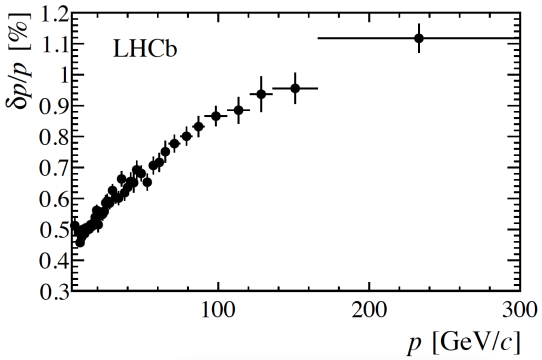


Figure 9.4: Relative momentum resolution achieved at LHCb as a function of momentum in $J/\psi \to \mu^+\mu^-$ decays [184].

9.2 $b \to s\ell^+\ell^-$ processes

Flavour Changing Neutral Currents (FCNC) are forbidden at tree level in the Standard Model. Instead, FCNC processes occur at loop level and are suppressed due to the presence of more vertices and the mass of the mediating particle. The decay $B^0 \to K^{*0}\mu^+\mu^-$ is one such process involving the quark level transition $b \to s\ell^+\ell^-$. Figure 9.5 shows example diagrams of this decay. This suppression makes the process an ideal place to look for NP which may not require loop processes and could therefore enter at a similar order to the SM itself even if the interaction strength of the NP is very small. The presence of NP could then measurably affect the decay rate of these $b \to s\ell^+\ell^-$ processes, distort the angular distributions of decay products, or could favour one lepton flavour over another leading to breakdown of lepton universality in the SM. LHCb has observed inconsistencies in all three of these areas, although no single

measurement is yet precise enough to confirm discovery [6, 12, 185–187]. Angular analyses are of particular interest as they can provide the best model discrimination. Best fit values for Wilson coefficients, describing the strength of various contributions to a decay process, can be obtained and compared to Standard Model predictions [6, 188]. While angular observables can be constructed for which SM predictions are relatively clean with small uncertainties, it is still possible that some uncertainties in the strength of non-local contributions are causing part of the observed anomalies. This is the so called *charm loop*, where the same final state is achieved through a $c\bar{c}$ loop, see Fig. 9.6. New results for the extremely theoretical clean ratios R_K and R_{K^*}, where $R_K = \mathcal{B}(B \to K\mu^+\mu^-)/\mathcal{B}(B \to Ke^+e^-)$, may clear up this debate if they reveal disparities to the SM significant to 5σ as charm loop contributions will cancel. If this is the case the most likely explanation to describe the emerging pattern of results is a *leptoquark* that couples more strongly to the 2$^{\text{nd}}$ and 3$^{\text{rd}}$ than lepton generations that that of 1$^{\text{st}}$ generation [189, 190].

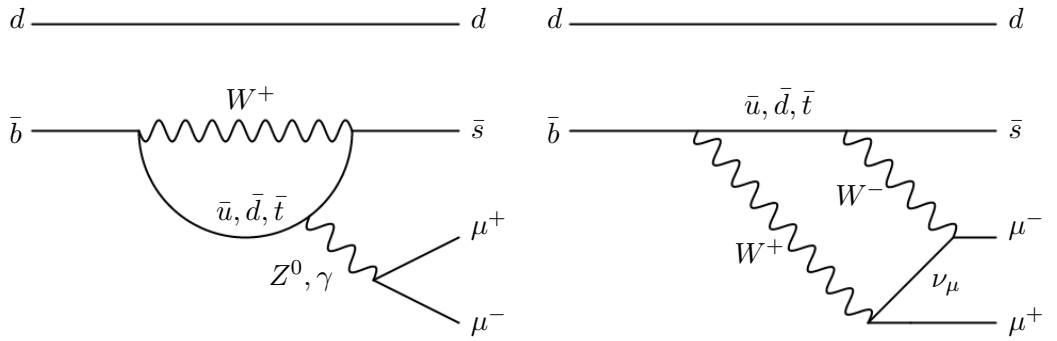


Figure 9.5: Feynman diagrams of leading order contributions to $B^0 \to K^{*0}\,\mu^+\,\mu^-$.

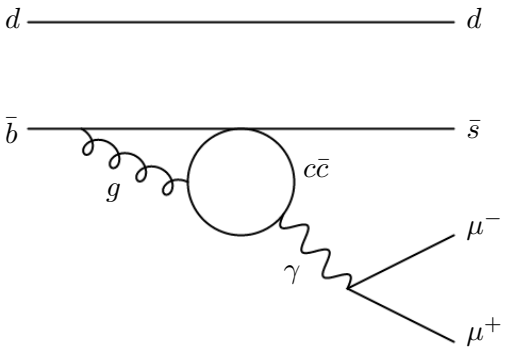


Figure 9.6: Feynman diagrams of charm loop contribution to $B^0 \to K^{*0}\,\mu^+\,\mu^-$.

9.3 Effective field theory and operator product expansion

An effective field theory (EFT) can be used to describe the characteristics of $b \to s\ell^+\ell^-$ processes. There are two main advantages that come with using an EFT, it is a model independent approach where no assumptions are made about the the NP involved and different energy scales can be handled separately [191]. There are however challenges in this approach. Firstly, the factorisation of physical effects at different energy scales must be correct. Secondly, the low energy hadronic system must be parameterised with *Form Factors* and these come with large uncertainties.

The EFT is expressed as an operator product expansion (OPE), this is a summation of physical effects which are split into long and short distance effects according to an energy scale μ. The long distance effects are low energy contributions such as QCD processes, to model these effects we define operators $\mathcal{O}_i$ which account for effects below μ. The short distance effects are the high energy contributions such as weak interactions and NP effects, these processes involve energies greater than μ. These contributions are described with scalar couplings called the Wilson Coefficients C_i.

The full Lagrangian density $\mathcal{L}$ is the sum of the Standard Model $\mathcal{L}_{SM}$ and any NP contribution $\Delta\mathcal{L}_{NP}$. This can be parameterised with the OPE and the terms described above as an effective Lagrangian $\mathcal{L}_{\text{eff}}$,

$$\mathcal{L} = \mathcal{L}_{SM} + \Delta\mathcal{L}_{NP} \quad \to \quad \mathcal{L}_{\text{eff}} = \sum_i C_i(\mu)\mathcal{O}_i(\mu). \tag{9.1}$$

The equivalent effective Hamiltonian $\mathcal{H}_{\text{eff}}$ for the transition from state $|i\rangle$ to state $|f\rangle$ at the energy scale μ is then,

$$\langle f|\mathcal{H}_{\text{eff}}|i\rangle = \sum_i C_i(\mu)\ \langle f|\mathcal{O}_i|i\rangle\Big|_\mu . \tag{9.2}$$

Note that the Wilson coefficients are often split up as,

$$C_i = C_i^{SM} + C_i^{NP}, \tag{9.3}$$

where C_i^{SM} are Standard Model predictions and any measured deviation in C_i from these predictions would imply the presence of some NP interaction C_i^{NP}.

The effective Hamiltonian of the $B^0 \to K^{*0}\,(\to K^+\pi^-)\,\mu^+\mu^-$ interaction is described with 12 Wilson Coefficients $C_{1,\ldots 10,S,P}$ and 12 long distance operators $\mathcal{O}_{1,\ldots 10,S,P}$, however many of

these contributions are suppressed or well constrained [192]. The only relevant contributions to this process are $\mathcal{O}_7$, $\mathcal{O}_9$ and $\mathcal{O}_{10}$ and their chiral partners (e.g. $\mathcal{O}'_7$). $\mathcal{O}_7$ is the electromagnetic operator and is associated with processes involving photon currents, $\mathcal{O}_9$ is the vector semi-leptonic operator and $\mathcal{O}_{10}$ the axial-vector semi-leptonic operator. The regions in q^2 in which each of these Wilson Coefficients dominates is depicted in Fig. 9.7.

An additional step is made to define effective Wilson Coefficients, these are convenient combinations of Wilson Coefficients. The relevant effective Wilson Coefficients are built as follows,

$$C_7^{\text{eff}} = \frac{4\pi}{\alpha_s}C_7 - \frac{1}{3}C_3 - \frac{4}{9}C_4 - \frac{20}{3}C_5 - \frac{80}{9}C_6, \tag{9.4}$$

$$C_9^{\text{eff}} = \frac{4\pi}{\alpha_s}C_9 + Y(q^2), \tag{9.5}$$

$$C_{10}^{\text{eff}} = \frac{4\pi}{\alpha_s}C_{10}, \tag{9.6}$$

$$C_{7,9,10}^{\prime\text{eff}} = \frac{4\pi}{\alpha_s}C'_{7,9,10}. \tag{9.7}$$

The function $Y(q^2)$ is constructed from other Wilson coefficient and describes the contribution from non-local effects and corrects for the presence of charm-loop diagrams, as mentioned in Section 9.2. In the SM, C_7^{eff} is approximately -0.3, $C_7^{\prime\text{eff}}$ is approximately -0.006, C_9^{eff} is approximately 4.27 and C_{10}^{eff} is approximately -4.16. In the SM there are no right handed vector or axial-vector currents so $C_9^{\text{eff}\prime} = 0$ and $C_{10}^{\text{eff}\prime} = 0$.

The Wilson Coefficients are not directly observable and are instead constructed from observables, as will be introduced in Section 9.4, and the aforementioned Form Factors. The Form Factors parameterise the $B \to K^*$ system, their behaviour and their theoretical treatment vary with q^2. They are calculated using non-perturbative techniques, such as Light Cone Sum Rules at low-q^2 and Lattice-QCD at high-q^2 [193], resulting in large uncertainties. These uncertainties often dominate the theoretical uncertainties in observable predictions.

9.4 Decay rate formalism

To provide a minimal but full description of the angular distribution of the $B^0 \to K^{*0}\,\mu^+\,\mu^-$ decay the following variables are defined. Firstly, q^2 is the invariant mass of the di-muon pair

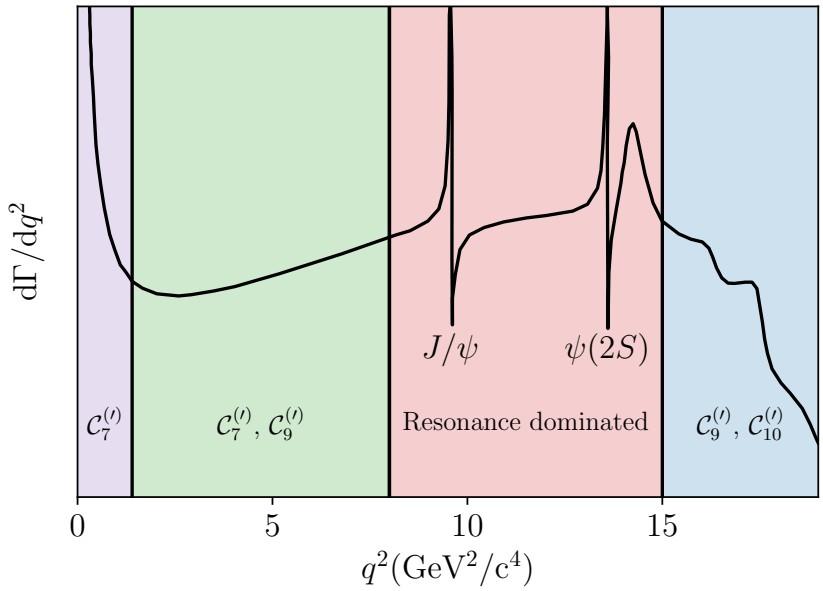


Figure 9.7: Differential decay rate of $B^0 \to K^{*0}\mu^+\mu^-$ [191]. Different regions of the q^2 spectrum are labelled to indicate the Wilson coefficients that dominate the physics. The photon pole dominates the low q^2 region, coloured in purple.

squared. Then $m_{K\pi}$ is the $K^+\pi^-$ invariant mass providing information about the K^{*0} resonance involved, where the K^{*0} can be in both a P-wave $J = 1$ or an S-wave $J = 0$ state. Then finally, three angular terms, $cos(\theta_K)$, $cos(\theta_l)$ and $cos(\phi)$. Where θ_K is the angle between K^+ in the $K^+\pi^-$ rest frame and $K^+\pi^-$ in the rest frame of the B^0, θ_l is the angle between μ^+ in the rest frame of $\mu^-\mu^+$ and the $\mu^-\mu^+$ object in the rest frame of the B^0. Finally ϕ is the angle between the planes of $\mu^-\mu^+$ and $K^+\pi^-$ in the rest frame of the B^0. Diagrams of these angles are presented in Fig. 9.8.

The form of the $\bar{B}^0$ angular decay rate for the component of the rate that has K^{*0} in a P-wave configuration is then

$$\begin{aligned}\frac{\mathrm{d}\Gamma_P}{\mathrm{d}m_{K\pi}\mathrm{d}\cos\theta_\ell\mathrm{d}\cos\theta_K\mathrm{d}\phi} = \frac{9}{32\pi}[&J_{1s}\sin^2\theta_K + J_{1c}\cos^2\theta_\ell + J_{2s}\sin^2\theta_K\cos 2\theta_\ell + J_{2c}\cos^2\theta_K\cos\theta_\ell \\ &+J_3\sin^2\theta_K\sin^2\theta_\ell\cos 2\phi + J_4\sin 2\theta_K\sin 2\theta_\ell\cos\phi \\ &+J_5\sin 2\theta_K\sin\theta_\ell\cos\phi + J_{6s}\sin^2\theta_K\cos\theta_\ell \\ &+J_{6c}\cos^2\theta_K\cos\theta_\ell + J_7\sin 2\theta_K\sin\theta_\ell\sin\phi + J_8\sin 2\theta_K\sin 2\theta_\ell\sin\phi \\ &+J_9\sin^2\theta_K\sin^2\theta_\ell\sin 2\phi] \times |BW_P(m_{K\pi})|^2, \end{aligned} \tag{9.8}$$

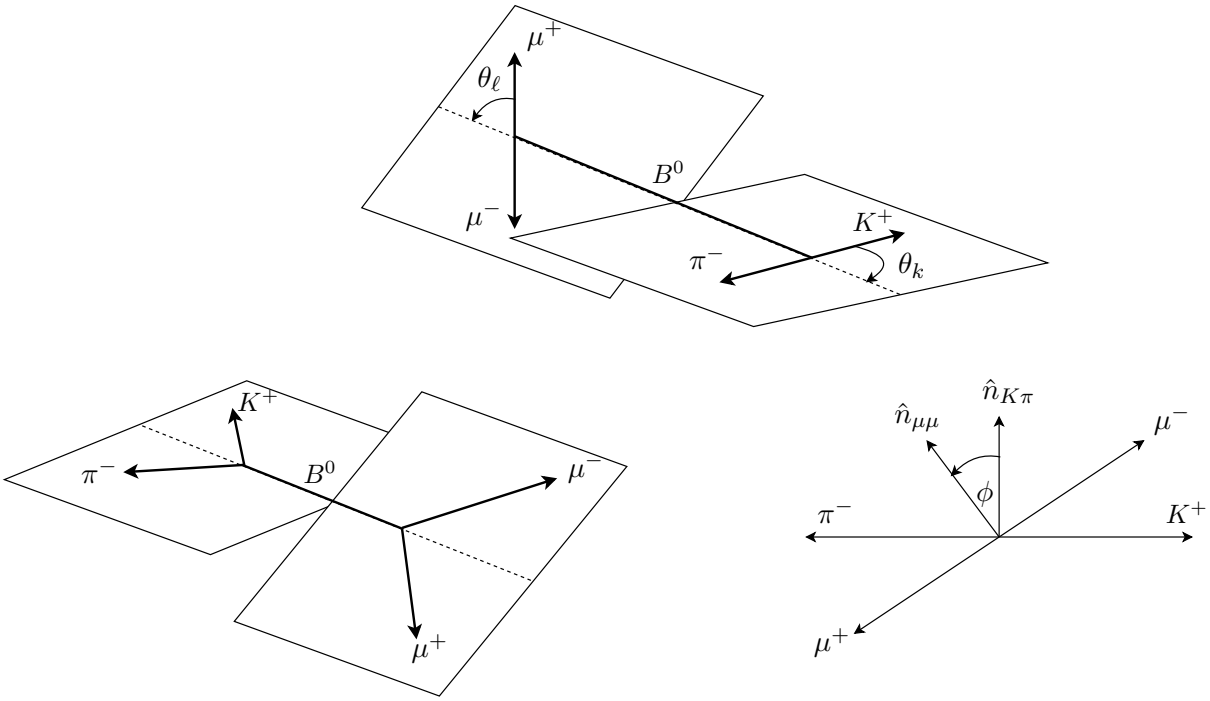


Figure 9.8: Diagrams displaying the definition of each angle used to describe the $B^0 \to K^{*0}\,\mu^+\,\mu^-$ decay θ_ℓ, θ_k and ϕ [194].

and similarly for the B^0 [195]. The J terms can be viewed as factors controlling the size of each orthogonal angular term in equation 9.8. An example of a J term would be,

$$J_5 = \sqrt{2}\beta \mathrm{Re}(\mathcal{A}_0^{\mathrm{L}}\mathcal{A}_\perp^{\mathrm{L}*} - \mathcal{A}_0^{\mathrm{R}}\mathcal{A}_\perp^{\mathrm{R}*}). \tag{9.9}$$

The $m_{K\pi}$ dependence is included in equation 9.8 with the relativistic Breit-Wigner amplitude $BW_P(m_{K\pi})$, given by

$$BW_P(m_{K\pi}) = B'_{L_B}(p, p_R, d) \cdot \left(\frac{k}{m_B}\right)^{L_B} \cdot B'_{L_{K^{*0}}}(k, k_R, d) \cdot \left(\frac{k}{m_{K\pi}}\right)^{L_{K^{*0}}} \cdot \frac{1}{m^2_{K^{*0}} - m^2_{K\pi} - i m_{K^{*0}}\Gamma_{K^{*0}}(m_{K\pi})}, \tag{9.10}$$

where k (p) is the momentum of K^+ (K^{*0}) in the rest frame of the K^{*0} (B^0) evaluated at a given $m_{K\pi}$, k_R and p_R are the equivalent quantities evaluated at the $K^*(892)^0$ resonance, $L_B = 0$ and $L_{K^{*0}} = 1$ are the orbital angular momenta, d is meson radius parameter, and B'_L is a Blatt-Weisskopf barrier factor, where

$$B'_{L=0}(p, p_R, d) = 1, \qquad B'_{L=1}(p, p_R, d) = \sqrt{\frac{1 + (pd)^2}{1 + (p_R d)^2}}, \tag{9.11}$$

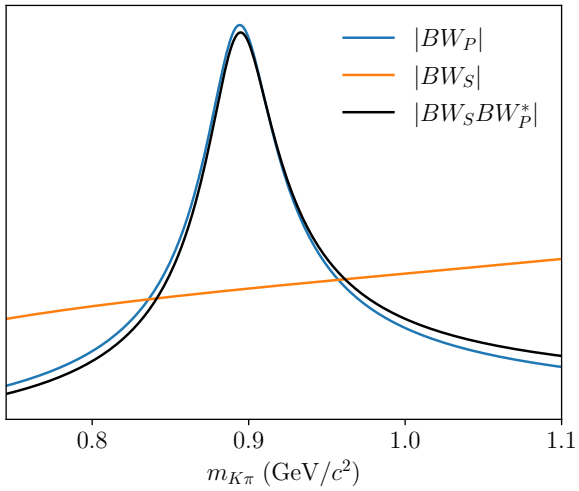


Figure 9.9: Lineshapes of the P-wave ($|BW_P|$), S-wave ($|BW_S|$) and interference ($|BW_S * BW_P|$) contributions to the decay. The lineshapes are normalised such that they have the same area for this plot.

and running width of the resonance $\Gamma_J(m_{K\pi})$ is

$$\Gamma_J(m_{K\pi}) = \Gamma_R \cdot B'^2_{L_R}(k, k_R, d) \left(\frac{k}{k_R}\right)^{2L_R+1} \left(\frac{m_R}{m_{K\pi}}\right), \tag{9.12}$$

where Γ_R and m_R are the pole mass and width of the resonance, in this case $K^*(892)^0$ and $L_R = L_{K^{*0}}$. The $|BW_P|$ lineshape in $m_{K\pi}$ is presented in Fig. 9.9.

The K^{*0} can also be in an S-wave configuration. This contribution adds a further two transversity amplitudes, $\mathcal{A}_S^{L,R}$, which lead to S-wave, and P- and S-wave interference terms $\tilde{J}_i$. These make up the following S-wave contribution to the angular decay rate,

$$\begin{aligned}
\frac{\mathrm{d}\Gamma_S}{\mathrm{d}\cos\theta_\ell \mathrm{d}\cos\theta_K \mathrm{d}\phi \mathrm{d}m_{K\pi}} = \frac{1}{4\pi} \Big[& (\tilde{J}^c_{1a} + \tilde{J}^c_{2a}\cos 2\theta_\ell)|BW_S|^2 \\
& + [\tilde{J}^{c,r}_{1b}\mathrm{Re}(BW_S BW_P^*) - \tilde{J}^{c,i}_{1b}\mathrm{Im}(BW_S BW_P^*)]\cos\theta_K \\
& + [\tilde{J}^{c,r}_{2b}\mathrm{Re}(BW_S BW_P^*) - \tilde{J}^{c,i}_{2b}\mathrm{Im}(BW_S BW_P^*)]\sin 2\theta_\ell \cos\theta_K \\
& + [\tilde{J}^{r}_{4}\mathrm{Re}(BW_S BW_P^*) - \tilde{J}^{i}_{4}\mathrm{Im}(BW_S BW_P^*)]\sin 2\theta_l \sin\theta_K \cos\phi \\
& + [\tilde{J}^{r}_{5}\mathrm{Re}(BW_S BW_P^*) - \tilde{J}^{i}_{5}\mathrm{Im}(BW_S BW_P^*)]\sin\theta_l \sin\theta_K \cos\phi \\
& + [\tilde{J}^{r}_{7}\mathrm{Im}(BW_S BW_P^*) + \tilde{J}^{i}_{7}\mathrm{Re}(BW_S BW_P^*)]\sin\theta_l \sin\theta_K \sin\phi \\
& + [\tilde{J}^{r}_{8}\mathrm{Im}(BW_S BW_P^*) + \tilde{J}^{i}_{8}\mathrm{Re}(BW_S BW_P^*)]\sin 2\theta_l \sin\theta_K \sin\phi \Big] .
\end{aligned} \tag{9.13}$$

Here BW_S is the S-wave $m_{K\pi}$ amplitude which is described with the LASS parameterisation [196–201]. The LASS lineshape is described as,

$$BW_S(m_{K\pi}) = B'_{L_B}(p, p_{1430}, d) \cdot \left(\frac{p}{p_{1430}}\right)^{L_B} \cdot \sqrt{kp} \cdot \frac{m_{K\pi}}{k} \cdot \left(\frac{1}{cot\delta_B - i} + \frac{1}{cot\delta_R - i}\right), \quad (9.14)$$

where $L_B = 1$ and $cot\delta_B$, describing the non-resonant component of the LASS model, is defined as

$$cot\delta_B = \frac{1}{ak} + \frac{rk}{2}, \quad (9.15)$$

where the parameters a and r are obtained empirically, and $cot\delta_R$, the contribution from the $K^*(1430)^0$ resonance, is defined as

$$cot\delta_R = \frac{m_{1430}^2 - m_{K\pi}^2}{m_{1430}\Gamma_{1430}(m_{K\pi})}, \quad (9.16)$$

where $\Gamma_{1430}(m_{K\pi})$ is the running width as defined in equation 9.12. The values used in this analysis for the parameters a and r in this analysis are taken from Ref. [197]. The $|BW_S|$ lineshape in $m_{K\pi}$ is presented in Fig. 9.9.

From J_i and $\bar{J}_i$ terms we can form q^2 dependent CP-averaged, S, and CP-asymmetric, A, terms as, for the example of P-wave observables,

$$S_i^P = \left(J_i + \bar{J}_i\right) \Big/ \left(\frac{d\Gamma_P}{dq^2} + \frac{d\bar{\Gamma}_P}{dq^2}\right), \qquad A_i^P = \left(J_i - \bar{J}_i\right) \Big/ \left(\frac{d\Gamma_P}{dq^2} + \frac{d\bar{\Gamma}_P}{dq^2}\right), \quad (9.17)$$

it is these values which are extracted from the fit. From these S and A observables the following definitions are made by convention. The longitudinal polarisation of the K^{*0}, F_L, is then

$$F_L = S_{1c} = \frac{|\mathcal{A}_0^L|^2 + |\mathcal{A}_0^R|^2}{|\mathcal{A}_0^L|^2 + |\mathcal{A}_0^R|^2 + |\mathcal{A}_\parallel^L|^2 + |\mathcal{A}_\parallel^R|^2 + |\mathcal{A}_\perp^L|^2 + |\mathcal{A}_\perp^R|^2}, \quad (9.18)$$

the forward-backward asymmetry of the di-muon system, A_{FB},

$$A_{FB} = \frac{3}{4} S_{6s}. \quad (9.19)$$

Then additionally the fraction of the system in an S-wave configuration is denoted F_S, this is defined in Ref. [185].

Integrating out the angular and $m_{K\pi}$ dependence in equation 9.8 and 9.13 yields the following expressions,

$$\frac{d\Gamma_P}{dq^2} = \frac{3}{4}(2J_{1s} + J_{1c}) - \frac{1}{4}(2J_{2s} + J_{2c}), \qquad \frac{d\Gamma_S}{dq^2} = 2\tilde{J}_{1a}^c - \frac{2}{3}\tilde{J}_{2a}^c. \quad (9.20)$$

Combining the P-wave expression with equation 9.17 leads to the following expression relating CP-averaged observables,

$$1 = \frac{3}{4}(2S_{1s} + S_{1c}) - \frac{1}{4}(2S_{2s} + S_{2c}). \tag{9.21}$$

This can then be used to express one of $\{S_{1s}, S_{1c}, S_{2s}, S_{2c}\}$ in terms of the others, simplifying the fit. For the majority of the q^2 range explored an assumption of massless leptons can be made, giving $\beta = 1$. This assumption leads to the following further relations,

$$\begin{aligned} S_{1c} &= -S_{2c}, & S_{1s} &= 3S_{2s}, \\ A_{1c} &= -A_{2c}, & A_{1s} &= 3A_{2s}, \end{aligned} \tag{9.22}$$

which reduces the number of angular observables in bins where the massless assumption holds.

The naming conventions used in this analysis and a note on the origin of each observable is provided in Table 9.1.

Observable		Observable	
FL	from J_{1c} in 9.8 and 9.21 and 9.18	SS5re	from $\tilde{J}_7^r$ in 9.13 and 9.21
S3	from J_3 in 9.8 and 9.21	SS1im	from $\tilde{J}_{2b}^{c,r}$ in 9.13 and 9.21, $\tilde{J}_{2b}^{c,r} = -\frac{3}{8}$SS1im
S4	from J_4 in 9.8 and 9.21	SS2im	from $\tilde{J}_4^i$ in 9.13 and 9.21
S5	from J_5 in 9.8 and 9.21	SS3im	from $\tilde{J}_5^i$ in 9.13 and 9.21
AFB (S6)	from J_6 in 9.8 and 9.21 and 9.19	SS4im	from $\tilde{J}_6^i$ in 9.13 and 9.21
S7	from J_7 in 9.8 and 9.21	SS5im	from $\tilde{J}_7^i$ in 9.13 and 9.21
S8	from J_8 in 9.8 and 9.21	S1s	from J_{1s} in 9.8 and 9.21
S9	from J_9 in 9.8 and 9.21	S1c	from J_{1c} in 9.8 and 9.21
FS	see Ref. [185]	S1ac	from $\tilde{J}_{1a}^c$ in 9.13 and 9.21
SS1re	from $\tilde{J}_{2b}^{c,r}$ in 9.13 and 9.21, $\tilde{J}_{2b}^{c,r} = -\frac{3}{8}$SS1re	S1bcre	from $\tilde{J}_{1b}^{c,r}$ in 9.13 and 9.21
SS2re	from $\tilde{J}_4^r$ in 9.13 and 9.21	S1bcim	from $\tilde{J}_{1b}^{c,i}$ in 9.13 and 9.21
SS3re	from $\tilde{J}_5^r$ in 9.13 and 9.21	S2s	from J_{2s} in 9.8 and 9.21 (removed with 9.21)
SS4re	from $\tilde{J}_6^r$ in 9.13 and 9.21		

Table 9.1: A guide to the naming convention used in the analysis and origin of each observables. Note the asymmetric observables just either have S swapped with an A, or an A appended to the front of the name.

9.5 Analysis overview

This iteration of the q^2-binned angular analysis of $B^0 \to K^{*0}\,(\to K^+\pi^-)\,\mu^+\mu^-$ will for the first time provide best fit values for P- and S-wave interference observables and CP-asymmetric observables, the analysis will also provide updated values for the branching fraction for this decay mode. This analysis is an updated version of a previous analysis that used a data set with

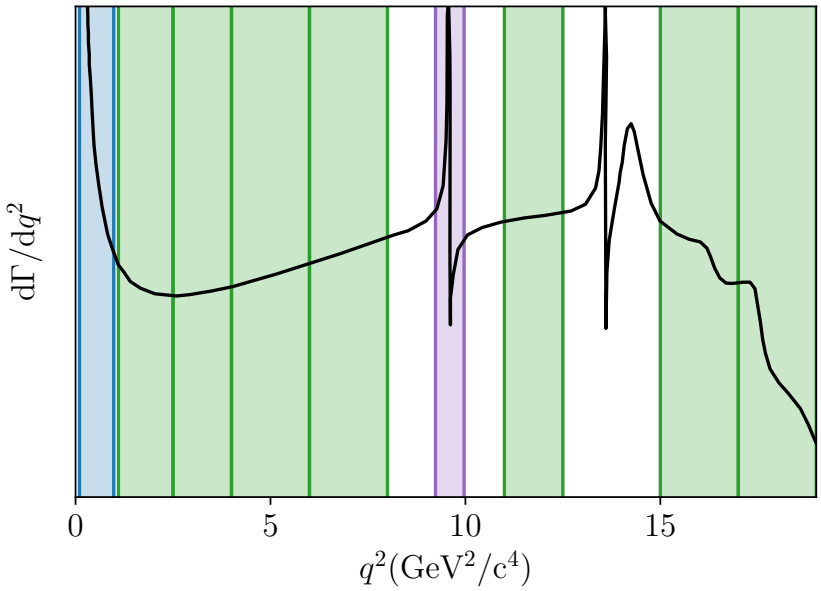


Figure 9.10: Differential decay rate of $B^0 \to K^{*0}\mu^+\mu^-$ [191]. The binning scheme used in this analysis is overlayed. In blue and green are the signal massive and massless lepton bins respectively, and in purple is the control mode window.

an integrated luminosity of 4.7fb^{-1} which itself super-seeded an analysis of 3fb^{-1}, presented in Refs. [8, 195]. The 4.7fb^{-1} analysis measured notable anomalies with respect to SM predictions. The measured angular observables were combined to produce best fit values of the Wilson coefficients C_9 and C_{10}. These results were consistent with a shift in $Re(C_9)$ away from the SM significant to 3.3 standard deviations. These strong hints of NP have highlighted the forthcoming update with the larger 9fb^{-1} data set as an analysis of particular interest.

Within the available data set the B^0 and $\bar{B}^0$ data are separated, discrimination is made using the kaon charge. These samples are further split into three sub-samples, split by data taking period. The splits made are 2011 + 2012 (Run 1), 2016, and 2017 + 2018. A single full five-dimensional fit is performed in each q^2-bin to all data where each data taking period has unique acceptance and background descriptions. The binning scheme used is depicted in Fig. 9.10.

In order to get the size of the background PDF relative to the signal PDF a simultaneous fit is performed to the m_{B^0} distribution, an example of this fit is presented in Fig. 9.11. The signal distribution is fit with two Crystal ball functions, and the background with a single exponential

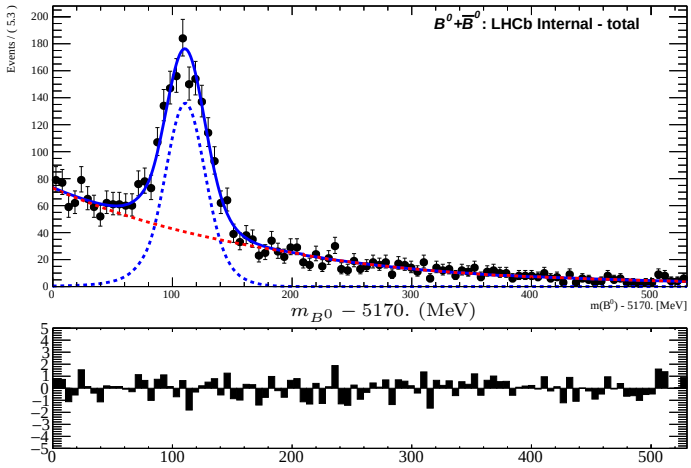


Figure 9.11: An example fit of a toy m_{B^0} distribution in a single rare mode bin. The dashed blue and red lines show the signal and background components respectively, the solid blue line is then the sum of these components.

function.

The background PDF is built up of 1$^{\text{st}}$ and 2$^{\text{nd}}$ order Legendre polynomials, $f(m_{B^0})$ and $f(x)$. These are all assumed to factorise and therefore the background PDF takes the following form,

$$f(m_{B^0}, \cos\theta_\ell, \cos\theta_K, \phi, m_{K\pi}) = f(m_{B^0}) \cdot f(\cos\theta_\ell) \cdot f(\cos\theta_K) \cdot f(\phi) \cdot f(m_{K\pi}).$$

It is assumed there is no CP asymmetry in the background and therefore B^0 and $\bar{B}^0$ share the same background PDF. The background PDF is then described with 9 free parameters for each data taking period. This was justified by demonstrating that this parameterisation successfully captures the angular distributions of the background dominating upper mass side band, see Refs. [194, 202].

The acceptance function is also constructed with Legendre polynomials L_a of order a,

$$\epsilon(\cos\theta_l, \cos\theta_K, \phi, q^2, m_{K\pi}) = \sum_{ijmnp} c_{ijmnp} L_i(\cos\theta_l) L_j(\cos\theta_K) L_m(\phi) L_n(q^2) L_p(m_{K\pi}). \quad (9.23)$$

This function accounts for any distortion of the angular distribution originating from angular acceptance, reconstruction and selection cuts. The efficiency function is fit to simulated signal events with varying orders, from the resulting fits toy samples are generated. It is desirable that the orders of the acceptance function are large enough to provide an accurate description but

not too large to allow for any overfitting to any statistical fluctuations in the simulated samples. Acceptance functions are obtained for B^0 and $\bar{B}^0$ separately on simulated signal samples.

A signal PDF is constructed for each data sub-sample. This is parameterised in terms of J_i and $\tilde{J}_i$ observables as in equations 9.8 and 9.13. These themselves are functions of S_i and A_i as,

$$J_i = S_i + A_i \qquad\qquad \bar{J}_i = S_i - A_i.$$

Values for S_i and A_i are common for across the signal PDFs of all sub-samples. The $m_{K\pi}$ propagators of the P-wave and S-wave lineshapes ($BW(m_{K\pi})$) are separately normalised (to $BW'(m_{K\pi})$) such that their integral over the $m_{K\pi}$ region being analysed is 1. That is,

$$BW'(m_{K\pi}) = \frac{BW(m_{K\pi})}{\sqrt{\int_{min(m_{K\pi})}^{max(m_{K\pi})} |BW(m_{K\pi})|^2 dm_{K\pi}}}, \tag{9.24}$$

such that

$$\int_{min(m_{K\pi})}^{max(m_{K\pi})} |BW'_P(m_{K\pi})|^2 dm_{K\pi} = 1, \qquad \int_{min(m_{K\pi})}^{max(m_{K\pi})} |BW'_S(m_{K\pi})|^2 dm_{K\pi} = 1. \tag{9.25}$$

This has the effect that the fitted values for the observables are independent of $m_{K\pi}$ window compared to Refs. [8, 195]. This analysis is carried out in a new wider $m_{K\pi}$ window. The previous analyses used $0.7959 < m_{K\pi} < 0.995$ GeV/c^2 which will be referred to as the narrow window, this analysis uses $0.745 < m_{K\pi} < 1.1$ GeV/c^2. The wider window extends into $m_{K\pi}$ regions that are dominated by background events, however the wider window should provide better control on the S-wave component of the fit.

A control mode is used to validate the fit, this is the $B^0 \to J/\psi K^{*0}$ channel where data is selected from within a window of ± 60 MeV around the J/ψ mass. A diagram of this decay is provided in Fig. 9.12. In this control mode fit a $B_s^0 \to J/\psi \bar{K}^{*0}$ component is added. This component is modelled with the same angular distribution and same shape in m_{B^0} as the $B^0 \to J/\psi K^{*0}$ channel but is shifted by the difference between $m_{B_s^0}$ and m_{B^0}. The width and mean parameters of the signal shape in m_{B^0} are floated along with the width and mean of the K^{*0} and its hadron radius d. Additionally the parameters of the LASS S-wave lineshape model are floated in this fit [196–201].

Then finally systematic errors are investigated, although not in this thesis, which cover the following sources: efficiency effects of the `L0` trigger, kinematic corrections to the MC used to fit

the acceptance function, tracking efficiencies in MC sample itself, the limited MC sample size, observables q^2 dependence across each bin, inclusion of peaking backgrounds, the exclusion of higher K^{*0} P- and D-wave states, exclusion of a $B_s^0 \to \bar{K}^{*0}\mu^+\mu^-$ component in the rare mode fit, parameters of the m_{B^0} mass model, any tracking asymmetries, the modelling of the S-wave $m_{K\pi}$ model and biases in the fit. These systematics are estimated with toys.

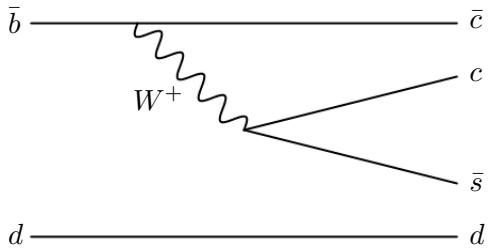


Figure 9.12: A tree-level Feynman diagram contributing to $B^0 \to J/\psi K^{*0}$.

9.6 Fitting asymmetry observables

As mentioned the data is split by run period and $\mathcal{CP}$ state into six sub-samples. To fit the data a negative log likelihood is defined for each sub-sample $\mathcal{NLL}^i$. These are then normalised simultaneously with the values for the observables being shared across sub-samples. The parameterisation of the acceptance and the fit to the background is handled separately for each sub-sample.

The default fit set up to obtain best fit values for all CP-averaged observables involves normalising the PDF such that $\Gamma_P + \bar{\Gamma}_P + \Gamma_S + \bar{\Gamma}_S = 1$, setting all CP-asymmetric terms to 0 and employing equation 9.21 to reduce the number of observables. This approach is not sufficient for a fit that floats CP-asymmetric terms. A relation equivalent to equation 9.21 cannot be obtained for the CP-asymmetric terms, and there are therefore too many fit parameters compared to the number of orthogonal terms in the angular distribution. The only way to fit all the CP-asymmetric terms is by measuring the total CP-asymmetry manifesting in the measured yields. This can be achieved via an extended maximum likelihood fit ($S+A$ fit). The components of the negative log likelihood of each of the six sub-samples, each of three run periods split by $\mathcal{CP}$ state, are adapted as follows.

For the standard non-extended fit (S fit), for each $\mathcal{CP}$ state and run period combination i, the signal (PDF_{sig}) and background (PDF_{bkg}) PDFs are combined as,

$$PDF^i = f^i_{sig} \times PDF^i_{sig} + (1 - f^i_{sig}) \times PDF^i_{bkg}, \tag{9.26}$$

where f^i_{sig} is the signal fraction. This is then combined with a Poisson extended term P^i that will constrain the total yield.

$$PDF^i \times P^i = (f^i_{sig} \times PDF^i_{sig} + (1 - f^i_{sig}) \times PDF^i_{bkg}) \times P^i. \tag{9.27}$$

The form of P^i is as follows,

$$P^i = \frac{(N^i_{bkg} + N^i_{sig})^{N^i_D} e^{-(N^i_{bkg}+N^i_{sig})}}{N^i_D!}, \tag{9.28}$$

where N^i_{bkg} and N^i_{sig} are the expected number of background and signal events in the sub-sample respectively, and N^i_D is the total measured number of events in the sub-sample. The value of N^i_{bkg} is fit as a nuisance parameter. The value of N^i_{sig} is estimated as follows,

$$N^i_{sig} = \int \mathcal{L}^i \,\mathrm{d}t \times 2\sigma_{b\bar{b}} \times \epsilon^i_{sig}(\Omega, m_{K\pi}) \times \mathcal{B}(b \to B^0/\bar{B}^0) \times \mathcal{B}(B^0 \to K^{*0}\mu^+\mu^-), \tag{9.29}$$

where $\int \mathcal{L}^i \,\mathrm{d}t$ is the integrated luminosity of the relevant run period, $\sigma_{b\bar{b}}$ is the $b\bar{b}$ production fraction, ϵ^i_{sig} the acceptance and efficiency of the detector, $\mathcal{B}(b \to B^0/\bar{B}^0)$ the production fraction of $B^0/\bar{B}^0$ mesons, and $\mathcal{B}(B^0 \to K^{*0}\mu^+\mu^-)$ is the CP-averaged branching fraction of the signal channel. This can then be expressed in terms of the total CP-averaged width of the $B^0/\bar{B}^0$ meson Γ_B, and the integrated angular distribution which includes the detector acceptance,

$$N^i_{sig} = \int \mathcal{L}^i \,\mathrm{d}t \times 2\sigma_{b\bar{b}} \times \mathcal{B}(b \to B^0/\bar{B}^0) \times \frac{\int \epsilon^i_{sig}(\Omega, m_{K\pi}) \frac{\mathrm{d}\Gamma}{\mathrm{d}\Omega\,\mathrm{d}m_{K\pi}} \,\mathrm{d}\Omega\,\mathrm{d}m_{K\pi}}{\Gamma_B}. \tag{9.30}$$

To simplify this in the fit we can fit relative to the control mode. So similarly for the $B^0 \to J/\psi K^{*0}$ channel,

$$N^i_{J/\psi} = \int \mathcal{L}^i \,\mathrm{d}t \times 2\sigma_{b\bar{b}} \times \mathcal{B}(b \to B^0/\bar{B}^0) \times \frac{\int \epsilon^i_{J/\psi}(\Omega, m_{K\pi}) \left.\frac{\mathrm{d}\Gamma}{\mathrm{d}\Omega\,\mathrm{d}m_{K\pi}} \,\mathrm{d}\Omega\,\mathrm{d}m_{K\pi}\right|_{J/\psi}}{\Gamma_B}. \tag{9.31}$$

Then combining the above to cancel $\int \mathcal{L}^i \,\mathrm{d}t \times 2\sigma_{b\bar{b}} \times \mathcal{B}(b \to B^0/\bar{B}^0)$ gives,

$$N^i_{sig} = N^i_{J/\psi} \times \frac{\int \epsilon^i_{sig}(\Omega, m_{K\pi}) \frac{\mathrm{d}\Gamma}{\mathrm{d}\Omega\,\mathrm{d}m_{K\pi}} \,\mathrm{d}\Omega\,\mathrm{d}m_{K\pi}}{\int \epsilon^i_{J/\psi}(\Omega, m_{K\pi}) \left.\frac{\mathrm{d}\Gamma}{\mathrm{d}\Omega\,\mathrm{d}m_{K\pi}} \,\mathrm{d}\Omega\,\mathrm{d}m_{K\pi}\right|_{J/\psi}} \times R_{BF} = N^i_{J/\psi} \times \frac{\mathcal{I}^i_{sig}}{\mathcal{I}^i_{J/\psi}} \times R_{BF}, \tag{9.32}$$

where R_{BF} is the relative branching fraction equal to ratio of CP-averaged branching fractions,

$$R_{BF} = \frac{\mathcal{B}(B^0 \to K^{*0}\mu^+\mu^-)}{\mathcal{B}(B^0 \to J/\psi K^{*0})}. \tag{9.33}$$

For each sub-sample the values for $N_{J/\psi}$ and the integral of the control mode angular distribution, $\mathcal{I}^i_{sig}$, can be obtained in a one-off external fit to $B^0 \to J/\psi K^{*0}$ data. During this external fit, and the rare mode fit, the angular distributions are normalised. The relative branching fraction accounts for this normalisation. In the final rare mode fit R_{BF} and N^i_{bkg} are floated. Measured R_{BF} values can later be combined with external measurements of the $B^0 \to J/\psi K^{*0}$ branching fraction to obtain a useful measurement of the rare mode branching fraction. The total negative log likelihood to be minimised for each $\mathcal{CP}$ state and run period combination is then,

$$\mathcal{NLL}^i = -log(P^i) + \sum_{j=0}^{N^i_D} -log(PDF^i_j). \tag{9.34}$$

$$\mathcal{NLL}^i = -(N^i_D \times log(N^i_{bkg} + N^i_{sig}) - (N^i_{bkg} + N^i_{sig})) + \sum_{j=0}^{N^i_D} -log(PDF^i_j). \tag{9.35}$$

In the non-extended fit f^i_{sig} was floated, now in the extended fit f^i_{sig} is reconstructed from N^i_{sig} and N^i_{bkg}. Then for the full fit the $\mathcal{NLL}^i$ terms are all minimised simultaneously.

9.6.1 External $B^0 \to J/\psi K^{*0}$ fit

Sub-sample i	$N^i_{J/\psi}$	$\mathcal{I}^i_{sig}$
Run 1 B^0	188,463	1.072495
Run 1 $\bar{B}^0$	183,035	1.072883
2016 B^0	166,410	1.081046
2016 $\bar{B}^0$	161,718	1.089758
2017+2018 B^0	346,394	1.079925
2017+2018 $\bar{B}^0$	337,625	1.083851

Table 9.2: Values of $\mathcal{I}^i_{sig}$, as in equation 9.32, and $N^i_{J/\psi}$ obtained from the external fit.

An external non-extended S fit is performed on the control mode $B^0 \to J/\psi K^{*0}$ data. From this values for $\mathcal{I}_{sig}$ and $N_{J/\psi}$ are obtained, these are presented in Table 9.2. In the external fit $A_{CP} = 0$ is set, as for the $B^0 \to J/\psi K^{*0}$ channel there should be no angular asymmetries beyond the detector asymmetry provided by the acceptance function and B^0

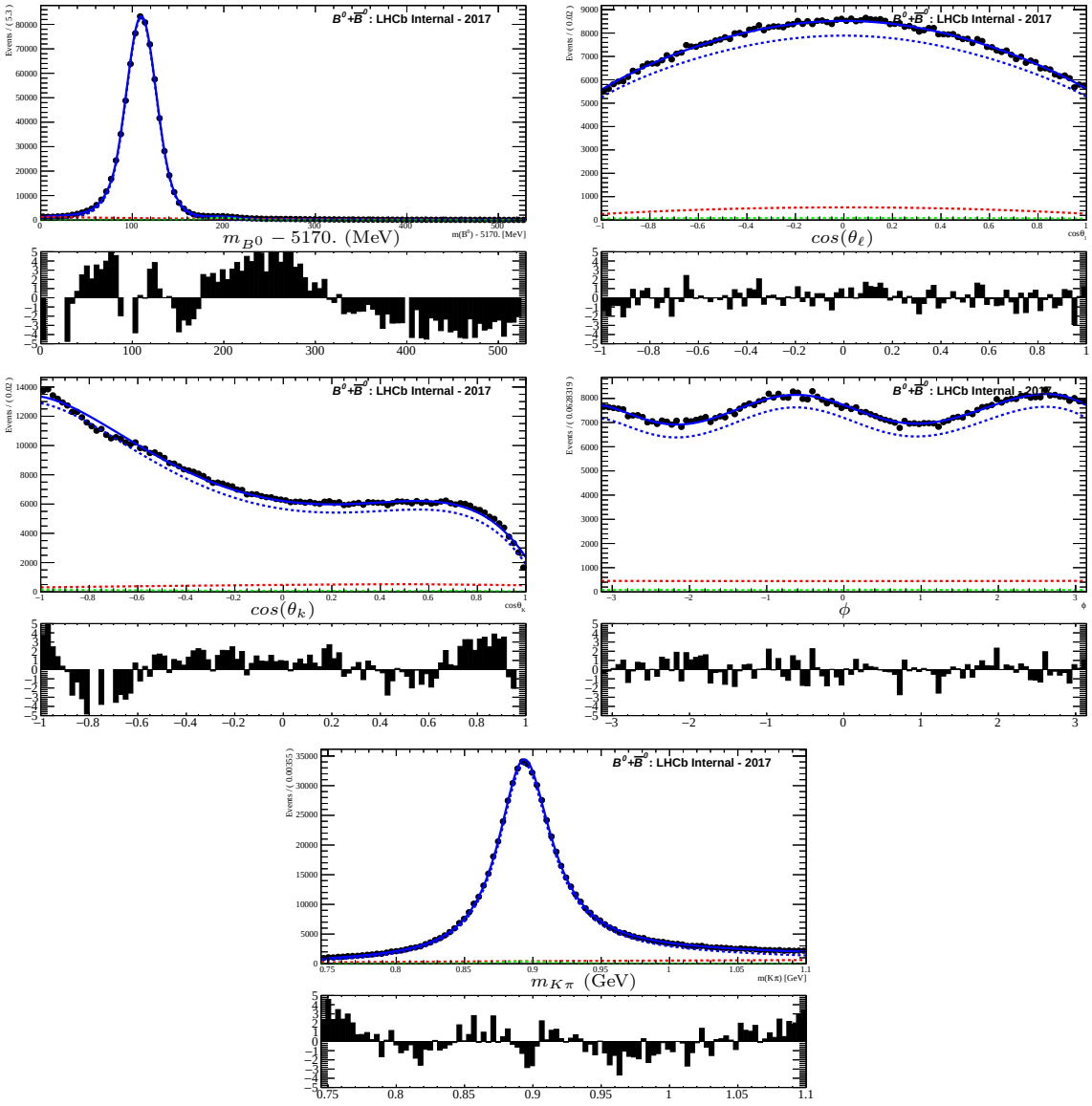


Figure 9.13: Projections of the 2017+2018 control mode data and external fit model. The dashed blue and red lines show the signal and background components respectively, the green dashed line is the B_s component, the solid blue line is then the total fit.

production asymmetry reflected in the measured yields. This was separately shown to be consistent, every CP-asymmetric parameter was consistent with 0 to within 2σ.

Projections of the 2017+2018 fit are provided in Fig. 9.13. The full fit is plotted with a solid blue line, the blue, red and green dashed lines indicate contributions from signal, background and $B_s^0 \to J/\psi K^{*0}$ respectively. A significant deviation is observed in the $cos(\theta_k)$ projection. This is understood to be caused by the contribution from the exotic resonance $Z(4430)$, a model

of which is not included in this fit [197].

With values for $\mathcal{I}^i_{sig}$ and $N^i_{J/\psi}$ obtained, the extended fit can now be tested. An initial test is performed by fitting back to the same J/ψ control mode data. The fit converges to a result with R_{BF} consistent with 1, therefore it appears from this preliminary check that the fitting procedure is working correctly.

9.6.2 SM toys

The extended $S + A$ fit can now be tested with toys using SM values for the CP-averaged observables and and CP-asymmetric observables. For the toys, the parameters of the background shape from a fit to the control mode data are used to generate background events. The background yields for each sub-sample are obtained for each bin by scaling the Run 1 yields obtained in the previous analysis, presented in Table 9.3, by the relative integrated luminosity [194]. These yields are scaled again to estimate the new yields expected for the wider $m_{K\pi}$ window, this is achieved using the ratio of integrals of the $m_{K\pi}$ distributions.

q^2 bin (GeV/c^2)^2	Run 1 N_{sig}	Run 1 N_{bkg}	2016 N_{sig}	2016 N_{bkg}
[0.1, 0.98]	337.7 ± 19.6	57.8 ± 10.2	289.6 ± 18.2	59.9 ± 10.1
[1.1, 2.5]	179.4 ± 15.4	120.1 ± 13.3	185.1 ± 15.7	113.4 ± 13.2
[2.5, 4.0]	164.6 ± 15.8	200.9 ± 16.9	166.7 ± 15.1	126.8 ± 13.7
[4.0, 6.0]	278.6 ± 20.1	290.9 ± 20.4	244.9 ± 18.0	161.6 ± 15.5
[6.0, 8.0]	343.6 ± 22.1	337.9 ± 21.9	339.3 ± 21.3	189.1 ± 17.4
[11.0, 12.5]	329.7 ± 21.0	206.8 ± 17.8	296.9 ± 19.5	130.6 ± 14.6
[15.0, 17.0]	448.3 ± 23.8	185.2 ± 17.4	408.8 ± 22.3	120.7 ± 14.5
[17.0, 19.0]	300.1 ± 19.8	139.4 ± 15.2	255.6 ± 17.6	71.9 ± 11.2
Total	2392.7 ± 56.5	1578.8 ± 48.7	2225.3 ± 53.3	979.2 ± 39.9

Table 9.3: Values of N_{sig} and N_{bkg} obtained from previous analysis [194].

The signal yields are obtained using equation 9.32 and estimated values of R_{BF}. For each q^2 bin, R_{BF} is calculated via the following formulation. Starting with a measurement of the CP-averaged P-wave component of $B^0 \to J/\psi K^{*0}$, $\mathcal{B}(B^0 \to J/\psi\ K^*(892)^0) = 1.19 \times 10^{-3}$ [203]. This is translated into a branching fraction for the P-wave component of the control mode by multiplying by the branching fraction of $J/\psi \to \mu^+\mu^-$, $\mathcal{B}_{J/\psi\to\mu\mu} = 0.06$ [204]. The S-wave component of the $B^0 \to J/\psi K^{*0}$ channel is then included using values for $F_S(J/\psi) = 0.08$ [205]. This information is combined as

$$\mathcal{B}(B^0 \to J/\psi(\to \mu^+\mu^-)K^{*0}) = \mathcal{B}(B^0 \to J/\psi K^*(892)^0) \times (1 + F_S) \times \mathcal{B}_{J/\psi\to\mu\mu}, \tag{9.36}$$

this provides a CP-averaged estimate for $\mathcal{B}(B^0 \to J/\psi K^{*0})$ in equation 9.33. For each q^2 bin the CP-averaged rare mode branching fraction $\mathcal{B}(B^0 \to K^{*0}\mu^+\mu^-)$ is calculated using P-wave differential branching fraction measurements, $d\mathcal{B}_P/\,\mathrm{d}q^2$, and S-wave fractions, F_S, from Ref. [194]. Values for these quantities used and final R_{BF} values obtained for each q^2 bin are presented in Table. 9.4.

q^2 bin $(\mathrm{GeV/c^2})^2$	$d\mathcal{B}_P/\,\mathrm{d}q^2 \times 10^{-7}$ $(\mathrm{GeV/c^2})^{-2}$	F_S	$R_{BF} \times 10^{-3}$
[0.1, 0.98]	1.02	0.02	1.19
[1.1, 2.5]	0.33	0.14	0.68
[2.5, 4.0]	0.33	0.03	0.66
[4.0, 6.0]	0.35	0.12	1.02
[6.0, 8.0]	0.43	0.03	1.15
[11.0, 12.5]	0.49	0.02	0.97
[15.0, 17.0]	0.53	0.00	1.37
[17.0, 19.0]	0.36	0.02	0.95

Table 9.4: Values used to calculated R_{BF} for the narrow $m_{K\pi}$ window.

The $d\mathcal{B}_P/\,\mathrm{d}q^2$ and F_S values in Table 9.4 are measured in the narrow $m_{K\pi}$ window. The default for this analysis is to perform the extended fit in the wide window. This miss-match must be addressed to correctly calculate N_{sig} for the toy generation. Therefore, using the control mode fit model, $N_{J/\psi}$ is scaled to the expected value for the narrow window using the ratio of $m_{K\pi}$ lineshape integrals over the required intervals. Using this and values from Table 9.4 we can provide an estimated value for N_{sig} for the narrow window. This is then similarly scaled up using the ratio of the same integrals of the rare mode model.

q^2 bin $(\mathrm{GeV/c^2})^2$	Run 1 N_{sig}	Run 1 N_{bkg}	2016 N_{sig}	2016 N_{bkg}	2017+2018 N_{sig}	2017+2018 N_{bkg}
[0.1, 0.98]	441.473	106.264	362.915	113.348	758.453	262.117
[1.1, 2.5]	201.498	220.806	173.509	235.526	362.616	544.655
[2.5, 4.0]	195.479	369.486	168.671	394.119	380.769	911.399
[4.0, 6.0]	316.72	535.135	276.085	570.811	576.989	1320.0
[6.0, 8.0]	389.667	621.773	342.24	663.224	715.246	1533.71
[11.0, 12.5]	373.056	380.77	319.198	406.155	667.089	939.233
[15.0, 17.0]	506.787	341.676	405.689	364.454	847.846	842.8
[17.0, 19.0]	326.063	257.5	249.709	274.667	521.864	635.167
Total	2750.743	2833.41	2298.016	3022.304	4830.872	6989.081

Table 9.5: Values of N_{sig} and N_{bkg} used to generate toy events in the wide $m_{K\pi}$ window.

For the generation of each toy data set N_{sig} and N_{bkg} values are Poisson fluctuated around the values presented in Table 9.5. Some projections of example rare mode toy generated data sets and corresponding extended $S + A$ fits are presented in Fig. 9.14 and Fig. 9.15 for the bins

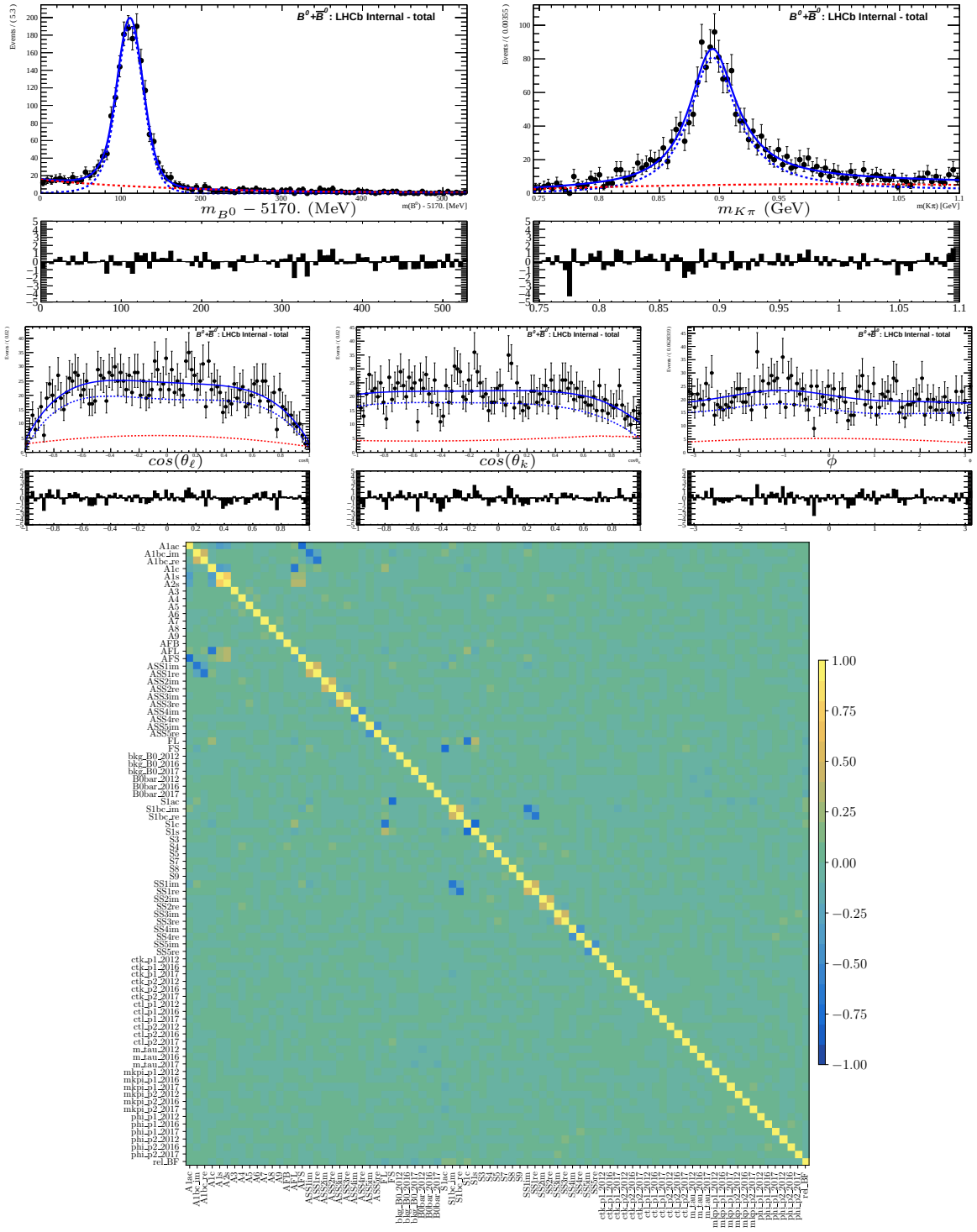


Figure 9.14: Example toy data set, extended $S + A$ fit and correlations between observables for bin 0 which has $0.1 < q^2 < 0.98$ $(\mathrm{GeV/c^2})^2$. The dashed blue and red lines show the signal and background components respectively, the solid blue line is then the sum of these components.

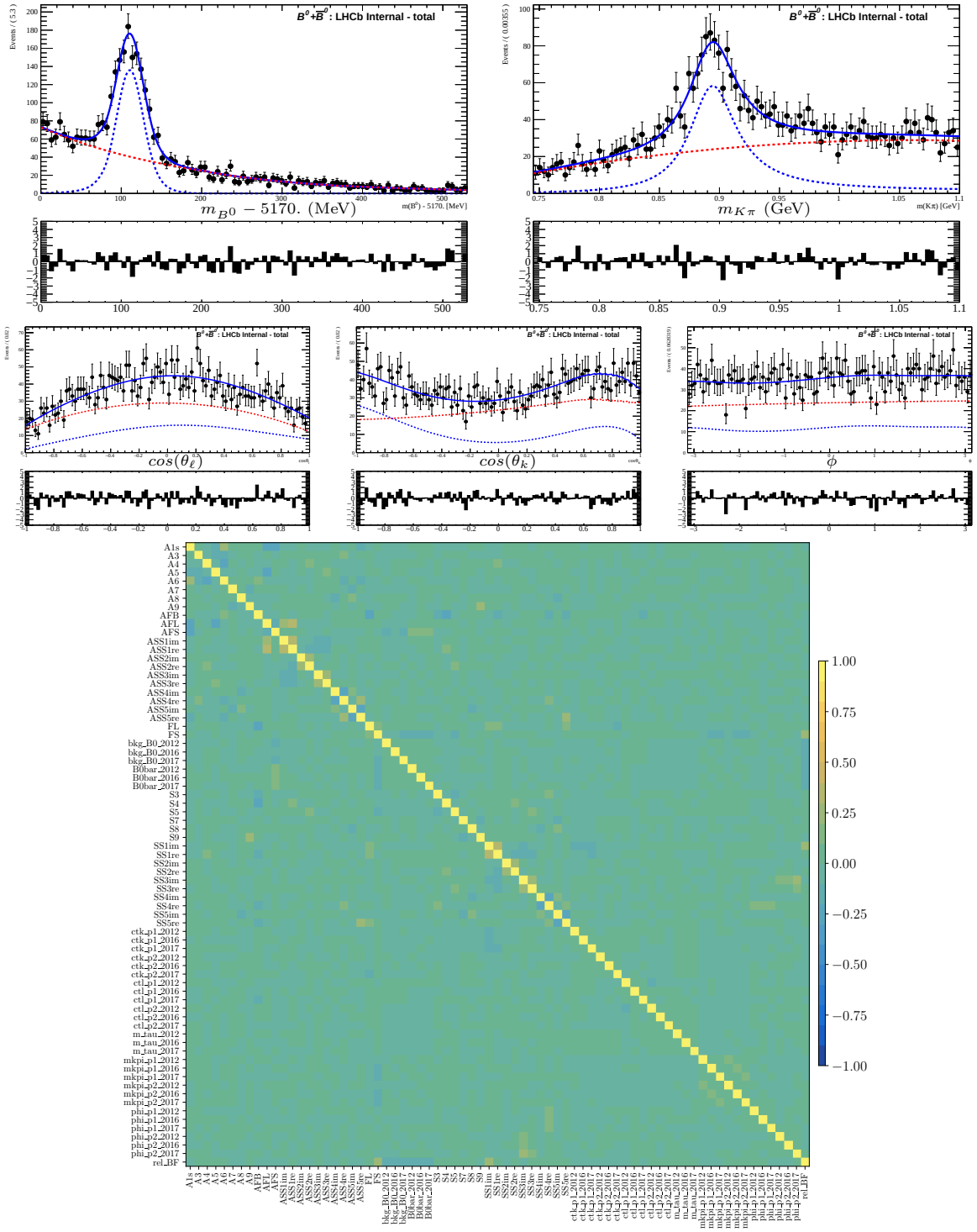


Figure 9.15: Example toy data set, extended $S + A$ fit and correlations between observables for bin 0 which has $4.0 < q^2 < 6.0\ (\mathrm{GeV/c^2})^2$. The dashed blue and red lines show the signal and background components respectively, the solid blue line is then the sum of these components.

of $0.1 < q^2 < 0.98$ $(\mathrm{GeV/c^2})^2$ (massive leptons) and $4.0 < q^2 < 6.0$ $(\mathrm{GeV/c^2})^2$ (massless leptons) respectively. The projections look good for both bins shown, the behaviour of the other bins is unchanged. The correlations between observables in the massless bin all appear well behaved. This is however not true for the massive bin where there exists large anti-correlations between the additional observables and significant correlations between some of the interference observables. This behaviour is understood to originate from interactions with the physical boundary of the angular distribution. This effect will be overcome in the fit final by computing statistical intervals of the observables using the Feldman-Cousins method [206].

For each q^2 bin 2500 toys are run for the extended $S + A$ fit in the wide $m_{K\pi}$ window. For each observable λ_i in each fit the pull is calculated as $(\lambda_i^{\mathrm{fit}} - \lambda_i^{\mathrm{toy}})/\sigma(\lambda_i^{\mathrm{fit}})$, where λ_i^{fit} is the fit result and λ_i^{toy} is the observable value used to generate the toy data and $\sigma(\lambda_i^{\mathrm{fit}})$ is the error from the Hessian matrix. The pull distributions for each q^2 bin are presented in green in Figs. 9.16-9.23. For an unbiased fit whose Hessian errors have the correct statistical coverage, the pull distributions should have a mean of 0 and a width of 1. Therefore, overlayed on each pull plot is a unit normal distribution with these properties to be used as a comparison. The right panels of Figs. 9.16-9.23 present a summary of the pull study results.

The first thing to note is well behaved pull distributions for all CP-asymmetric observables across all bins. The fit results for R_{BF} are also well behaved, the bias here is $\sim 10\%$ of the statistical error at worse. There are however significant biases in some of the CP-averaged observables, including S_4, S_5 and F_L which are three observables of particular theoretical interest. To confirm that there is nothing wrong with the fitter these pull studies are repeated with toys with a factor 10 more events per toy, these results are presented in purple in Figs. 9.16-9.23. These higher statistic toys show largely improved behaviour with the exception of bin 0 for which some observables still show large biases, still originating from the PDF being close to the physical boundary. The standard fit to only the CP-averaged observables (the S fit), that acted as the nominal before the development of the $S + A$ extended fit, provides better behaviour for the CP-averaged observables. The S fit results are presented in orange in Figs. 9.16-9.23.

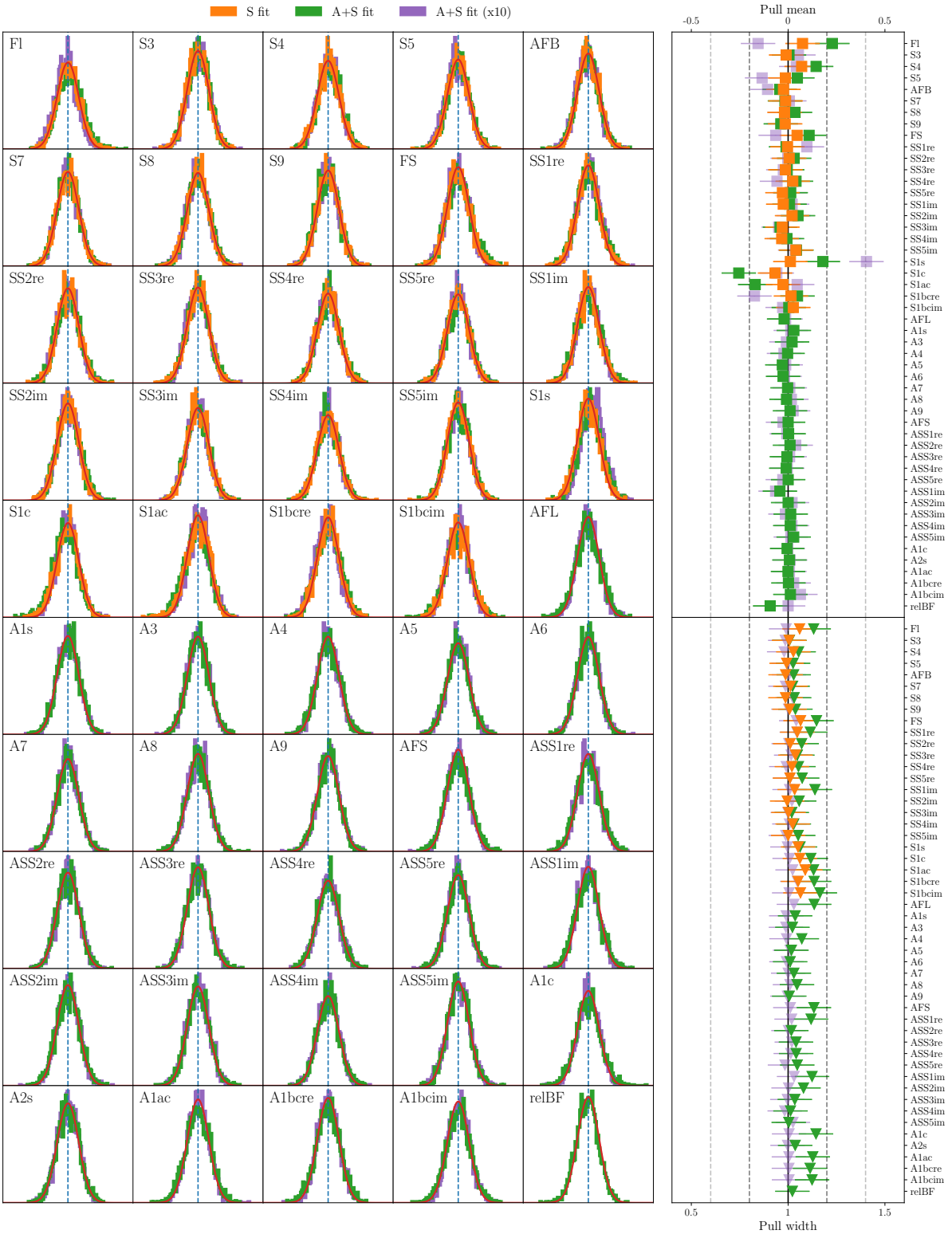


Figure 9.16: Pull results for SM toy studies in bin of $0.1 < q^2 < 0.98$ $(\mathrm{GeV/c^2})^2$.

9.6.3 Conclusions from toy studies

With the expected data from an integrated luminosity of 9fb^{-1} it is expected that a simultaneous fit of both CP-averaged and CP-asymmetric observables will result in sizeable biases. Toy studies show that this will be possible in the future with larger data sets, at least for the bins where the massless lepton assumption can be made. For the 9fb^{-1} result the plan therefore is to obtain the full set of observables in two separate fits. One to determine the CP-averaged observables and another to determine the CP-asymmetric observables. Relative to the previous analysis of 4.7fb^{-1} of LHCb data the data set used in this analysis is ~ 2 times the size. As the results will still be statistically dominated it is expected that the uncertainties obtained will be $\sim \sqrt{2}$ times smaller.

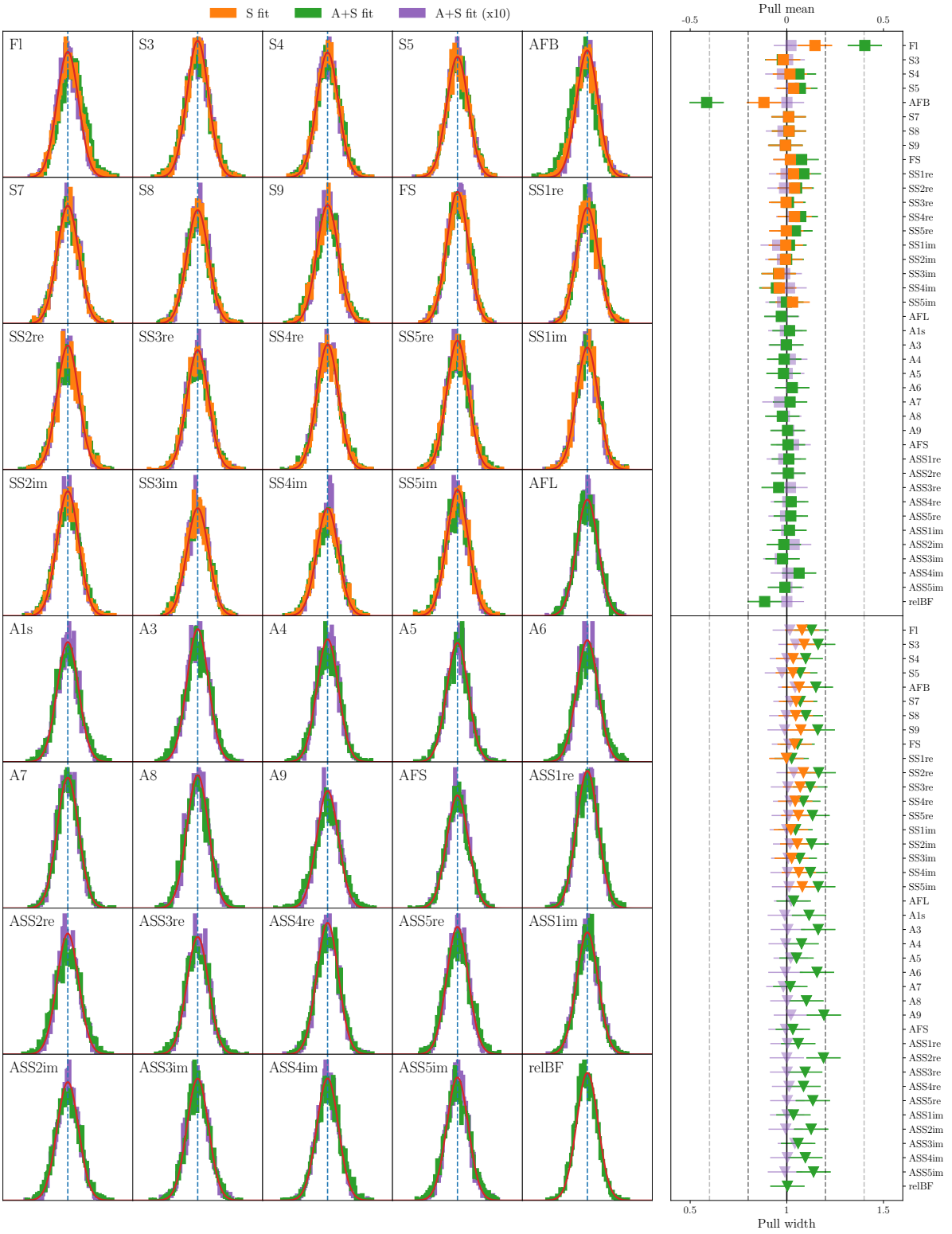


Figure 9.17: Pull results for SM toy studies in bin of $1.1 < q^2 < 2.5$ $(\mathrm{GeV/c^2})^2$.

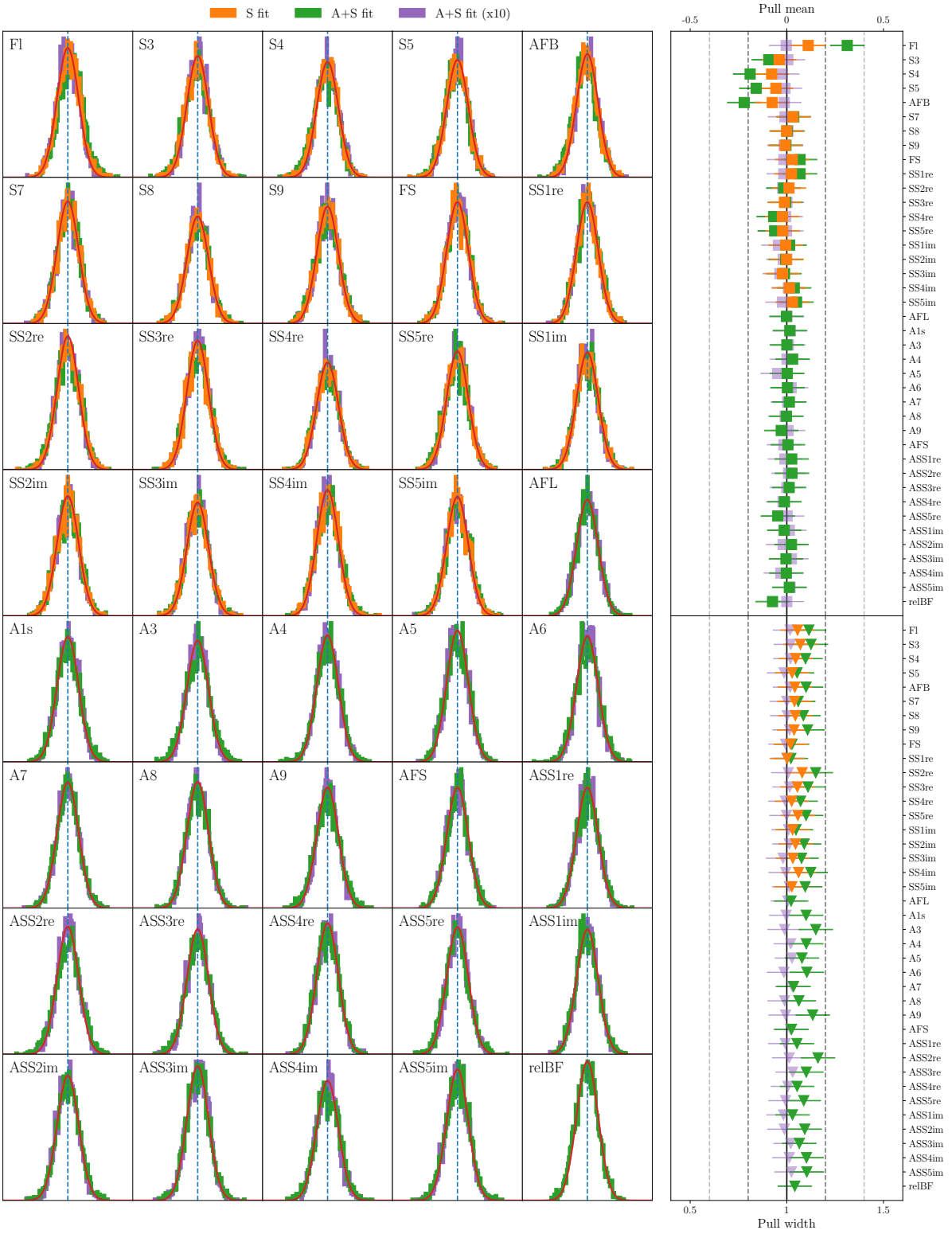


Figure 9.18: Pull results for SM toy studies in bin of $2.5 < q^2 < 4.0$ $(\mathrm{GeV/c^2})^2$.

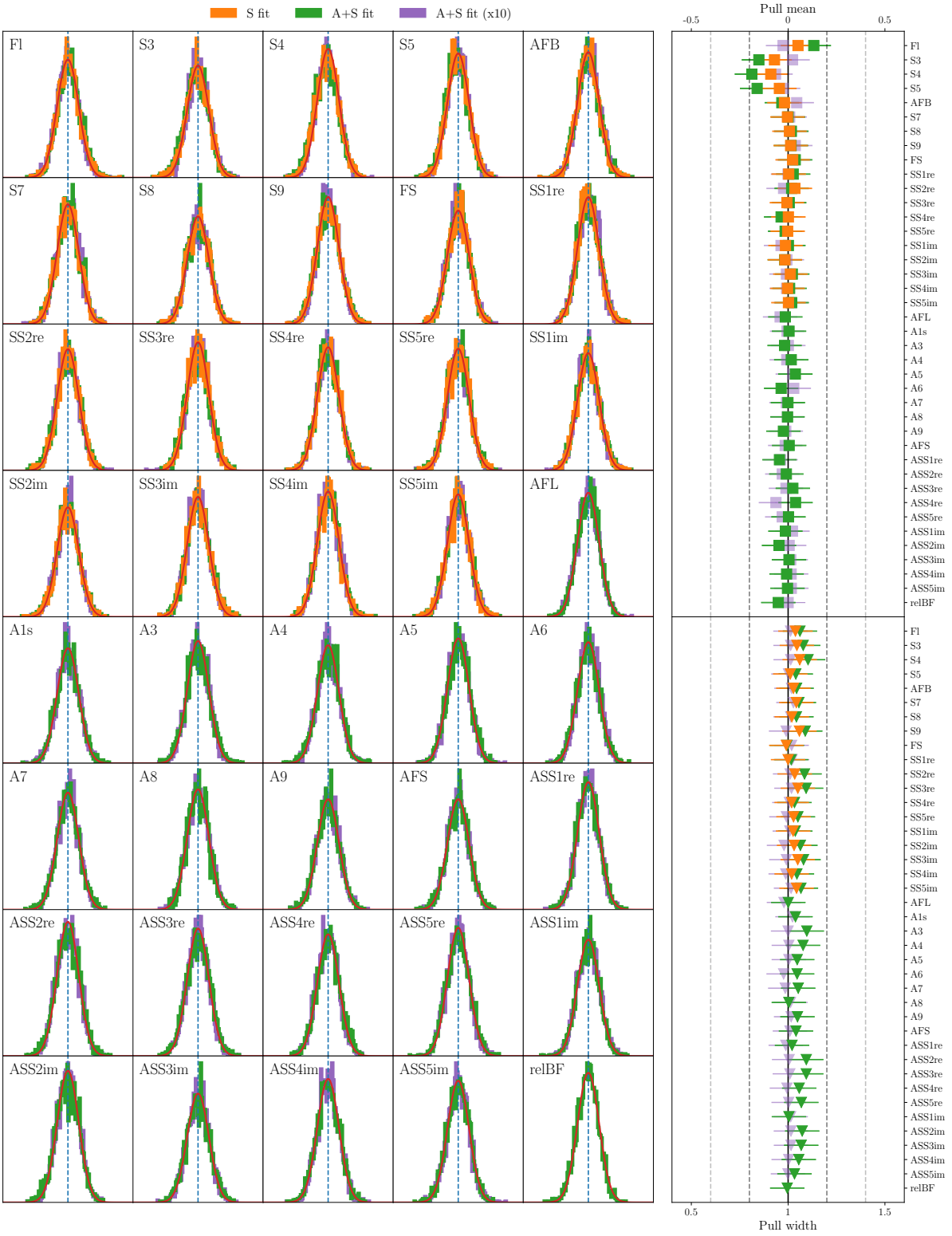


Figure 9.19: Pull results for SM toy studies in bin of $4.0 < q^2 < 6.0$ $(\mathrm{GeV/c^2})^2$.

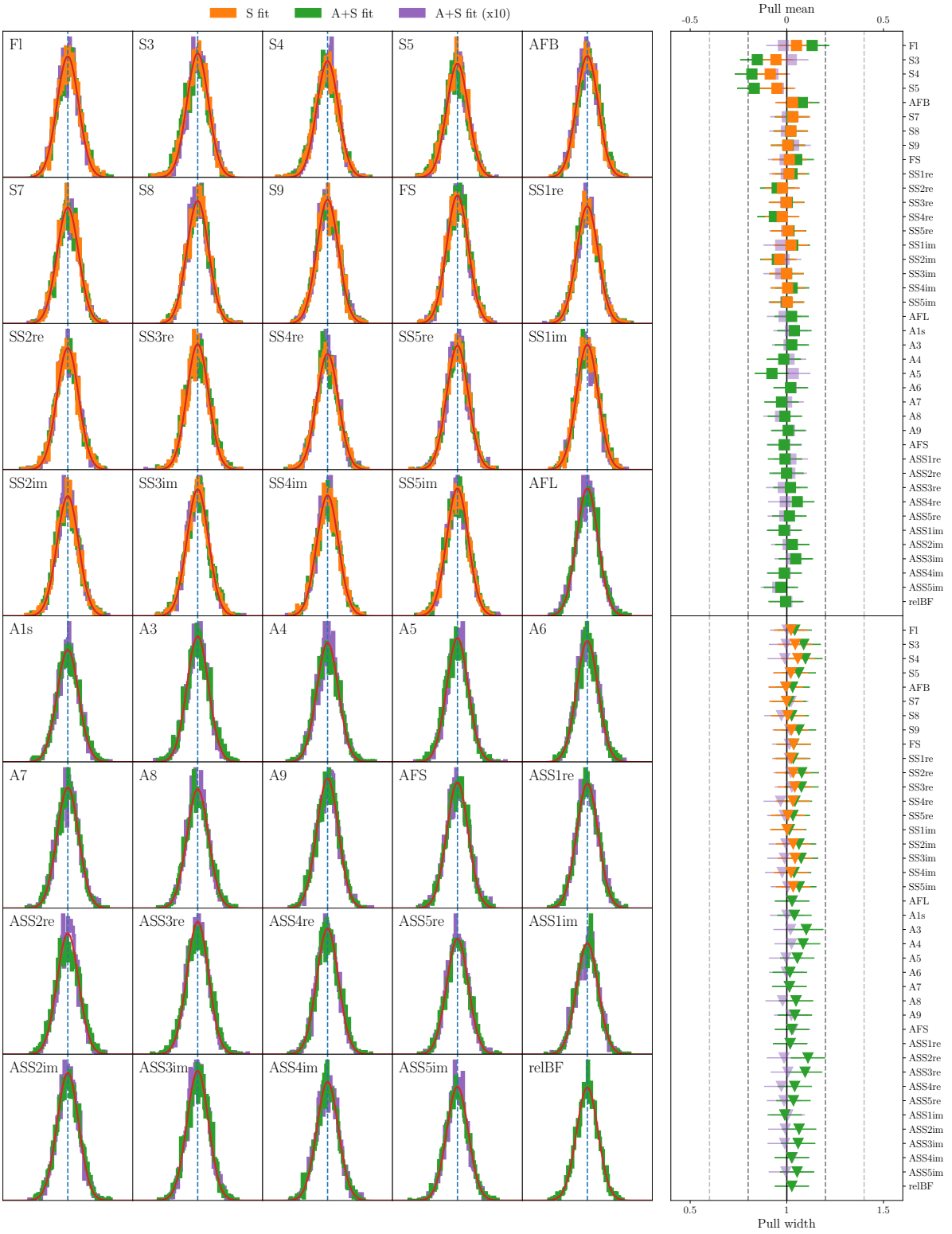


Figure 9.20: Pull results for SM toy studies in bin of $6.0 < q^2 < 8.0$ $(\mathrm{GeV/c^2})^2$.

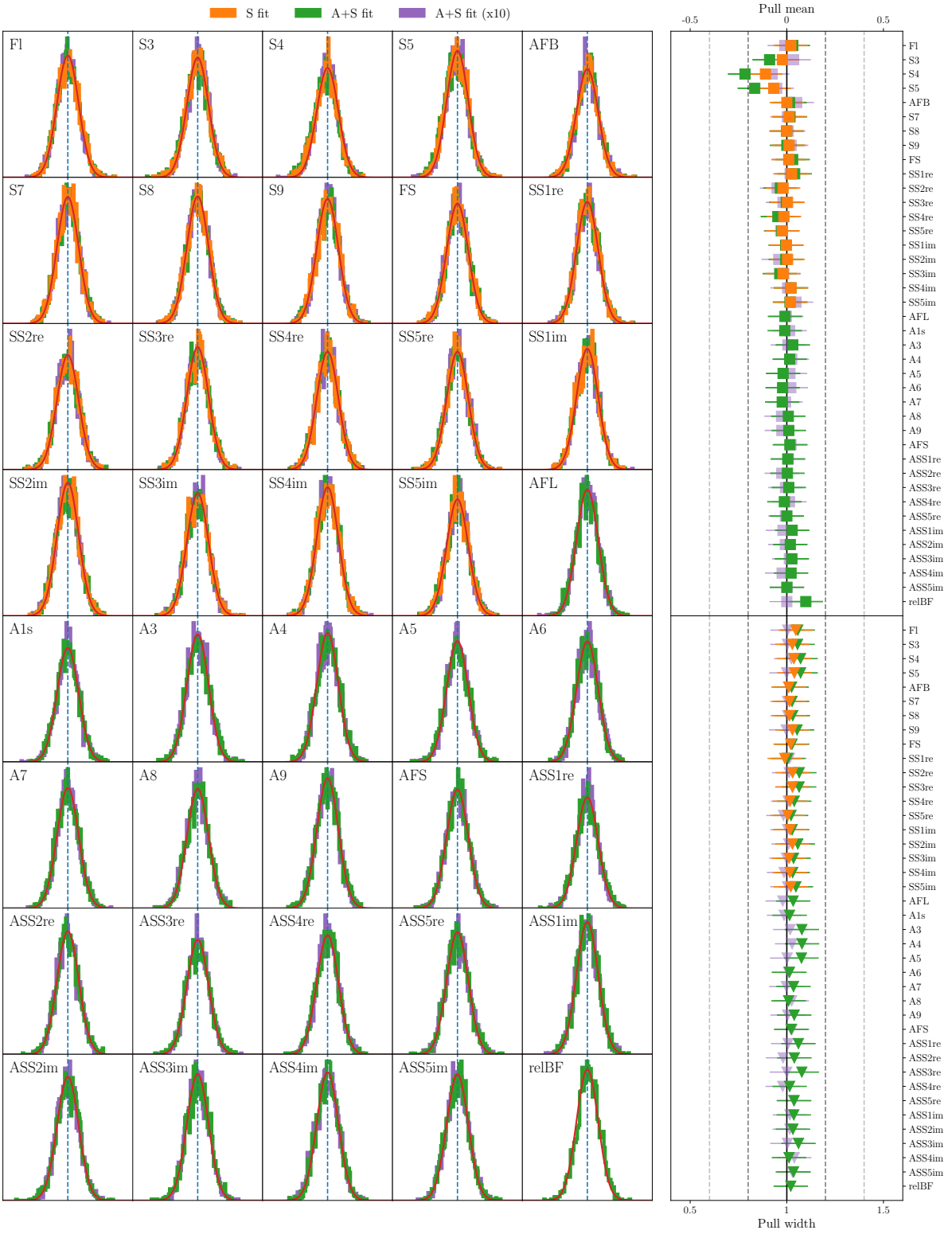


Figure 9.21: Pull results for SM toy studies in bin of $11.0 < q^2 < 12.5$ $(\mathrm{GeV/c^2})^2$.

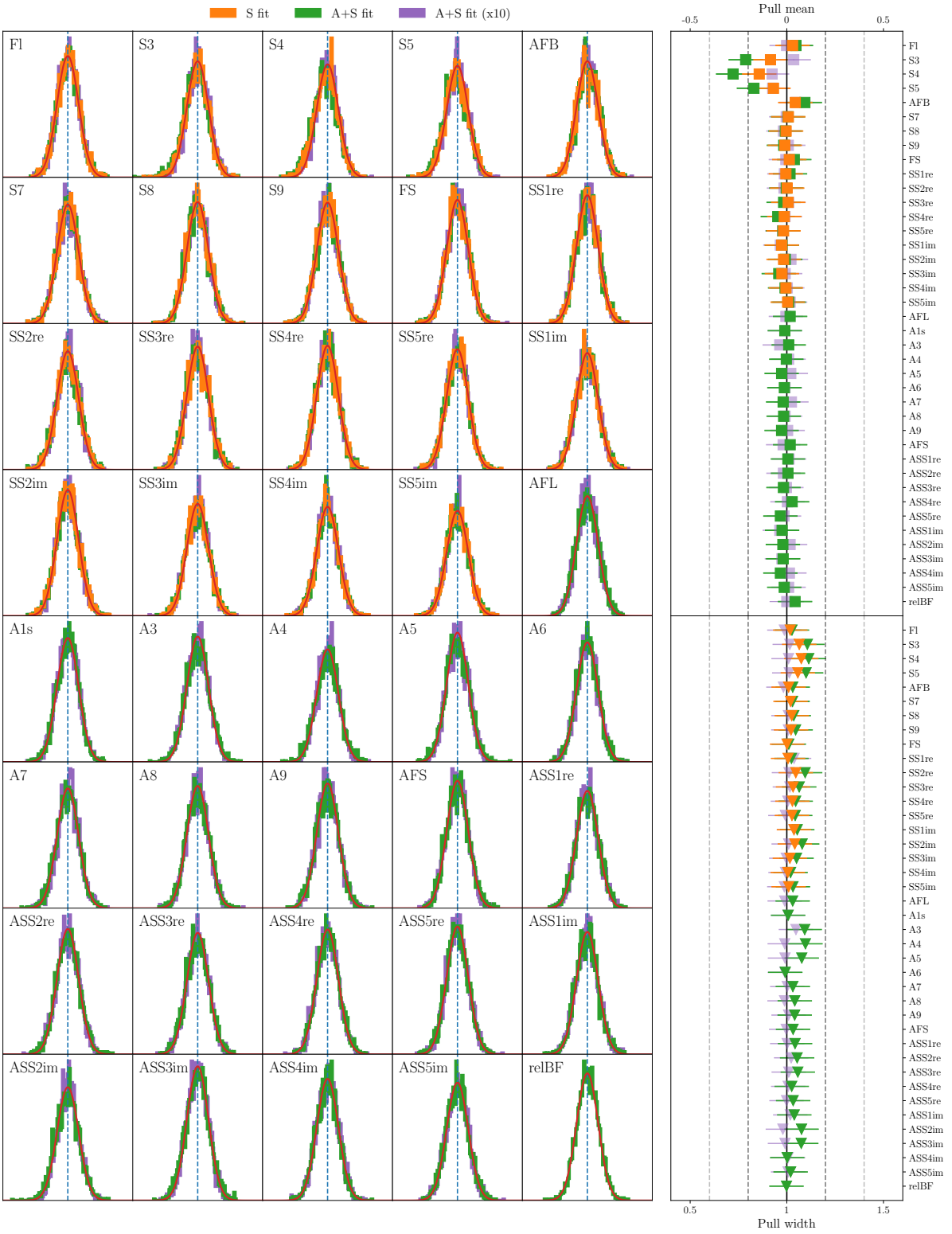


Figure 9.22: Pull results for SM toy studies in bin of $15.0 < q^2 < 17.0$ $(\mathrm{GeV/c^2})^2$.

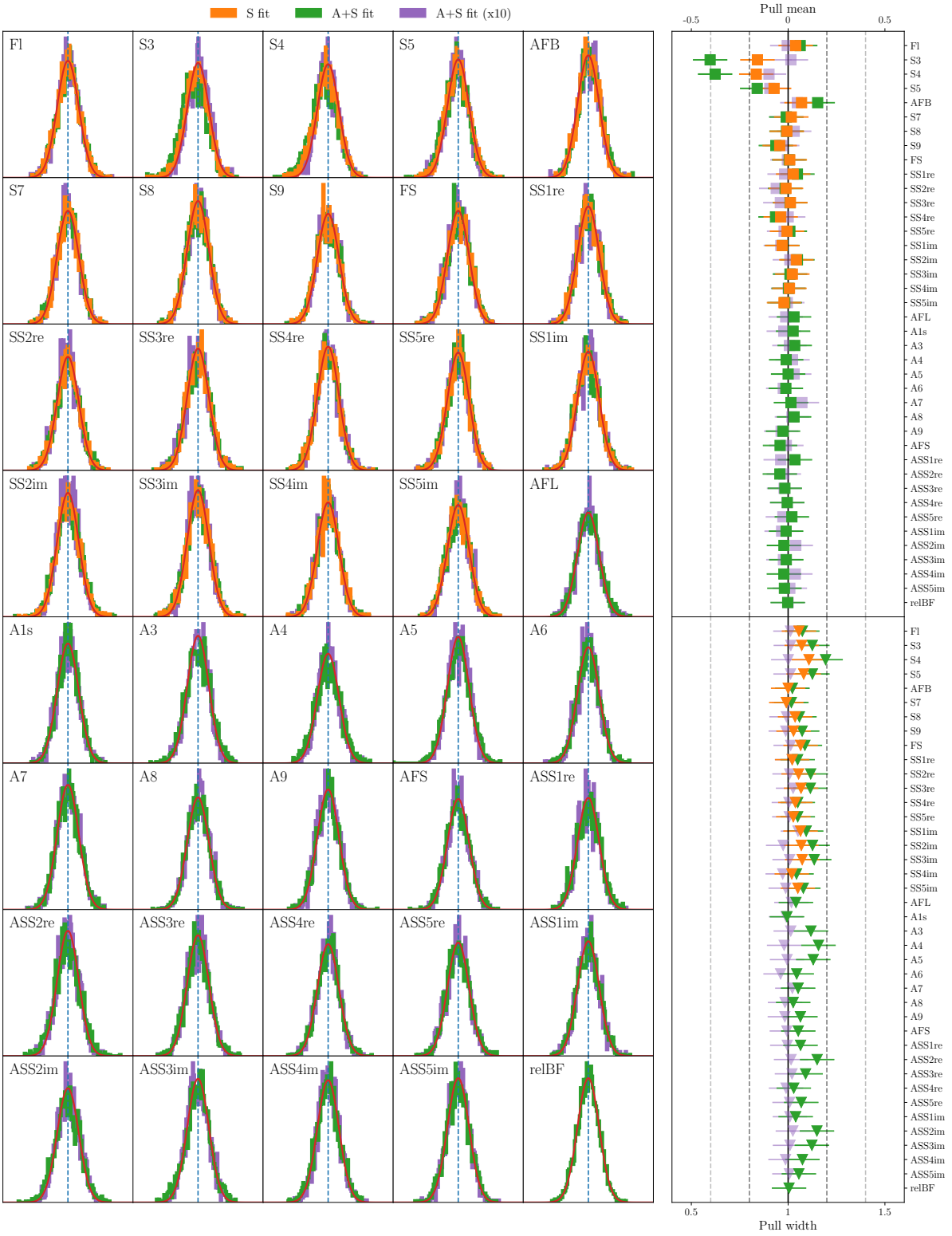


Figure 9.23: Pull results for SM toy studies in bin of $17.0 < q^2 < 19.0$ $(\mathrm{GeV/c^2})^2$.

Chapter 10

Conclusion

This thesis explored a variety of topics, with work on both the SHiP and LHCb experiments. The following provides a summary of this work.

A novel fast generative approach of producing muon kinematics for the study of muon induced backgrounds at SHiP was developed. Whilst this new approach has some limitations regarding the fidelity and population of muons in the tails of the distribution it can produce useful background samples $\mathcal{O}(10^6)$ times faster than the full target simulation [17].

A refined implementation of this GAN based event generation approach was then later in the thesis presented as a proof on concept. The new architecture provides improved performance in the tails of the distribution and new flexibility to control the generated output after training. This flexibility comes from auxiliary parameters that are shown to the generator network as conditional inputs. These auxiliary parameters encode the *rarity* of each muon in the training sample. Examples of this flexibility and a discussion of potential applications within SHiP were provided.

Rates of combinatorial muon background at SHiP were estimated. This is an extremely important background at SHiP. With a fully simulated muon sample it was shown that the background rate can be suppressed, though this required employing an assumption that the suppression efficiencies from the selection cuts and from the vetoing systems factorise. With a

larger generated sample this assumption was found not to hold. However, avoiding the assumption was shown to significantly improve background suppression. The fully simulated result is therefore taken as an upper limit. The systems in place to suppress this background are sufficient even when employing a loose impact parameter cut of 2.5 m, such a wide cut would be chosen to keep as much partially reconstructed signal as possible [19].

Sensitivity estimates of the SHiP experiment to a range of standardised RPV neutralino benchmark scenarios were presented. For the first time these take into account cascade production of heavy flavour in the target, reconstruction efficiencies and background suppression. Estimates were also presented for new benchmarks involving production from B_s and B_c mesons, and benchmarks where the neutralino decays into fully leptonic final states. Comparisons to similar proposed facilities showed SHiP to have unprecedented sensitivity to neutralinos produced from charm decays. For sensitivity to neutralinos from beauty decays SHiP is surpassed by the MATHUSLA experiment where gains come from the larger beauty production at the LHC due to the higher centre of mass energy.

An investigation was made into the performance of the modern computer architecture, Graphcore IPU chips, and the potential for these IPUs to speed up particle physics generative networks. The performance of the IPU relative to a GPU option for both GAN inference and GAN training was presented. The IPUs dominate for low batch sizes as are often chosen during the training of GANs, though at higher batch sizes the GPU remains competitive. The IPUs therefore may be especially useful for model development where faster training times would allow for more rigorous testing of various architectures or hyperparameter optimisation [93].

For the LHCb experiment, modifications were made to the q^2 binned $B^0 \to K^{*0}\mu^+\mu^-$ analysis strategy allowing for a simultaneous fit for all CP-averaged and CP-asymmetric observables. This involved an extended fit where the branching fraction of $B^0 \to K^{*0}\mu^+\mu^-$ is measured relative to the $B^0 \to J/\psi K^{*0}$ control mode. The extended fit was shown to behave correctly for the CP-asymmetric observables in toys studies. However, fitting with the expected yields of the data set sizeable biases exist for the CP-averaged observables. This study suggested that CP-averaged and CP-asymmetric observables should be obtained in two separate fits to avoid these biases. This may be avoided in the future with more events where it may be possible to fit

all the observables in one simultaneous fit, this was demonstrated with toys that used enhanced yields. The results of this angular analysis, to come in the near future, will make a significant contribution to the search for new physics in the semi-leptonic rare decays of B mesons. If new measurements of lepton flavour universality measurements R_K and R_{K^*} reveal indisputable anomalies then muon only measurements that provide q^2 dependent information, such as this $B^0 \to K^{*0}\mu^+\mu^-$ angular analysis, will make vital contributions to global fits constraining the properties of this new physics.

Bibliography

[1] P. W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". *Phys. Rev. Lett.* (1964). doi: 10.1103/PhysRevLett.13.508.

[2] **CMS** Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". *Physics Letters B* (2012). doi: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235.

[3] **ATLAS** Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". *Phys. Lett.* (2012). doi: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[4] C. Langenbruch. "Rare decays at LHCb". In: *Proceedings, 51st Rencontres de Moriond on QCD and High Energy Interactions*. 2016. arXiv: 1605.07075 [hep-ex].

[5] **LHCb** Collaboration. "Search for New Physics in rare decays at LHCb". *Nucl. Phys. Proc. Suppl.* (2013). doi: 10.1016/j.nuclphysbps.2013.06.009. arXiv: 1209.1208 [hep-ex].

[6] M. Algueró et al. "Emerging patterns of New Physics with and without Lepton Flavour Universal contributions". *Eur. Phys. J. C* (2019). [Addendum: Eur.Phys.J.C 80, 511 (2020)]. doi: 10.1140/epjc/s10052-019-7216-3. arXiv: 1903.09578 [hep-ph].

[7] F. Archilli et al. "Flavour-changing neutral currents making and breaking the standard model". *Nature* (2017). doi: 10.1038/nature21721.

[8] R. Aaij et al. "Measurement of CP-Averaged Observables in the $B^0 \to K^{*0}\mu^+\mu^-$ Decay". *Phys. Rev. Lett.* (2020). doi: 10.1103/PhysRevLett.125.011802. arXiv: 2003.04831 [hep-ex].

[9] **LHCb** Collaboration. "Search for lepton-universality violation in $B^+ \to K^+\ell^+\ell^-$ decays". *Phys. Rev. Lett.* (2019). DOI: 10.1103/PhysRevLett.122.191801. arXiv: 1903.09252 [hep-ex].

[10] **LHCb** Collaboration. "Measurement of the ratio of branching fractions $\mathcal{B}(\bar{B}^0 \to D^{*+}\tau^-\bar{\nu}_\tau)/\mathcal{B}(\bar{B}^0 \to D^{*+}\mu^-\bar{\nu}_\mu)$". *Phys. Rev. Lett.* (2015). DOI: 10.1103/PhysRevLett.115.159901,10.1103/PhysRevLett.115.111803. arXiv: 1506.08614 [hep-ex].

[11] X. Cid Vidal et al. "Beyond the Standard Model Physics at the HL-LHC and HE-LHC" (2018). arXiv: 1812.07831 [hep-ph].

[12] A. Cerri et al. "Report from Working Group 4: Opportunities in Flavour Physics at the HL-LHC and HE-LHC". *CERN Yellow Rep. Monogr.* (2019). Ed. by A. Dainese et al. DOI: 10.23731/CYRM-2019-007.867. arXiv: 1812.07638 [hep-ph].

[13] F. Kahlhoefer. "Review of LHC Dark Matter Searches". *Int. J. Mod. Phys.* (2017). DOI: 10.1142/S0217751X1730006X. arXiv: 1702.02430 [hep-ph].

[14] A. Cakir. "Searches for Beyond Standard Model Physics at the LHC: Run1 Summary and Run2 Prospects". *PoS* (2015). DOI: 10.22323/1.248.0024. arXiv: 1507.08427 [hep-ph].

[15] L. Lee et al. "Collider Searches for Long-Lived Particles Beyond the Standard Model". *Prog. Part. Nucl. Phys.* (2019). DOI: 10.1016/j.ppnp.2019.02.006. arXiv: 1810.12602 [hep-ph].

[16] **SHiP** Collaboration. "A facility to Search for Hidden Particles (SHiP) at the CERN SPS" (2015). arXiv: 1504.04956 [physics.ins-det].

[17] **SHiP** Collaboration. "Fast simulation of muons produced at the SHiP experiment using Generative Adversarial Networks" (2019). DOI: 10.1088/1748-0221/14/11/P11028. arXiv: 1909.04451.

[18] **SHiP** Collaboration. "SHiP Experiment - Progress Report". CERN-SPSC-2019-010 ; SPSC-SR-248 (2019).

[19] **SHiP** Collaboration. "SHiP Experiment - Comprehensive Design Study report". CERN-SPSC-2019-049 ; SPSC-SR-263 (2019).

[20] A. Pich. "The Standard model of electroweak interactions". *arXiv preprint arXiv:1201.0537* (2012).

[21] W. N. Cottingham and D. A. Greenwood. *An introduction to the standard model of particle physics*. Cambridge university press, 2007.

[22] D. Griffiths. *Introduction to elementary particles*. John Wiley & Sons, 2020.

[23] G. Bertone, D. Hooper, and J. Silk. "Particle dark matter: Evidence, candidates and constraints". *Phys. Rept.* (2005). doi: 10.1016/j.physrep.2004.08.031. arXiv: hep-ph/0404175.

[24] E. Corbelli and P. Salucci. "The extended rotation curve and the dark matter halo of M33". *Monthly Notices of the Royal Astronomical Society* (2000). doi: 10.1046/j.1365-8711.2000.03075.x. arXiv: astro-ph/9909252.

[25] R. Massey, T. Kitching, and J. Richard. "The dark matter of gravitational lensing". *Reports on Progress in Physics* (2010). doi: 10.1088/0034-4885/73/8/086901. arXiv: 1001.1739 [astro-ph.CO].

[26] G. Hinshaw et al. "Five-year wilkinson microwave anisotropy probe* observations: data processing, sky maps, and basic results". *The Astrophysical Journal Supplement Series* (2009). doi: 10.1088/0067-0049/180/2/225. arXiv: 0803.0732.

[27] K. Garrett and G. Duda. "Dark matter: A primer". *Advances in Astronomy* 2011 (2011).

[28] S. W. Randall et al. "Constraints on the self-interaction cross section of dark matter from numerical simulations of the merging galaxy cluster 1E 0657–56". *The Astrophysical Journal* (2008). doi: 10.1086/587859. arXiv: 0704.0261.

[29] G. Busoni et al. "Making the most of the relic density for dark matter searches at the LHC 14 TeV run". *Journal of Cosmology and Astroparticle Physics* (2015). doi: 10.1088/1475-7516/2015/03/022. arXiv: 1410.7409 [hep-ph].

[30] S. Funk. "Indirect detection of dark matter with γ rays". *Proceedings of the National Academy of Sciences* (2015). DOI: 10.1073/pnas.1308728111. arXiv: 1310.2695.

[31] **LZ** Collaboration. "The LZ dark matter experiment". *JPhCS* (2016). DOI: 10.1088/1742-6596/718/4/042039.

[32] **XENON** Collaboration. "Constraining the spin-dependent WIMP-nucleon cross sections with XENON1T". *Physical Review Letters* (2019). DOI: 10.1103/PhysRevLett.122.141301. arXiv: 1902.03234.

[33] **LHCb** Collaboration. "Observation of CP Violation in Charm Decays". *Phys. Rev. Lett.* (2019). DOI: 10.1103/PhysRevLett.122.211803. arXiv: 1903.08726 [hep-ex].

[34] **NA48** Collaboration. "A New measurement of direct CP violation in two pion decays of the neutral kaon". *Physics Letters B* (1999). DOI: 10.1016/S0370-2693(99)01030-8. arXiv: hep-ex/9909022.

[35] **KTeV** Collaboration. "Observation of Direct CP Violation in $K_{S,L} \to \pi\pi$ Decays" (1999). arXiv: hep-ex/9905060.

[36] **LHCb** Collaboration. "First Observation of CP Violation in the Decays of B_s Mesons". *Physical review letters* (2013). DOI: 10.1103/PhysRevLett.110.221601. arXiv: 1304.6173.

[37] **LHCb** Collaboration. "Observation of CP violation in charm decays". *Physical review letters* (2019). DOI: 10.1103/PhysRevLett.122.211803. arXiv: 1903.08726.

[38] **DUNE** Collaboration. "Long-Baseline Neutrino Facility (LBNF) and Deep Underground Neutrino Experiment (DUNE) conceptual design report" (2015). arXiv: 1512.06148.

[39] **T2K** Collaboration. "Constraint on the matter-antimatter symmetry-violating phase in neutrino oscillations" (2019). arXiv: 1910.03887.

[40] **T2K** Collaboration. "Measurements of neutrino oscillation in appearance and disappearance channels by the T2K experiment with 6.6×10^{20} protons on target". *Phys. Rev.* (2015). DOI: 10.1103/PhysRevD.91.072010. arXiv: 1502.01550 [hep-ex].

[41] **SHiP** Collaboration. "Sensitivity of the SHiP experiment to Heavy Neutral Leptons". *JHEP* (2019). DOI: 10.1007/JHEP04(2019)077. arXiv: 1811.00930 [hep-ph].

[42] P. Minkowski. "$\mu \to e\gamma$ at a Rate of One Out of 1×10^9 Muon Decays". *Phys. lett. B* (1977). DOI: https://doi.org/10.1016/0370-2693(77)90435-X.

[43] T. Yanagida. "Proc. Workshop on Unified theory and the baryon number in the universe". *KEK Report No. 79-18* (1979).

[44] O. Sawada et al. "Neutrino mass and spontaneous parity violation". *Phys. Rev. Lett* (1980). DOI: 10.1103/PhysRevLett.44.912.

[45] J. Beacham et al. "Physics Beyond Colliders at CERN: Beyond the Standard Model Working Group Report". *J. Phys. G* (2020). DOI: 10.1088/1361-6471/ab4cd2. arXiv: 1901.09966 [hep-ex].

[46] S. Alekhin et al. "A facility to Search for Hidden Particles at the CERN SPS: the SHiP physics case". *Rept. Prog. Phys.* (2016). DOI: 10.1088/0034-4885/79/12/124201. arXiv: 1504.04855 [hep-ph].

[47] **SHiP** Collaboration. "Sensitivity of the SHiP experiment to light dark matter" (2020). arXiv: 2010.11057.

[48] G. Lanfranchi, M. Pospelov, and P. Schuster. "The Search for Feebly-Interacting Particles" (2020). DOI: 10.1146/annurev-nucl-102419-055056. arXiv: 2011.02157 [hep-ph].

[49] G. Lanfranchi. ""Sensitivity of the SHiP experiment to a light scalar particle mixing with the Higgs"". CERN-SHiP-NOTE-2017-001 (2017).

[50] W. Buchmüller, P. Di Bari, and M. Plümacher. "Leptogenesis for pedestrians". *Annals of Physics* (2005). DOI: https://doi.org/10.1016/j.aop.2004.02.003.

[51] A. Pilaftsis. "The little review on leptogenesis". In: *Journal of Physics: Conference Series*. IOP Publishing. 2009. DOI: https://doi.org/10.1088/1742-6596/171/1/012017. arXiv: 0904.1182.

[52] M. Shaposhnikov. "Is there a new physics between electroweak and Planck scales?" (2007). arXiv: 0708.3550.

[53] M. Shaposhnikov. "A possible symmetry of the νMSM". *Nuclear Physics B* (2007). arXiv: hep-ph/0605047.

[54] T. Asaka and M. Shaposhnikov. "The νMSM, dark matter and baryon asymmetry of the universe". *Physics Letters B* (2005). arXiv: hep-ph/0505013.

[55] T. Asaka, S. Blanchet, and M. Shaposhnikov. "The nuMSM, dark matter and neutrino masses". *Phys. Lett. B* (2005). DOI: 10.1016/j.physletb.2005.09.070. arXiv: hep-ph/0503065.

[56] B. Döbrich et al. "ALPtraum: ALP production in proton beam dump experiments". *Journal of High Energy Physics* (2016). arXiv: 1512.03069.

[57] R. Barbier et al. "R-Parity Violating Supersymmetry". *Phys. Rept.* (2005). arXiv: hep-ph/0406039.

[58] P. Martin. "A Supersymmetry Primer" (1997). arXiv: hep-ph/9709356.

[59] H. K. Dreiner. "An Introduction to Explicit R-Parity Violation". *Pramana* (1998). arXiv: hep-ph/9707435.

[60] R. N. Mohapatra. "Supersymmetry and R-Parity: An Overview". *Physica Scripta* (2015). arXiv: 1503.06478.

[61] V. Khachatryan et al. "Searches for supersymmetry using the M T2 variable in hadronic events produced in pp collisions at 8 TeV". *Journal of High Energy Physics* (2015). arXiv: 1502.04358.

[62] D. Dercks et al. "R-parity Violation and Light Neutralinos at CODEX-b, FASER, and MATHUSLA". *Physical Review D* (2019). DOI: 10.1103/PhysRevD.99.055039. arXiv: 1810.03617.

[63] H. K. Dreiner et al. "Mass Bounds on a Very Light Neutralino". *Eur. Phys. J. C* (2009). arXiv: 0901.3485.

[64] G. Bélanger et al. "LHC constraints on light neutralino dark matter in the MSSM". *Phys. Lett. B* (2013). arXiv: 1308.3735.

[65] D. Hooper and T. Plehn. "Supersymmetric dark matter: How light can the LSP be?" *Phys. Lett. B* (2003). arXiv: hep-ph/0212226.

[66] B. C. Allanach, A. Dedes, and H. K. Dreiner. "R-parity violating minimal supergravity model". *Physical Review D* (2004). doi: 10.1103/PhysRevD.69.115002. arXiv: hep-ph/0309196.

[67] J. de Vries, H. K. Dreiner, and D. Schmeier. "R-parity violation and light neutralinos at SHiP and the LHC". *Physical Review D* (2016). arXiv: 1511.07436.

[68] Y. Kao and T. Takeuchi. "Single-Coupling Bounds on R-parity violating Supersymmetry, an update" (2009). arXiv: 0910.4980.

[69] C. Ahdida et al. "SPS Beam Dump Facility–Comprehensive Design Study" (2019). doi: 10.23731/CYRM-2020-002. arXiv: 1912.06356.

[70] H. Dijkstra and T. Ruf. "Heavy Flavour Cascade Production in a Beam Dump". CERN-SHiP-NOTE-2015-009 (2015).

[71] E. L. Sola et al. "Design of a high power production target for the Beam Dump Facility at CERN". *Physical Review Accelerators and Beams* (2019). doi: 10.1103/PhysRevAccelBeams.22.113001.

[72] **SHiP** Collaboration. "The active muon shield in the SHiP experiment". *JINST* (2017). doi: 10.1088/1748-0221/12/05/P05011. arXiv: 1703.03612.

[73] T. Sjostrand, S. Mrenna, and P. Z. Skands. "A Brief Introduction to PYTHIA 8.1". *Comput. Phys. Commun.* (2008). arXiv: 0710.3820.

[74] O. J. I. Lantwin. "Optimisation of the SHiP experimental design". PhD Thesis. Imperial Coll., London, CERN-THESIS-2019-157 (2019).

[75] **SHiP** Collaboration. "The Magnet of the Scattering and Neutrino Detector for the SHiP experiment at CERN". *Journal of Instrumentation* (2020). doi: 10.1088/1748-0221/15/01/P01027. arXiv: 1910.02952.

[76] R. Acquafredda et al. "The OPERA experiment in the CERN to Gran Sasso neutrino beam". *Journal of Instrumentation* (2009). DOI: 10.1088/1748-0221/4/04/P04018.

[77] M. Al-Turany et al. "The FairRoot framework". *J. Phys. Conf. Ser.* (2012). DOI: 10.1088/1742-6596/396/2/022001.

[78] C. Andreopoulos et al. "The GENIE Neutrino Monte Carlo Generator". *Nucl. Instrum. Meth.* (2010). DOI: 10.1016/j.nima.2009.12.009. arXiv: 0905.2517.

[79] T. Sjostrand, S. Mrenna, and P. Z. Skands. "PYTHIA 6.4 Physics and Manual". *JHEP* (2006). DOI: 10.1088/1126-6708/2006/05/026. arXiv: hep-ph/0603175.

[80] E. Van Herwijnen et al. "Simulation and Pattern Recognition for the SHiP Spectrometer Tracker". CERN-SHiP-NOTE-2015-002 (2015).

[81] B. Hosseini and W. M. Bonivento. "Particle Identification Tools and Performance in the SHiP Experiment". CERN-SHiP-NOTE-2017-002 (2017).

[82] L. de Oliveira, M. Paganini, and B. Nachman. "Learning Particle Physics by Example: Location-Aware Generative Adversarial Networks for Physics Synthesis". *Comput. Softw. Big Sci.* (2017). DOI: 10.1007/s41781-017-0004-6. arXiv: 1701.05927.

[83] S. Amoroso et al. "Challenges in Monte Carlo event generator software for High-Luminosity LHC" (2020). Ed. by A. Valassi, E. Yazgan, and J. McFayden. arXiv: 2004.13687 [hep-ph].

[84] E. Karavakis et al. "Common accounting system for monitoring the ATLAS Distributed Computing resources". *J. Phys. Conf. Ser* ATL-SOFT-PROC-2013-009 (2014).

[85] M. Paganini, L. de Oliveira, and B. Nachman. "CaloGAN : Simulating 3D high energy particle showers in multilayer electromagnetic calorimeters with generative adversarial networks". *Phys. Rev. D* (2018). DOI: 10.1103/PhysRevD.97.014021. arXiv: 1712.10321.

[86] A. M. Sirunyan et al. "Observation of Higgs boson decay to bottom quarks". *Phys. Rev. Lett.* (2018). DOI: 10.1103/PhysRevLett.121.121801. arXiv: 1808.08242 [hep-ex].

[87] G. Apollinari et al. "High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1". CERN-2017-007-M (2017).

[88] **LHCb** Collaboration. "Physics case for an LHCb Upgrade II - Opportunities in flavour physics, and beyond, in the HL-LHC era" (2018). arXiv: 1808.08865 [hep-ex].

[89] P. Canal et al. "GeantV: from CPU to accelerators". *PoS* (2016). doi: 10.22323/1.282.0177.

[90] G. Amadio et al. "GeantV: Results from the prototype of concurrent vector particle transport simulation in HEP" (2020). arXiv: 2005.00949 [physics.comp-ph].

[91] J. Albrecht et al. "A Roadmap for HEP Software and Computing R&D for the 2020s". *Comput. Softw. Big Sci.* (2019). doi: 10.1007/s41781-018-0018-8. arXiv: 1712.06982 [physics.comp-ph].

[92] P. Musella and F. Pandolfi. "Fast and accurate simulation of particle detectors using generative adversarial networks". *Comput. Softw. Big Sci.* (2018). doi: 10.1007/s41781-018-0015-y. arXiv: 1805.00850.

[93] L. R. M. Mohan et al. "Studying the potential of Graphcore IPUs for applications in Particle Physics" (2020). arXiv: 2008.09210 [physics.comp-ph].

[94] T. Ruf and E. Van Herwijnen. "Description of the SHiP FixedTargetGenerator". CERN-SHiP-INT-2017-001 (2017).

[95] T. Karras et al. "Progressive growing of GANs for improved quality, stability, and variation" (2017). arXiv: 1710.10196.

[96] J. Yu et al. "Generative Image Inpainting with Contextual Attention" (2018). arXiv: 1801.07892.

[97] H. Zhang et al. "Self-attention generative adversarial networks" (2019). arXiv: 1805.08318.

[98] H. Zhang et al. "Stackgan: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks" (2016). arXiv: 1612.03242.

[99] C. Ledig et al. "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network". *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* (2017). arXiv: 1609.04802.

[100] T. Wang et al. "High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs". *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* (2018). arXiv: 1711.11585.

[101] P. Isola et al. "Image-to-Image Translation with Conditional Adversarial Networks" (2016). arXiv: 1611.07004.

[102] R. Di Sipio et al. "DijetGAN: a Generative-Adversarial Network approach for the simulation of QCD dijet events at the LHC". *Journal of High Energy Physics* (2019). doi: 10.1007/JHEP08(2019)110. arXiv: 1903.02433.

[103] A. Butter, T. Plehn, and R. Winterhalder. "How to GAN Event Subtraction" (2019). arXiv: 1912.08824 [hep-ph].

[104] J. Arjona Martínez et al. "Particle Generative Adversarial Networks for full-event simulation at the LHC and their application to pileup description". *J. Phys. Conf. Ser.* (2020). doi: 10.1088/1742-6596/1525/1/012081. arXiv: 1912.02748 [hep-ex].

[105] S. Carrazza and F. A. Dreyer. "Lund jet images from generative and cycle-consistent adversarial networks". *Eur. Phys. J. C* (2019). doi: 10.1140/epjc/s10052-019-7501-1. arXiv: 1909.01359 [hep-ph].

[106] A. Butter, T. Plehn, and R. Winterhalder. "How to GAN LHC Events". *SciPost Phys.* (2019). doi: 10.21468/SciPostPhys.7.6.075. arXiv: 1907.03764 [hep-ph].

[107] S. Otten et al. "Event Generation and Statistical Sampling for Physics with Deep Generative Models and a Density Information Buffer" (2019). arXiv: 1901.00875 [hep-ph].

[108] M. Paganini, L. de Oliveira, and B. Nachman. "Accelerating science with generative adversarial networks: an application to 3D particle showers in multilayer calorimeters". *Physical review letters* (2018). doi: 10.1103/PhysRevLett.120.042003. arXiv: 1705.02355.

[109] A. Maevskiy et al. "Fast Data-Driven Simulation of Cherenkov Detectors Using Generative Adversarial Networks". In: *19th International Workshop on Advanced Computing and Analysis Techniques in Physics Research*. 2019. DOI: 10.1088/1742-6596/1525/1/012097. arXiv: 1905.11825 [physics.ins-det].

[110] M. Erdmann, J. Glombitza, and T. Quast. "Precise simulation of electromagnetic calorimeter showers using a Wasserstein Generative Adversarial Network". *Comput. Softw. Big Sci.* (2019). DOI: 10.1007/s41781-018-0019-7. arXiv: 1807.01954 [physics.ins-det].

[111] E. Buhmann et al. "Getting High: High Fidelity Simulation of High Granularity Calorimeters with High Speed" (2020). arXiv: 2005.05334 [physics.ins-det].

[112] M. Bellagente et al. "How to GAN away Detector Effects". *SciPost Phys.* (2020). DOI: 10.21468/SciPostPhys.8.4.070. arXiv: 1912.00477 [hep-ph].

[113] A. Ghosh. "Deep generative models for fast shower simulation in ATLAS". *J. Phys. Conf. Ser.* (2020). DOI: 10.1088/1742-6596/1525/1/012077.

[114] F. Carminati et al. "Generative Adversarial Networks for fast simulation". *J. Phys. Conf. Ser.* (2020). DOI: 10.1088/1742-6596/1525/1/012064.

[115] D. Belayneh et al. "Calorimetry with Deep Learning: Particle Simulation and Reconstruction for Collider Physics". *Eur. Phys. J. C* (2020). DOI: 10.1140/epjc/s10052-020-8251-9. arXiv: 1912.06794 [physics.ins-det].

[116] D. Derkach et al. "Cherenkov Detectors Fast Simulation Using Neural Networks" (2019). DOI: 10.1016/j.nima.2019.01.031. arXiv: 1903.11788 [hep-ex].

[117] B. Hashemi et al. "LHC analysis-specific datasets with Generative Adversarial Networks" (2019). arXiv: 1901.05282.

[118] A. L. Maas, A. Y. Hannun, and A. Y. Ng. "Rectifier Nonlinearities Improve Neural Network Acoustic Models". *Proc. icml* (2013).

[119] D. E. Rumelhart, G. E. Hinton, R. J. Williams, et al. "Learning representations by back-propagating errors". *Cognitive modeling* (1988). DOI: 10.1038/323533a0.

[120] I. Goodfellow et al. "Generative Adversarial Networks". *Advances in neural information processing systems* (2014). arXiv: 1406.2661.

[121] C. Weisser and M. Williams. "Machine learning and multivariate goodness of fit" (2016). arXiv: 1612.07186.

[122] S. Ruder. *An Overview of Gradient Descent Optimization Algorithms*. arXiv:1609.04747.

[123] N. Srivastava et al. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". *The Journal of Machine Learning Research* (2014).

[124] S. Ioffe and C. Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift" (2015). arXiv: 1502.03167.

[125] D. P. Kingma and J. Ba. "Adam: A Method for Stochastic Optimization" (2014). arXiv: 1412.6980.

[126] S. J. Reddi, S. Kale, and S. Kumar. "On the Convergence of Adam and Beyond" (2019). arXiv: 1904.09237.

[127] D. W. Scott. "Multivariate density estimation: theory, practice, and visualization". *John Wiley & Sons* (2015).

[128] F. Chollet et al. *Keras*. https://keras.io. 2015.

[129] M. Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. 2015. URL: http://tensorflow.org/.

[130] Nvidia. *NVIDIA TESLA P100 specifications*. https://www.nvidia.com/en-gb/data-center/tesla-p100/. 2020 (accessed 2 December, 2020).

[131] N. V. Chawla et al. "SMOTE: synthetic minority over-sampling technique". *Journal of artificial intelligence research* (2002). DOI: 10.1613/jair.953. arXiv: 1106.1813.

[132] H. He et al. "ADASYN: Adaptive synthetic sampling approach for imbalanced learning". *IEEE International Joint Conference on Neural Networks* (2008). DOI: 10.1109/IJCNN.2008.4633969.

[133] K. T. Matchev and P. Shyamsundar. "Uncertainties associated with GAN-generated datasets in high energy physics" (Feb. 2020). arXiv: 2002.06307 [hep-ph].

[134] A. Butter et al. "GANplifying Event Samples" (Aug. 2020). arXiv: 2008.06545 [hep-ph].

[135] A. Butter et al. "Amplifying Statistics using Generative Models" (2020). URL: https://ml4physicalsciences.github.io/2020/files/NeurIPS_ML4PS_2020_59.pdf.

[136] **SHiP** Collaboration. "Measurement of the muon flux for the SHiP experiment" (2020). arXiv: 2001.04784 [physics.ins-det].

[137] T. Ruf. "Studying Multiple Scattering with GEANT4 v10.3.2". CERN-SHiP-NOTE-2017-003 (2018).

[138] V. Gligorov and M. Williams. "Efficient, reliable and fast high-level triggering using a bonsai boosted decision tree". *JINST* (2013). DOI: 10.1088/1748-0221/8/02/P02013. arXiv: 1210.6861 [physics.ins-det].

[139] H. K. Dreiner, J. Y. Günther, and Z. S. Wang. "R-parity Violation and Light Neutralinos at ANUBIS and MAPP" (2020). arXiv: 2008.07539 [hep-ph].

[140] **CODEX-b** Collaboration. "Expression of Interest for the CODEX-b Detector" (2019). arXiv: 1911.00481 [hep-ex].

[141] **FASER** Collaboration. "FASER: ForwArd Search ExpeRiment at the LHC" (2019). arXiv: 1901.04468 [hep-ex].

[142] **FASER** Collaboration. "FASER's physics reach for long-lived particles". *Phys. Rev. D* (2019). DOI: 10.1103/PhysRevD.99.095011. arXiv: 1811.12522 [hep-ph].

[143] **MATHSULA** Collaboration. "An Update to the Letter of Intent for MATHUSLA: Search for Long-Lived Particles at the HL-LHC" (2020). arXiv: 2009.01693 [physics.ins-det].

[144] P. Zyla et al. "Review of Particle Physics". *PTEP* (2020). DOI: 10.1093/ptep/ptaa104.

[145] A. Dedes, H. Dreiner, and P. Richardson. "Attempts at explaining the NuTeV observation of dimuon events". *Physical Review D* 65.1 (2001), p. 015001.

[146] D. Gorbunov and I. Timiryasov. "Decaying light particles in the SHiP experiment. II. Signal rate estimates for light neutralinos". *Physical Review D* (2015). DOI: 10.1103/PhysRevD.92.075015. arXiv: 1508.01780 [hep-ph].

[147] **LHCb** Collaboration. "Observation of $B_c^+ \to D^0 K^+$ Decays". *Phys. Rev. Lett.* (2017). DOI: 10.1103/PhysRevLett.118.111803. arXiv: 1701.01856.

[148] **LHCb** Collaboration. "Measurement of b-hadron Production Fractions in 7 TeV pp Collisions". *Phys. Rev. D* (2012). DOI: 10.1103/PhysRevD.85.032008. arXiv: 1111.2357.

[149] M. Baker et al. "B Meson Decay Constants f_{B_c}, f_{B_s} and f_B from QCD Sum Rules". *JHEP* (2014). arXiv: 1310.0941.

[150] V. Abazov et al. "Search for R-parity violating supersymmetry via the $LL\bar{E}$ couplings λ_{121}, λ_{122} or λ_{133} in $p\bar{p}$ collisions at $s = 1.96$ TeV". *Physics Letters B* (2006). DOI: 10.1016/j.physletb.2006.05.077.

[151] K. Bondarenko et al. "Phenomenology of GeV-scale Heavy Neutral Leptons". *JHEP* (2018). arXiv: 1805.08567.

[152] M. Bauer et al. "ANUBIS: Proposal to search for long-lived neutral particles in CERN service shafts" (2019). arXiv: 1909.13022.

[153] C. Leggett and I. Shapoval. "Simulating HEP Workflows on Heterogeneous Architectures". In: *14th International Conference on e-Science*. 2018. DOI: 10.1109/eScience.2018.00087.

[154] B. Yeo, M. Lee, and Y. Kuno. "GPU-Accelerated Event Reconstruction for the COMET Phase-I Experiment" (2019). arXiv: 1911.09340 [physics.ins-det].

[155] R. Cenci et al. "Real-time reconstruction of long-lived particles at LHCb using FPGAs". *J. Phys. Conf. Ser.* (2020). DOI: 10.1088/1742-6596/1525/1/012101. arXiv: 2006.11067 [physics.ins-det].

[156] F. Lazzari et al. "Real-time cluster finding for LHCb silicon pixel VELO detector using FPGA". *J. Phys. Conf. Ser.* (2020). DOI: 10.1088/1742-6596/1525/1/012044.

[157] R. Aaij et al. "Allen: A high level trigger on GPUs for LHCb". *Comput. Softw. Big Sci.* (2020). DOI: 10.1007/s41781-020-00039-7. arXiv: 1912.09161 [physics.ins-det].

[158] Graphcore. *Graphcore Benchmarks*. https://www.graphcore.ai/benchmarks. 2020 (accessed 2 December, 2020).

[159] D. Guest, K. Cranmer, and D. Whiteson. "Deep Learning and its Application to LHC Physics". *Ann. Rev. Nucl. Part. Sci.* (2018). DOI: 10.1146/annurev-nucl-101917-021019. arXiv: 1806.11484 [hep-ex].

[160] Graphcore. *Graphcore.ai*. https://www.graphcore.ai/. 2020 (accessed 2 December, 2020).

[161] A. Paszke et al. "PyTorch: An Imperative Style, High-Performance Deep Learning Library". In: *Advances in Neural Information Processing Systems 32*. 2019. URL: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

[162] Intel. *Intel Xeon Platinum 8168 specifications*. https://ark.intel.com/content/www/us/en/ark/products/120504/intel-xeon-platinum-8168-processor-33m-cache-2-70-ghz.html. 2020 (accessed 2 December, 2020).

[163] Graphcore. *private communication*.

[164] Z. Jia et al. "Dissecting the Graphcore IPU Architecture via Microbenchmarking" (2019). arXiv: 1912.03413 [cs.DC].

[165] **CERN**. *Environment Report*. https://hse.cern/environment-report-2017-2018. 2020 (accessed 2 December, 2020).

[166] Graphcore. *Graphcore Benchmarks - MK2*. https://www.graphcore.ai/mk2-benchmarks. 2020 (accessed 2 December, 2020).

[167] G. Antipov, M. Baccouche, and J.-L. Dugelay. "Face aging with conditional generative adversarial networks". In: *IEEE ICIP*. 2017. arXiv: 1702.01983.

[168] M. Mirza and S. Osindero. *Conditional Generative Adversarial Nets*. arXiv:1411.1784.

[169] W. R. Tan et al. "ArtGAN: Artwork synthesis with conditional categorical GANs". In: *2017 IEEE ICIP*. 2017. arXiv: 1702.03410.

[170] A. Odena, C. Olah, and J. Shlens. "Conditional image synthesis with auxiliary classifier gans". In: *International conference on machine learning*. 2017. arXiv: 1610.09585.

[171] A. Dash et al. "Tac-gan-text conditioned auxiliary classifier generative adversarial network" (2017). arXiv: 1703.06412.

[172] S. Bazrafkan and P. Corcoran. "Versatile auxiliary regressor with generative adversarial network (VAR+ GAN)" (2018). arXiv: 1805.10864.

[173] V. Laparra, G. Camps-Valls, and J. Malo. "Iterative gaussianization: from ICA to random rotations". *IEEE transactions on neural networks* (2011). DOI: 10.1109/TNN.2011.2106511.

[174] C. Meng et al. "Gaussianization Flows" (2020). arXiv: 2003.01941.

[175] A. Makhzani et al. "Adversarial autoencoders" (2015). arXiv: 1511.05644.

[176] D. P. Kingma and M. Welling. "Auto-encoding variational bayes" (2013). arXiv: 1312.6114.

[177] S. Shirobokov et al. "Black-Box Optimization with Local Generative Surrogates" (Feb. 2020). arXiv: 2002.04632 [cs.LG].

[178] A. B. L. Larsen et al. "Autoencoding beyond pixels using a learned similarity metric". In: *International conference on machine learning*. 2016. arXiv: 1512.09300.

[179] D. Ulyanov, A. Vedaldi, and V. Lempitsky. "It takes (only) two: Adversarial generator-encoder networks" (2017). arXiv: 1704.02304.

[180] A. Alanov et al. "Pairwise augmented gans with adversarial reconstruction loss" (2018). arXiv: 1810.04920.

[181] J. Alves A.Augusto et al. "The LHCb Detector at the LHC". *JINST* (2008). DOI: 10.1088/1748-0221/3/08/S08005.

[182] **LHCb** VELO group. "Performance of the LHCb vertex locator". *Journal of Instrumentation* (2014). arXiv: 1405.7808.

[183] **LHCb** Collaboration. *PID Plots for Conference*. https://twiki.cern.ch/twiki/bin/view/LHCb/PIDConferencePlots. 2017 (accessed 1 March, 2021).

[184] **LHCb** Collaboration. "LHCb detector performance". *International Journal of Modern Physics A* (2015). arXiv: 1412.6352.

[185] **LHCb** Collaboration. "Measurements of the S-wave fraction in $B^0 \to K^+\pi^-\mu^+\mu^-$ decays and the $B^0 \to K^*(892)^0\mu^+\mu^-$ differential branching fraction". *JHEP* (2016). [Erratum: JHEP 04, 142 (2017)]. DOI: 10.1007/JHEP11(2016)047. arXiv: 1606.04731 [hep-ex].

[186] R. Aaij et al. "Angular analysis and differential branching fraction of the decay $B_s^0 \to \phi\mu^+\mu^-$". *JHEP* (2015). DOI: 10.1007/JHEP09(2015)179. arXiv: 1506.08777 [hep-ex].

[187] A. Abdesselam et al. "Test of lepton flavor universality in $B \to K^*\ell^+\ell^-$ decays at Belle" (2019). arXiv: 1904.02440 [hep-ex].

[188] W. Altmannshofer et al. "Symmetries and Asymmetries of $B \to K^*\mu^+\mu^-$ Decays in the Standard Model and Beyond". *JHEP* (2009). DOI: 10.1088/1126-6708/2009/01/019. arXiv: 0811.1214 [hep-ph].

[189] R. Alonso, B. Grinstein, and J. Martin Camalich. "Lepton universality violation and lepton flavor conservation in B-meson decays". *JHEP* (2015). DOI: 10.1007/JHEP10(2015)184. arXiv: 1505.05164 [hep-ph].

[190] G. Hiller and M. Schmaltz. "R_K and future $b \to s\ell\ell$ physics beyond the standard model opportunities". *Phys. Rev. D* (2014). DOI: 10.1103/PhysRevD.90.054014. arXiv: 1408.1627 [hep-ph].

[191] T. Blake, T. Gershon, and G. Hiller. "Rare b hadron decays at the LHC". *Ann. Rev. Nucl. Part. Sci.* (2015). DOI: 10.1146/annurev-nucl-102014-022231. arXiv: 1501.03309 [hep-ex].

[192] W. Altmannshofer et al. "Symmetries and Asymmetries of $B \to K^*\mu^+\mu^-$ Decays in the Standard Model and Beyond". *Journal of High Energy Physics* (2009).

[193] A. Bharucha, D. M. Straub, and R. Zwicky. "$B \to V\ell^+\ell^-$ in the Standard Model from light-cone sum rules". *Journal of High Energy Physics* 2016 (2016).

[194] P. Alvarez Cartelle et al. "Angular analysis of $B^0 \to K^{*0}\mu^+\mu^-$ decays using Run 1 and 2016 data" (Oct. 2020). URL: https://cds.cern.ch/record/2286899.

[195] **LHCb** Collaboration. "Angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay using 3 fb^{-1} of integrated luminosity". *JHEP* (2016). DOI: 10.1007/JHEP02(2016)104. arXiv: 1512.04442 [hep-ex].

[196] D. Aston et al. "A Study of $K^-\pi^+$ Scattering in the Reaction $K^-p \to K^-\pi^+n$ at 11 GeV/c". *Nucl. Phys. B* (1988). DOI: 10.1016/0550-3213(88)90028-4.

[197] J. Bressieux et al. "Evidence for the resonant character of the $Z(4430)^- \to \psi(2S)\pi^-$ mass peak observed in $B^0 \to \psi(2S)K^+\pi^-$ decays, and determination of the $Z(4430)^-$ spin-parity" (June 2013). URL: https://cds.cern.ch/record/1553766.

[198] J. Back et al. "Laura^{++}: A Dalitz plot fitter". *Comput. Phys. Commun.* (2018). DOI: 10.1016/j.cpc.2018.04.017. arXiv: 1711.09854 [hep-ex].

[199] **LHCb** Collaboration. "Dalitz plot analysis of $B_s^0 \to \bar{D}^0K^-\pi^+$ decays". *Phys. Rev. D* (2014). DOI: 10.1103/PhysRevD.90.072003. arXiv: 1407.7712 [hep-ex].

[200] **BaBar** Collaboration. "Search for the $Z(4430)^-$ at BABAR". *Phys. Rev. D* (2009). DOI: 10.1103/PhysRevD.79.112001. arXiv: 0811.0564 [hep-ex].

[201] **BaBar** Collaboration. "Amplitude Analysis of the Decay $B^0 \to K^+\pi^-\pi^0$". In: *34th International Conference on High Energy Physics*. 2008. arXiv: 0807.4567 [hep-ex].

[202] T. Blake et al. "Angular analysis of $B^0 \to K^{*0}\mu^+\mu^-$ decays using 3 fb^{-1} of integrated luminosity" (Nov. 2013). URL: https://cds.cern.ch/record/1632246.

[203] **Belle** Collaboration. "Observation of a new charged charmonium-like state in $\bar{B}^0 \to J/\psi K^-\pi^+$ decays". *Phys. Rev. D* (2014). DOI: 10.1103/PhysRevD.90.112009. arXiv: 1408.6457 [hep-ex].

[204] K. Olive et al. "Review of Particle Physics". *Chin. Phys. C* (2014). DOI: 10.1088/1674-1137/38/9/090001.

[205] **LHCb** Collaboration. "Measurement of the polarization amplitudes in $B^0 \to J/\psi K^*(892)^0$ decays". *Phys. Rev. D* (2013). doi: 10.1103/PhysRevD.88.052002. arXiv: 1307.2782 [hep-ex].

[206] G. J. Feldman and R. D. Cousins. "Unified approach to the classical statistical analysis of small signals". *Physical Review D* (1998). arXiv: physics/9711021 [physics].