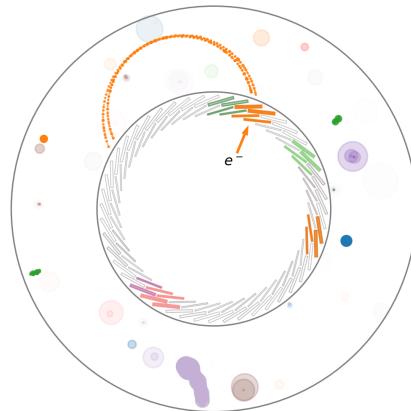


Sensitivity and Background Estimates towards Phase-I of the COMET Muon-to-Electron Conversion Search

Matthias Dubouchet

High Energy Physics Group
Department of Physics
Imperial College London



Submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

Abstract

COMET is a future high-precision experiment searching for charged lepton flavour violation through the muon-to-electron conversion process. It aims to push the intensity frontier of particle physics by coupling an intense muon beam with cutting-edge detector technology. The first stage of the experiment, COMET Phase-I, is currently being assembled and will soon enter its data acquisition period. It plans to achieve a single event sensitivity to $\mu-e$ conversion in aluminium of 3×10^{-15} .

This thesis presents a study of the sensitivity and backgrounds of COMET Phase-I using the latest Monte Carlo simulation data produced. The background contribution from cosmic ray-induced atmospheric muons is estimated using a backward Monte Carlo approach, which allows computational resources to be focused on the most critical signal-mimicking events.

Analysis of a $\mu-e$ conversion simulation sample suggests that COMET Phase-I will reach a single event sensitivity of 3.6×10^{-15} within 146 days of data acquisition. In that period, the background contribution from atmospheric muons is estimated to be 0.08 events under optimistic conditions, namely a 99.99 %-efficient Cosmic Ray Veto system and 99 % muon identification rate by the main detector.

Declaration

This dissertation is the result of my own work except where explicit reference is made to the work of others. The methods and results reported in Section 4.4 were investigated by a pair of students, Noam Mouelle and Irene Andreou, as part of a final year project toward their MSci degree. The backward Monte Carlo simulation method, presented and used in Chapters 5 and 6, was developed by Valentin Niess. The COMET experiment is designed and constructed by the COMET Collaboration. The ICEDUST software is maintained by the COMET Software Group and developed by contributors across the collaboration.

Matthias Dubouchet

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

Acknowledgements

I would like to thank my supervisor, Yoshi Uchida, for guiding me on this journey. I am grateful for your teachings and the freedom I was given in pursuing my objectives—sometimes from remote locations for extended periods of time! Thank you also to the other members of the COMET group at Imperial, Per, Phill, Kou and Roden, for the continuous feedback and constructive conversations.

Thanks to Ben and Ewen for the discussions that kickstarted my work on the COMET experiment. I have lost count of how many times I went back to your theses to find clarity. To Valentin, thank you for the discussions concerning your original cosmic ray study, from which mine is inspired in more ways than one.

And thank you to everyone from the COMET collaboration for undertaking such a difficult project, striving to reach the best achievable outcome despite many hurdles. It was a pleasure to discover some of your countries (and gastronomy), and a shame that we missed a few opportunities to meet (and eat) because of the global situation.

To the members of my family, thank you for your patience and care. Florie, thank you for proofreading this thesis, and well done managing to stay mostly awake in the process. I am grateful, and pleasantly surprised, that you have accompanied me throughout this adventure. Thanks also to all my close friends for the moral support and the good times.

I dedicate this work to my mother and my grandfather.

Contents

1	Lepton Flavour in the Standard Model and Beyond	1
1.1	Discovery of the muon	2
1.2	The muon in the Standard Model	2
1.2.1	Tensions between theory and experiment	3
1.2.2	Lepton flavour with massive neutrinos	4
1.3	Searching for charged lepton flavour violation	5
1.4	Muon-to-electron conversion	6
1.4.1	Standard Model backgrounds	7
1.5	Effective CLFV and the scale of new physics	10
2	The COMET Experiment	12
2.1	Proton beam	13
2.2	Pion-production section	16
2.3	Transport solenoid	16
2.4	Muon stopping target	19
2.5	Detector systems	20
2.5.1	StrECAL	20
2.5.2	CyDet	21
2.5.3	Cosmic ray veto	23
2.6	Conversion signature	23
2.7	Experimental backgrounds	24
2.8	Sensitivity and run time	25
3	Software and Simulation	27
3.1	The ICEDUST Software Suite	27

3.1.1	Data format	27
3.1.2	Simulation pipeline	29
3.1.3	Intermediate simulation file format	30
3.2	Monte Carlo simulation in ICEDUST	30
3.2.1	Geometry	32
3.2.2	Physics	32
3.2.3	Signal simulation	33
3.2.4	Representation of simulated CDC hits	33
3.3	Large-scale simulation: MC5	34
3.4	Animated CyDet event display	36
3.5	ICEDUST development	38
3.5.1	Version control	38
3.5.2	Continuous integration and deployment	39
3.5.3	Documentation	39
4	Data Augmentation with Generative Adversarial Networks	40
4.1	Generative Adversarial Networks	41
4.1.1	Wasserstein GAN	43
4.2	GANs in High Energy Physics	45
4.2.1	Event generation	45
4.2.2	Detector data generation	45
4.3	The CDC hit generator	46
4.3.1	Hit data structure	47
4.3.2	Event selection	48
4.3.3	Pre-processing	49
4.3.4	Hit sequence sampling and data augmentation	50
4.3.5	Network components	51
4.3.6	Network architectures	52
4.3.7	Training	55
4.3.8	Evaluation	56
4.4	Quality metrics	59
4.4.1	Inception score	59
4.4.2	Fréchet Inception Distance	60

4.4.3	Evaluating the CDC GAN performance	60
4.4.4	Interpretation of FID	62
5	Backward Monte Carlo Simulation of Atmospheric Muons	65
5.1	Backward Monte Carlo simulation	65
5.1.1	Principle	65
5.1.2	Method	66
5.2	Application to COMET Phase-I	68
5.2.1	Geometry	68
5.2.2	Event sampling	69
5.2.3	Atmospheric muon flux	70
5.2.4	Validations	71
5.2.5	Absolute rate estimation	72
5.2.6	μ - e conversion background rate	73
6	COMET Phase-I Sensitivity and Background Estimates	76
6.1	Data samples	76
6.1.1	μ - e conversion	77
6.1.2	Muon decay in orbit	78
6.1.3	Atmospheric muons	80
6.1.4	Sample weighting	81
6.2	Single event sensitivity	83
6.3	Signal and background event counts	84
6.3.1	Momentum spectrum	84
6.3.2	Particle identification	86
6.3.3	CRV efficiency	87
6.3.4	Summary	88
7	Discussion	90
7.1	Data augmentation with machine learning	90
7.1.1	Summary	90
7.1.2	Limitations and recommendations	91
7.2	Atmospheric muon backgrounds	93
7.2.1	Summary	93

7.2.2	Limitations and recommendations	95
8	Conclusion	98
A	Integration of backward Monte Carlo software into ICEDUST	101
A.1	Original implementation	101
A.2	Integration	102

Chapter 1

Lepton Flavour in the Standard Model and Beyond

The Standard Model (SM) of particle physics is possibly the most successful mathematical model of physical phenomena so far. It provides an accurate description of almost all observable interactions between known elementary particles. It yields predictions for what nature does and does not allow, and enables physicists to examine and test the fundamental laws of the universe. Furthermore, the SM establishes a rigid framework to make theoretical predictions, from which any measured deviations can then be interpreted as new physics.

The conservation of lepton flavour, which stems from an accidental symmetry of the SM, is currently under strict scrutiny by various experiments around the world. The observation of neutrino oscillations has already proven that lepton flavour is not conserved by neutral leptons, and this has further driven the search for lepton flavour violation among the charged leptons. If charged lepton flavour violation (CLFV) is observed, not only would it be paradigm-shifting evidence for physics beyond the Standard Model, but it will also help guide us toward the next theory of particle physics by hinting at the underlying processes which could be at play.

The most sensitive probe to search for charged lepton flavour violation is the muon, in large part because muons can be readily produced and focused into intense beams at accelerator facilities. This chapter describes how the muon has come from being a mysterious particle showering Earth from high in the atmosphere to becoming a tool in high energy physics experiments used to explore the frontier of our knowledge of nature.

1.1 Discovery of the muon

The first traces of muons were observed around 1937 by three experiments investigating the nature of cosmic ray-induced particle showers [1, 2, 3]. One of them was able to estimate the mass of the discovered particle at hundreds of times that of the electron: around the same as the strong-force-carrying meson predicted by Yukawa in 1935 [4]. Hence, the muon and Yukawa's particle were originally believed to be one and the same particle.

It was only a decade later, when the meson (now called π , for primary) was observed decaying into a muon, that the two particles were completely disambiguated. The discovery of the pion was attributed to the Lattes–Muirhead–Occhialini–Powell group [5], but it was Don Perkins, doctoral student at Imperial College London at the time, who recorded the first nuclear emulsion of a negative pion capture and set in motion the race to identify this new particle [6].

Subsequently, it was the fact that the muon appeared as nothing but a heavy electron which prompted Rabi to ask “who ordered that?” in response to its discovery. In fact, even the Standard Model of today remains unable to give a satisfactory answer to this question, since the SM does not motivate the existence of three generations of elementary particles. As far as the SM can explain, there is no fundamental reason for the existence of distinct flavours.

1.2 The muon in the Standard Model

The SM identifies the muon as the second-generation charged lepton, meaning it is a fermion with identical properties, aside from flavour and mass, as the electron and tau lepton. A muon in a vacuum can only decay through the weak force. The diagram for muon decay, $\mu^- \rightarrow e^- + \nu_\mu + \bar{\nu}_e$, is shown in Figure 1.1.

In the SM Lagrangian with massless neutrinos, none of the terms which involve leptons allow for flavour violation. The Lagrangian is invariant under transformations of the $U(1)_e \times U(1)_\mu \times U(1)_\tau$ group. Consequently, each lepton generation (e, μ, τ) has its own conserved number. In theory, this completely prevents a charged lepton from changing flavour without neutrinos being involved to balance the process.

These conservation laws do not correspond to a fundamental symmetry of nature;

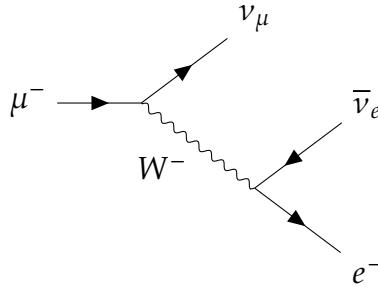


Figure 1.1: Feynman diagram for the weak decay of the muon. Note how lepton flavour is conserved by the presence of a muon neutrino and an electron anti-neutrino.

they are merely an accidental feature of the SM Lagrangian and have, so far, been observed to hold experimentally. For instance, the process $\mu \rightarrow e + \gamma$, in principle allowed by kinematics, has never been observed and the current upper limit on its branching ratio was set by the MEG experiment at 10^{-13} [7].

1.2.1 Tensions between theory and experiment

Various fundamental properties of the muon are currently under investigation by experiments around the world. Recent measurements have started to suggest that the charged lepton sector might be the next to reveal physics beyond the Standard Model (BSM). From only these results, predicting whether CLFV will also appear is not possible. However, CLFV searches are complementary in constraining the models of new physics which could lead to the observed discrepancies.

The Muon $g - 2$ experiment at Fermilab recently observed a 4.2σ discrepancy between the Standard Model prediction and the measured value for the magnetic moment of the muon [8]. This deviation, assuming it is a true sign of new physics¹, could be attributed to various scenarios where new particles, e.g. leptoquarks, a new Higgs doublet or neutral boson, interact with the muon.

The LHCb experiment at CERN has found evidence for lepton non-universality in the decay of b-quarks, with a significance of 3.1σ [10]. As with the $g - 2$ discrepancy, models which could explain the anomaly include a heavy neutral boson, leptoquarks, or an extended Higgs sector.

¹The theoretical expectation for the SM magnetic moment is disputed by a different approach using lattice QCD [9]. The latter calculation is indeed more consistent with the measurement, hence further investigation is necessary to understand the cause of the initial discrepancy and why the two calculations disagree.

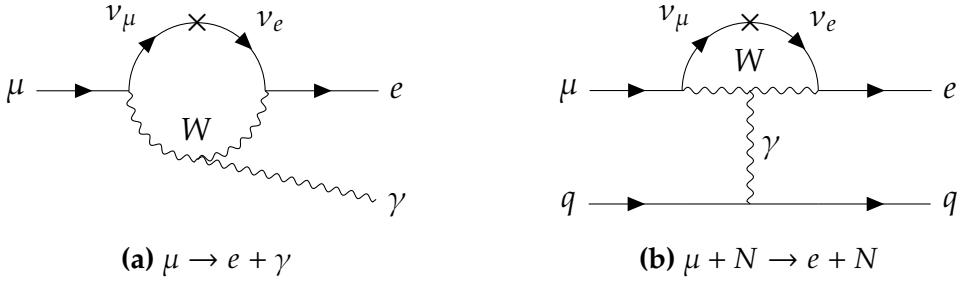


Figure 1.2: Feynman diagrams of processes allowing charged lepton flavour violation in the SM extended with neutrino masses. Although these processes enable CLFV, their branching ratios are heavily suppressed to unobservable levels because of the lightness of neutrinos compared to the weak scale [16].

Together, the Muon $g - 2$ and LHCb tensions suggest a special property of the muon which would differentiate it from the electron in a new, so far unexplained way. Although there are so far no tangible traces of CLFV, theoretical models attempting to resolve the $g - 2$ and LHCb anomalies often also allow CLFV at a level within experimental reach [11, 12, 13]. Hence, measurements such as the $\mu \rightarrow e$ conversion branching ratio are crucial to further constrain models and guide the theory toward the true nature of the new physics.

1.2.2 Lepton flavour with massive neutrinos

The observation of neutrino oscillations [14] means that the three accidental lepton flavour symmetries are not exact. It is now known that the flavour content of a neutrino changes as it propagates, hence the flavour quantum numbers are not true conserved quantities for leptons.

In the quark sector, flavour-changing neutral currents (FCNC) are one-loop order processes which allow a quark to change generations (e.g. from a strange quark to a down quark). The small rates of such processes are explained by the GIM (Glashow–Illiopoulos–Maiani) mechanism: interference between diagrams with different flavours of mediating quarks causes the branching ratio of FCNC to be heavily suppressed [15].

In the lepton sector, the fact that neutrinos are massless in the SM implies that FCNC among charged leptons is completely forbidden. With the addition of massive oscillating neutrinos, processes such as those shown in Figure 1.2 become allowed. These provide the only currently known way that charged lepton flavour can be violated.

However, the GIM suppression in this case is much more significant than in the quark sector because neutrinos are almost massless (compared to the weak scale). For instance, oscillating neutrinos allow the process shown in Figure 1.2a to occur, which gives rise to a non-zero rate for $\mu \rightarrow e + \gamma$. The branching ratio calculated for this process using the upper limit on neutrino masses is given by [16]

$$\mathcal{BR}(\mu \rightarrow e + \gamma) = \frac{3\alpha}{32\pi} \left| \sum_{i=2,3} U_{\mu i}^* U_{e i} \frac{\Delta m_{i1}^2}{m_W^2} \right|^2 \approx 10^{-54}, \quad (1.1)$$

where U is the PMNS matrix, Δm_{ij}^2 is the mass-squared difference between the i -th and j -th neutrino mass eigenstates, and m_W is the W -boson mass. A similar suppression also applies to the rates of other muon CLFV processes allowed in the SM extended with neutrino masses [17].

The vanishingly small branching ratio of Equation 1.1 implies that $\mu \rightarrow e\gamma$ cannot be observed experimentally unless some unexpected interaction is at play. Consequently, this and other CLFV processes are excellent probes to search for new physics. Specifically, many models for BSM physics give rise to direct mixing between charged lepton flavours at levels that may be observable in current and near-future experiments. Any experimental evidence that CLFV occurs at a rate higher than that of a neutrino-mediated FCNC would yield important clues as to what new physics lies beyond the SM.

1.3 Searching for charged lepton flavour violation

Three muon processes can be investigated in order to search for CLFV: $\mu \rightarrow e\gamma$, $\mu \rightarrow eee$ and $\mu N \rightarrow eN$ (muon-to-electron conversion) [16]. Observing any one of them would fundamentally contradict the Standard Model and thus provide undeniable evidence for a new interaction. Then, observing a second CLFV process would help us to determine which of the many theorised models for new physics is realised in nature.

CLFV has been sought after ever since the muon's discovery: the first investigation of whether nature allows $\mu \rightarrow e\gamma$ was performed in 1948 [22]. A multitude of experiments followed, but none so far have been able to find a signal [16]. Figure 1.3 shows experimentally-estimated upper limits on the branching ratios of $\mu \rightarrow e\gamma$, $\mu \rightarrow eee$ and

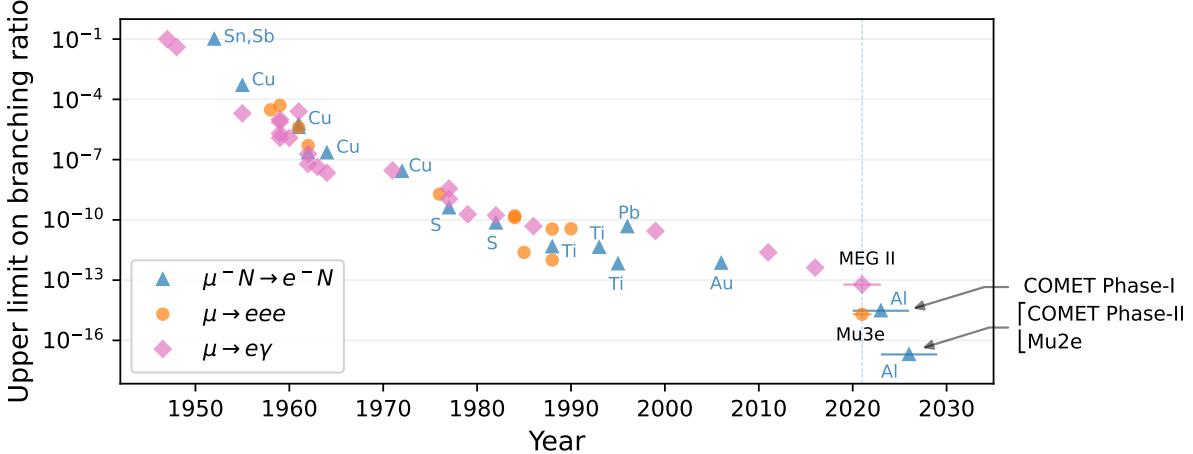


Figure 1.3: 90%-confidence upper limit on the branching ratio of three charged lepton flavour-violating processes over time. The target material N is indicated for μ - e conversion experiments. Past experiment results were tabulated by Bernstein and Cooper [16]. Future data points are the expected sensitivities quoted in the MEG II [18], Mu3e [19], COMET Phase-I [20] and Mu2e [21] design reports.

$\mu N \rightarrow e N$ over time, since the first experiment and into the next decade.

As higher and higher sensitivities are required, experiments must be able to produce a muon source which is increasingly intense while demonstrating precise control over every source of background. This is made possible by new technologies in both hardware and software applied across entire experiment designs. The next generation of CLFV-seeking precision experiments, which consists of MEG II, Mu3e, COMET and Mu2e, aims to be 10 to 10 000 times more sensitive than the last generation.

1.4 Muon-to-electron conversion

Muon-to-electron conversion is the neutrino-less decay of a muon bound to an atomic nucleus:

$$\mu^- + N(A, Z) \rightarrow e^- + N(A, Z), \quad (1.2)$$

where A is the mass number and Z the atomic number. Similarly to $\mu \rightarrow e\gamma$, this process is allowed in the SM extended with massive neutrinos via the diagram shown in Figure 1.2b, but suppressed to an experimentally unreachable level. Any signs of it occurring at current experimental sensitivities would suggest a BSM origin.

In order to search for this process, muons must be stopped in matter to form

muonic atoms. Initially bound in the outer atomic layers, the muon electromagnetically cascades down to the $1s$ orbital within close range of the nucleus over the following nanosecond [23]. When interacting with the nucleus, the reaction is called *coherent* if the nucleus remains unchanged and in its ground state. For elements heavier than magnesium, the ratio of coherent to incoherent $\mu-e$ conversion is expected to be around 9:1 [24].

In a coherent $\mu-e$ conversion, the kinematics are those of a straightforward two-body decay, hence the electron always has an energy

$$E_{\mu-e} = m_\mu - B_\mu - E_{\text{recoil}}, \quad (1.3)$$

where B_μ is the binding energy of the $1s$ -state muonic atom and E_{recoil} is the recoil energy of the nucleus. In aluminium, the target material of the COMET and Mu2e experiments, this yields

$$E_{\mu-e}^{\text{Al}} = 104.97 \text{ MeV}. \quad (1.4)$$

Since the signature of $\mu-e$ conversion is a single, mono-energetic electron, this process should be relatively easily identified by means of a momentum-measuring detector. However, this signal must also be discriminated from backgrounds originating from other processes, contamination of the beam and cosmic rays.

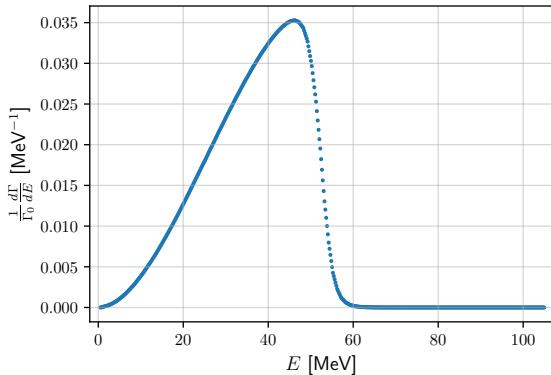
1.4.1 Standard Model backgrounds

Decay in orbit

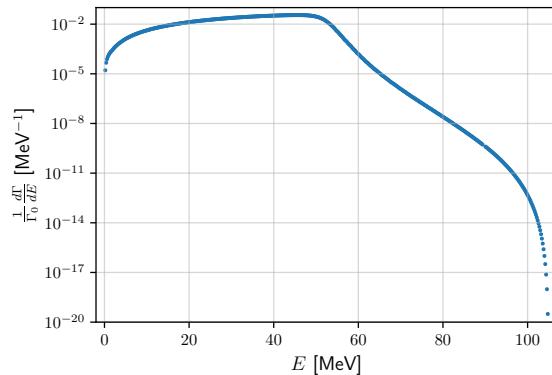
In the SM, a bound muon is allowed to decay in orbit (DIO):

$$\mu^- + N(A, Z) \rightarrow e^- + \nu_\mu + \bar{\nu}_e + N(A, Z). \quad (1.5)$$

Although the process is the same as a free decay, shown in Figure 1.1, here the nucleus is also involved in the kinematics of the process. In a free decay, the electron may carry at most half of the muon mass as kinetic energy (in the muon rest frame) when the two neutrinos are emitted in the opposite direction. In DIO, the nucleus can recoil and potentially provide more energy to the electron. If the two neutrinos are emitted back to back in the transverse direction to the electron, then the kinematics resemble that of



(a) Linear vertical axis.



(b) Logarithmic vertical axis.

Figure 1.4: Energy spectrum of decay-in-orbit electrons produced by muonic aluminium atoms, calculated by Czarnecki et al. [25]. The logarithmic scale reveals the shape of the high-energy tail up to the endpoint energy of 104.97 MeV.

$\mu-e$ conversion and the electron could easily be mistaken for the CLFV signal.

The energy spectrum of DIO electrons for various target nuclei is well understood theoretically [25], hence the expected background rate for $\mu-e$ conversion searches can be estimated. The DIO energy spectrum for aluminium is shown in Figure 1.4. Although it is highly unlikely that a DIO electron might carry as much energy as a conversion electron, DIO remains one of the main sources of background because of the extreme sensitivity aimed at by upcoming $\mu-e$ conversion searches, as well as the finite resolution achievable by the detectors given the high-intensity environment [20]. These effects are illustrated in Figure 1.5, which shows the expected momentum distributions for conversion and DIO electron tracks in COMET Phase-I.

Nuclear capture

A muon in proximity with a nucleus may also be captured via W -boson exchange:

$$\mu^- + N(A, Z) \rightarrow \nu_\mu + N(A, Z - 1). \quad (1.6)$$

This process is similar to an electron capture and transmutes the nucleus into a potentially unstable isotope, which can be the source of proton, photon and neutron emissions as it readjusts. Particles emitted by the nucleus after muon capture may fly into the detector system and cause unwanted occupancy of the readout. This is especially concerning if the detector is close to the target, as is the case in Phase-I. Hence, the

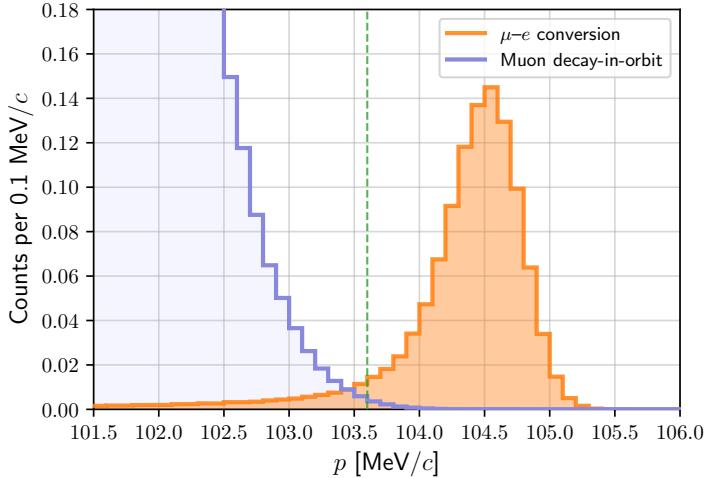


Figure 1.5: Momentum distribution of reconstructed signal and decay-in-orbit electrons in COMET Phase-I, assuming a conversion branching ratio of 3×10^{-15} . A momentum resolution of $200 \text{ keV}/c$ is assumed, and detector efficiency factors are extracted from the technical design report [20]. The vertical dashed line shows the lower bound of the momentum window at $103.6 \text{ MeV}/c$, which is used to discriminate background events from the conversion signal.

AlCap experiment was performed in 2013 at the Paul Scherrer Institute to measure the spectrum and yield of particles emitted after capture of stopped muons by aluminium nuclei [26]. The yield of highly-ionising particles was found to be half as much as the previous expectation, deduced from observations of nuclear capture by silicon nuclei, and the amount of detector shielding required in the Phase-I design was subsequently reduced.

Radiative muon capture (RMC) occurs when a photon is emitted during the capture process. Kinematically, the photon can carry energies close to the muon mass. Hence, if it goes on to produce an electron–positron pair with asymmetric momenta, the electron will mimic the conversion signal. The high-energy endpoint of the RMC spectrum is not as high as for DIO, hence an adequate momentum resolution from a tracking detector will discriminate most RMC electrons from the signal. In COMET Phase-I, the expected background count from RMC is 5 times lower than that from DIO [20].

Sources of background other than DIO and nuclear capture are expected in μ - e conversion-searching experiments, such as beam-related and cosmic ray-induced events. All backgrounds specific to the COMET experiment, and the associated design choices which were made to minimise their occurrence, are discussed in Section 2.7.

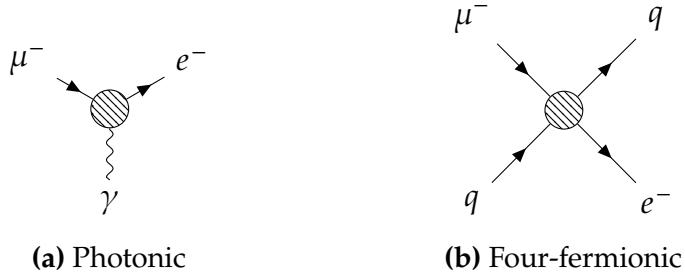


Figure 1.6: New vertices which arise in the effective Lagrangian of Equation 1.7 and allow CLFV. The photonic interaction can also mediate μ - e conversion by making the photon interact with an external quark.

1.5 Effective CLFV and the scale of new physics

Experiments searching for CLFV are sensitive to a wide variety of new physics, including a non-minimal Higgs, a Z' boson, leptoquarks, heavy neutrinos, and supersymmetric particles. In order to determine the scale of new physics to which future CLFV-searching experiments will be sensitive and the complementarity between experiments, we can consider a low-energy effective field theory derived from new interactions with generic massive ($m > 1$ TeV) particles. After integrating out heavy fields (see e.g. [27, Chapter IV]), one obtains the following effective Lagrangian [11], which allows CLFV to be mediated by the tree-level vertices shown in Figure 1.6:

$$\begin{aligned} \mathcal{L}_{\text{CLFV}}^{\text{eff}} = & \frac{1}{\kappa+1} \frac{m_\mu}{\Lambda^2} \bar{\mu}_R \sigma_{\mu\nu} e_L F^{\mu\nu} + \text{h.c.} \\ & + \frac{\kappa}{\kappa+1} \frac{1}{\Lambda^2} \bar{\mu}_L \gamma_\mu e_L (\bar{u}_L \gamma^\mu u_L + \bar{d}_L \gamma^\mu d_L) + \text{h.c.} \end{aligned} \quad (1.7)$$

where both interactions are suppressed by factors of Λ , the energy scale of the new physics, and κ determines whether the preferred channel is photonic ($\kappa \rightarrow 0$) or four-fermionic ($\kappa \rightarrow \infty$). The κ parameter allows this Lagrangian to be model-independent, i.e. it gives freedom for the new interaction to either enhance $\mu \rightarrow e\gamma$ or μ - e conversion more strongly. Searches for $\mu \rightarrow e\gamma$ will be more sensitive to CLFV than μ - e conversion searches if $\kappa \ll 1$, whereas the opposite is true if $\kappa \gg 1$.

From the estimated sensitivity of a future experiment, we can estimate the maximum scale of new physics Λ which the experiment will be able to probe. COMET Phase-II and Mu2e, which have a single-event sensitivity of around 10^{-17} [20, 21], will probe energy scales Λ up to 4000 TeV if κ is small, and up to 7000 TeV if κ is large [28, 29].

Even in the situation where no positive signal is observed, probing these enormous energy scales will heavily constrain any model which predicts new heavy particles whose interactions allow any significant amount of CLFV. On the other hand, if CLFV is observed in either the $\mu \rightarrow e\gamma$ or $\mu-e$ conversion channels, the value of κ can then be determined from a measurement in the other channel. This data will then further constrain and disambiguate models, and help to pinpoint the true nature of the new physics.

Chapter 2

The COMET Experiment

COMET (COherent Muon-to-Electron Transition) is a future muon-beam experiment designed to search for the muon-to-electron conversion process [20]. It is currently under construction at the Japan Proton Accelerator Research Complex (J-PARC) facility in Tokai, Japan. The goal of COMET is to be 10 000 times more sensitive to μ - e conversion than the current world-leading limit set by the SINDRUM II experiment [30].

Requirements

In order to reach its goal, the COMET experiment is designed with strict requirements defined to make the conversion signal as clear as possible, while efficiently rejecting background events. The essential requirements that define the COMET experiment are:

- An intense muon source to probe the rare conversion process;
- A pulsed beam such that timing information can be used to reject backgrounds;
- Strict selection of charge and momentum of beam particles prior to reaching the detector;
- A tracking detector to search for the 104.97 MeV conversion signature.

These requirements and the design choices that were made to address them are described in more detail throughout this chapter.

Strategy

COMET is planned to run in a staged approach so that the properties of the newly designed beam can be finely understood before making the measurement. COMET Phase-I has a double purpose, each fulfilled by a distinct detector system. The StrECAL detector, composed of a straw-tube tracker and electromagnetic calorimeter, will study the beam composition and timing properties and increase our understanding of potential backgrounds. The Cylindrical Detector, composed of a cylindrical drift chamber and a trigger hodoscope, will be used to perform a $\mu-e$ conversion search with a single-event sensitivity (see Section 2.8) of 3×10^{-15} , a factor-100 improvement over SINDRUM II. COMET Phase-II is a planned upgrade to Phase-I with higher beam intensity and better background rejection via a longer momentum-selecting beamline. Phase-II aims to improve on the single-event sensitivity of Phase-I by another factor 100 to reach a single event sensitivity of 2.6×10^{-17} .

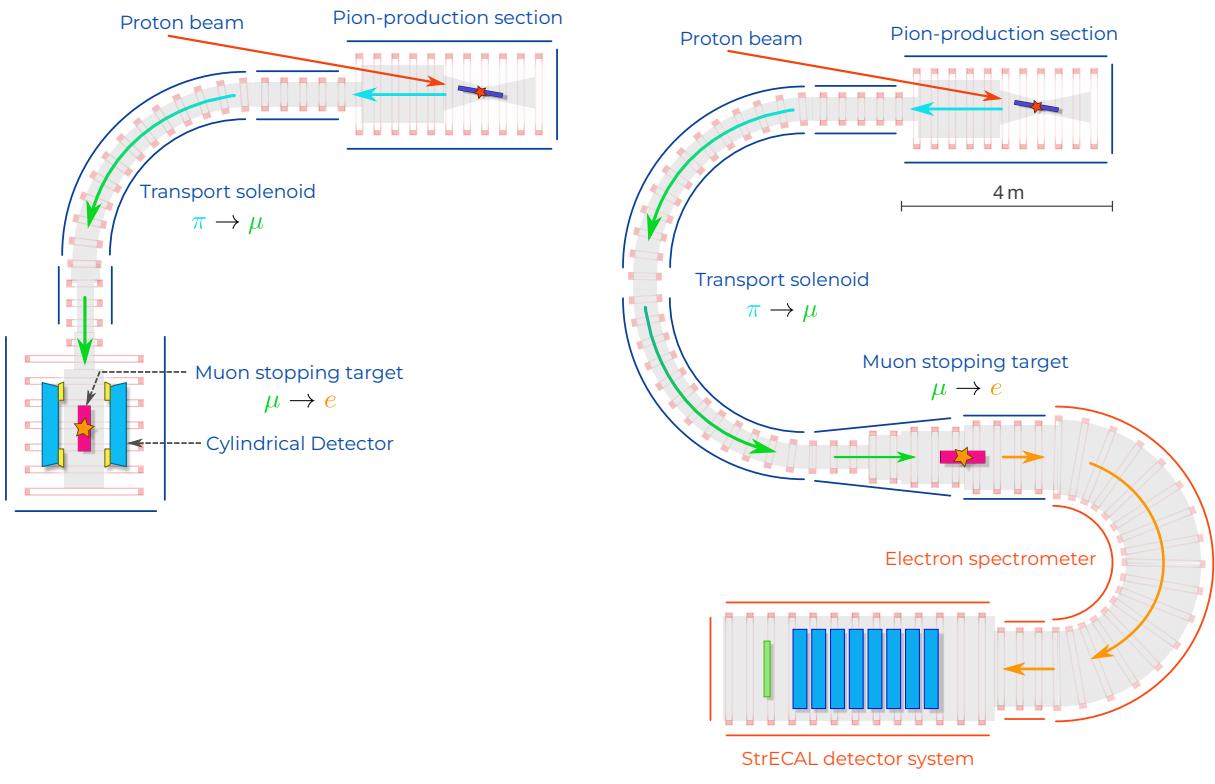
Design

Figure 2.1 shows a top-down schematic view of the COMET experiment, laying out the different sections that make up the beamline in Phase-I and Phase-II. In the latter, the transport solenoid is extended to allow more pions to decay into muons while tightening the momentum selection further. An additional curved solenoid, the *electron spectrometer*, further eliminates particles whose momenta do not match the 104.97 MeV conversion signature before they enter the detector system. While Phase-I will use the StrECAL to study the beam properties, Phase-II will use it as the conversion-searching detector system.

The following sections describe in more detail each component of the COMET beamline.

2.1 Proton beam

Muons in the COMET experiment are produced from the decay of pions created in proton collisions on a static solid target. The primary proton beam is provided by the J-PARC Main Ring synchrotron. Protons are delivered with an energy of 8 GeV, which is chosen to optimise between a high pion yield and a low yield of anti-protons, which



(a) Phase-I with the Cylindrical Detector.

(b) Phase-II.

Figure 2.1: Schematic top-down view of the COMET experiment. The beam pipe is represented in grey, and the light red rectangles along the beamline represent the solenoids that generate the magnetic field. Curved solenoids additionally help to select charge and momentum, as discussed in Section 2.3.

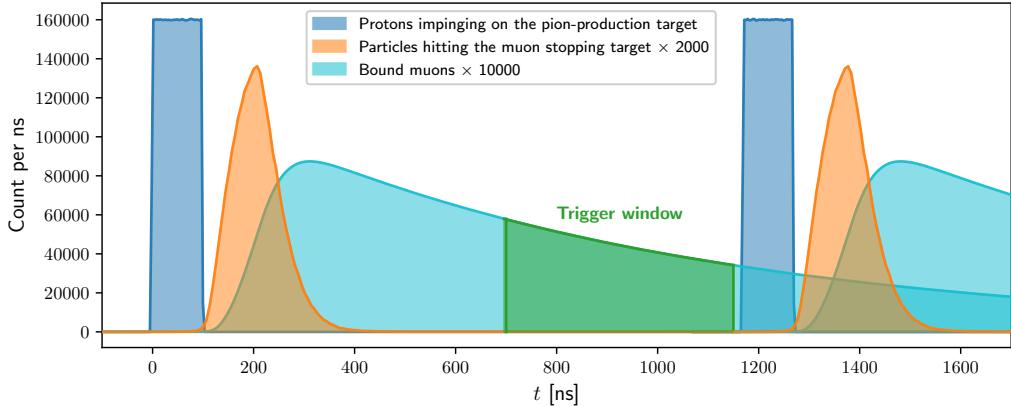


Figure 2.2: Timeline of a COMET event. Shortly after the proton collision, the beam flash floods the detector and muons begin to be bound in the muon stopping target. The Phase-I trigger window shown here starts at $t = 700$ ns and lasts until the next proton bunch collision. Electron tracks with a momentum around $105\text{ MeV}/c$ observed within this timing window are considered as signal candidates.

are a potential background source.

The timeline of a COMET event is shown in Figure 2.2. The beam has a pulsed time profile: protons are grouped into 100 ns bunches, each containing 16×10^6 protons. Bunches are separated by 1170 ns, and every fifth bunch is delayed by an additional 585 ns. Just after the collision, secondary particles will quickly move down the COMET beamline and produce numerous background hits in the detector system. This prompt and intense flooding of the detector is called the *beam flash*, and typically dies down within a few hundred nanoseconds. Muons bound by the muon stopping target have a lifetime of 864 ns (see Section 2.4). This, combined with the 1170 ns time span between two bunches, allows COMET to search for $\mu-e$ conversion after the beam flash has ended, and until the next collision occurs.

Stray protons arriving in the time between two bunches can contribute to the experimental backgrounds by sending particles toward the detector region at unexpected timings. COMET requires the J-PARC proton beam to have fewer than one such stray proton for every 600 bunches in order to reach its sensitivity goals. This corresponds to an extinction factor

$$R_{\text{extinction}} \equiv \frac{\text{protons between bunches}}{\text{protons per bunch}} \approx 10^{-10}. \quad (2.1)$$

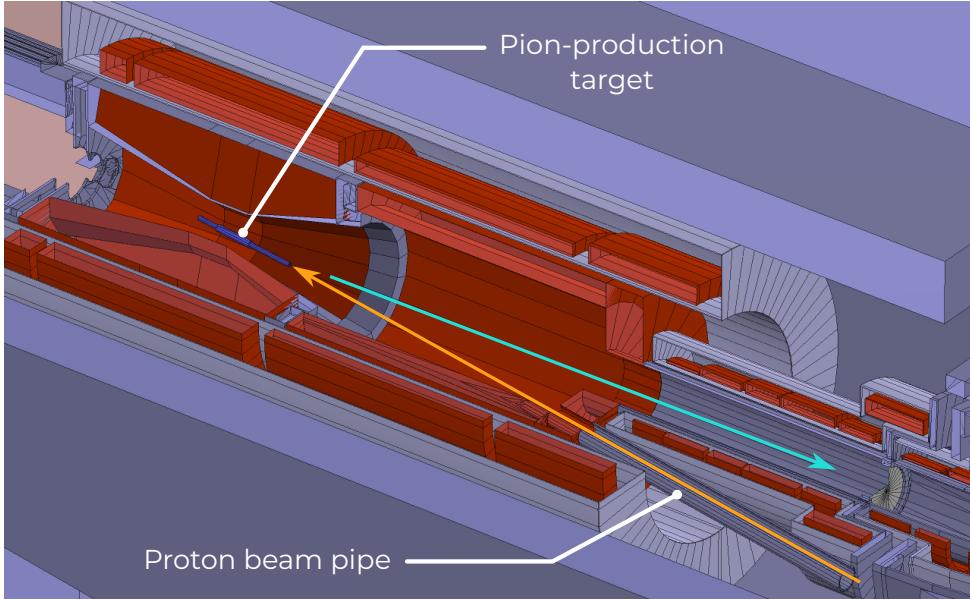


Figure 2.3: Cutaway view of the pion-production section. The orange arrow indicates the path of the proton beam while the teal arrow shows the direction of backward-going pions captured by the magnetic field.

2.2 Pion-production section

Pions are produced by the collision of the proton beam on a solid target made of graphite in Phase-I, and tungsten in Phase-II. This region is permeated by a 5 T magnetic field generated by a superconducting solenoid, which confines the pions and directs them toward the transport solenoid. Figure 2.3 shows a cutaway view of this region.

Pions produced moving backward with respect to the proton beam have a lower energy than those produced going forward, although they are not as numerous. In COMET, it is crucial to eliminate high-energy particles in the muon beam that could produce secondaries mimicking the conversion signal. Hence, the beamline is positioned in the opposite direction to the proton beam such that only low-energy, backward-moving pions are allowed into the COMET beamline. Figure 2.4 illustrates this by showing that pions moving backward with respect to the proton beam have a much lower momentum cut-off than forward-moving pions.

2.3 Transport solenoid

The transport solenoid is a curved magnet connecting the pion-production section to the muon stopping section. Its purpose is twofold. Firstly, it allows a larger fraction of

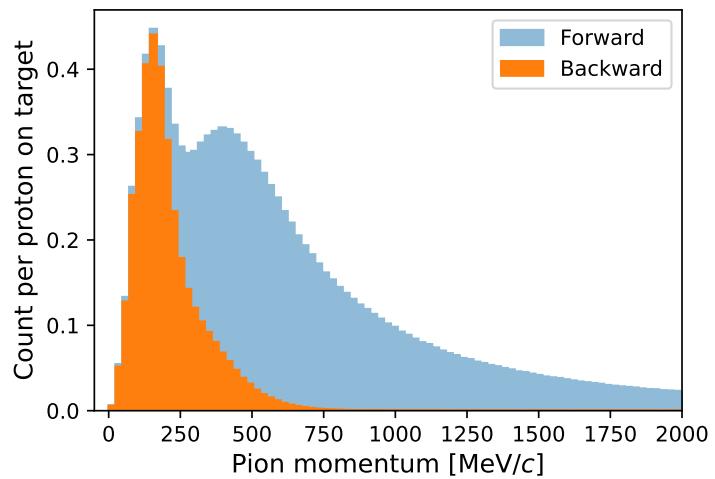


Figure 2.4: Momentum distribution of pions produced by simulating proton collisions with Geant4, depending on whether their initial direction is forward or backward with respect to the proton beam direction.

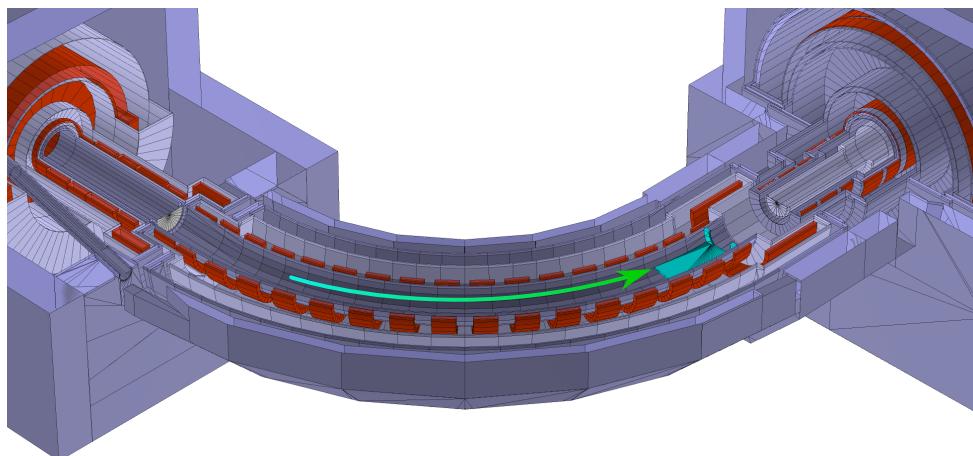
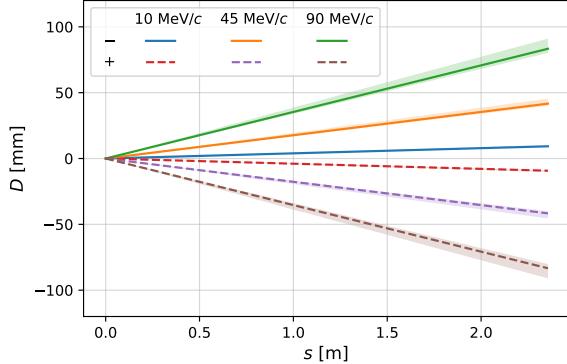
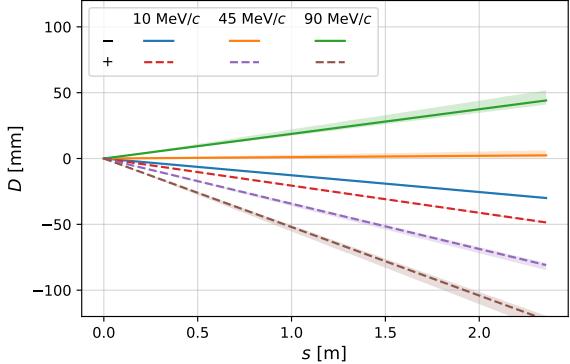


Figure 2.5: Cutaway view of the Phase-I transport solenoid. The curved solenoid (in red) combined with collimators (in teal) select particles depending on their charge and momentum.



(a) No dipole field.



(b) 0.05 T vertical dipole field.

Figure 2.6: Effective drift of particles of various momenta and either charge as they progress along the Phase-I transport solenoid, calculated using Equation 2.2. The pitch angle is assumed to be 45° and the bands show the effect of a $\pm 10^\circ$ difference. As shown on the right, a 0.05 T dipole field allows negative particles around 45 MeV/c to stay on axis while higher- and lower-momentum particles drift in opposite directions and positive particles all drift one way. By use of collimators at the end of the transport section, only particles of the right momentum and charge are allowed through.

pions to decay along the length of the beamline. Secondly, the curved shape combined with its magnetic field and collimators allows it to select negatively-charged particles of a specific momentum.

The magnetic field of a curved solenoid is slightly stronger on the inside of the curve than on the outside. Since charged particles follow helical trajectories, this has the net effect of making them drift vertically and the amount of drift D depends on momentum p according to the equation

$$D = \frac{1}{qB} \frac{s}{R} \frac{2p_L^2 + p_T^2}{2p_L}, \quad (2.2)$$

where q is the charge, B is the strength of the field along the gyration axis, s is the distance travelled along the solenoid, R is the radius of the curve, and p_L and p_T respectively denote momentum longitudinal and transverse to the solenoid axis [28]. From this expression, one can see that oppositely charged particles drift in opposite directions, and that drift is overall stronger for higher-momentum particles. The ratio $\frac{p_L}{p_T}$, which defines a helical trajectory's *pitch angle*, is also a major factor in a particle's drift.

The drift caused by the curved solenoid makes all particles of the same charge move in the same direction to varying degrees depending on momentum. In order to select

particles of a specific momentum, a vertical component is added to the magnetic field to counterbalance the drift. Selected particles thus stay on axis, while higher- and lower-momentum particles drift off axis. With the addition of a collimator at the end of the transport solenoid, particles with unwanted momenta are effectively eliminated from the beam.

Figure 2.6 shows drift D as a function of s , the longitudinal distance travelled by particles along the solenoid. This illustrates how, by adding a dipole field, particles with a momentum outside the required range can be efficiently eliminated by collimators at the top and bottom of the beam pipe.

2.4 Muon stopping target

The purpose of the muon stopping target is to slow down and stop as many muons as possible while not blocking the path of converted electrons. It is composed of a series of 17 thin aluminium disks placed in the way of the muon beam. The disks are 20 cm in diameter, 0.2 mm thick, and separated by 5 cm. The stopping target is shown in Figure 2.8, surrounded by the Cylindrical Detector.

The more aluminium there is, the higher the number of muons that will be stopped and allowed to undergo conversion. However, more material also means more energy lost by electrons flying outward, hence the design of the target optimises between muon stopping rate and acceptance of conversion electrons by the detector system.

The material of the stopping target influences the conversion rate, but also the lifetime of a muon caught in orbit around a nucleus. A heavy target favours the expected rate of $\mu-e$ conversion, however it also causes the nuclear capture rate to be higher. In COMET, beam bunches are separated by 1.17 ns and prompt backgrounds typically die off within a few hundred nanoseconds. A muonic atom with an iron or heavier nucleus has a lifetime less than 200 ns [28], which would be too short to allow muons to stay bound and convert after the beam flash is over. Hence, a light target such as aluminium, with a longer stopped muon lifetime of 864 ns [31], is better suited to the COMET conversion search.

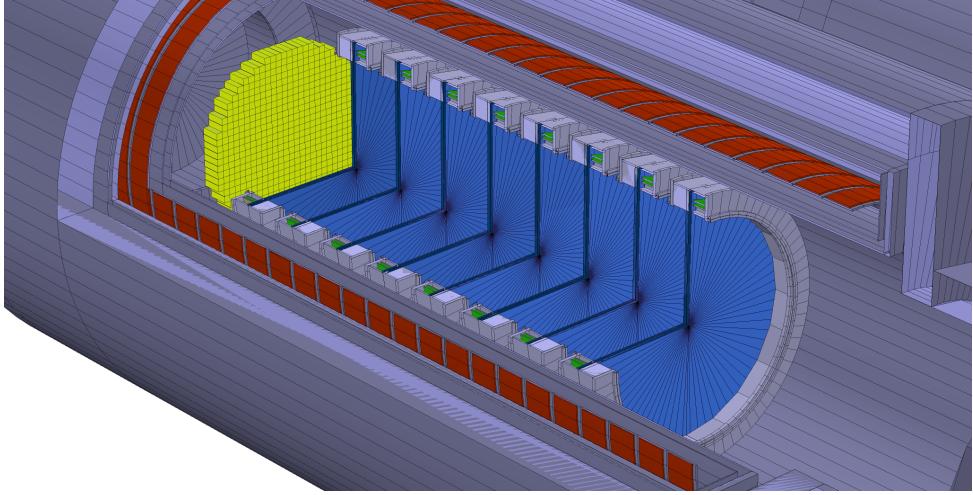


Figure 2.7: Cutaway view of the StrECAL detector. The Straw Tracker stations are shown in blue and the ECAL in yellow.

2.5 Detector systems

2.5.1 StrECAL

The StrECAL combines a straw-tube tracking detector with an electromagnetic calorimeter for energy measurement. In COMET Phase-I, the StrECAL will serve as a beam characterisation apparatus. It will be placed directly at the end of the transport solenoid, without a muon stopping target, to study the composition of the COMET beam and gain a more thorough understanding of potential backgrounds. The collected data can also serve to validate and refine the Monte Carlo simulation in preparation for the conversion measurement.

In COMET Phase-II, the StrECAL will serve as the detector system for the conversion measurement. It will be placed after the electron spectrometer, a curved solenoid section designed to select conversion electrons, as shown in Figure 2.1. Figure 2.7 shows a rendering of the StrECAL detector system in Phase-II.

The Straw Tracker uses long, thin-walled PET tubes as gaseous drift chambers. The tubes are first arranged into circular planes, four of which constitute a station. Each station is able to measure the horizontal and vertical position of a particle and has a spatial resolution of $120\text{ }\mu\text{m}$. Using time-of-flight information, the Straw Tracker can therefore reconstruct trajectories and estimate particle momenta.

The ECAL is a crystal electromagnetic calorimeter which supplements the Straw Tracker in measuring energy and thus in identifying electrons. The ECAL uses

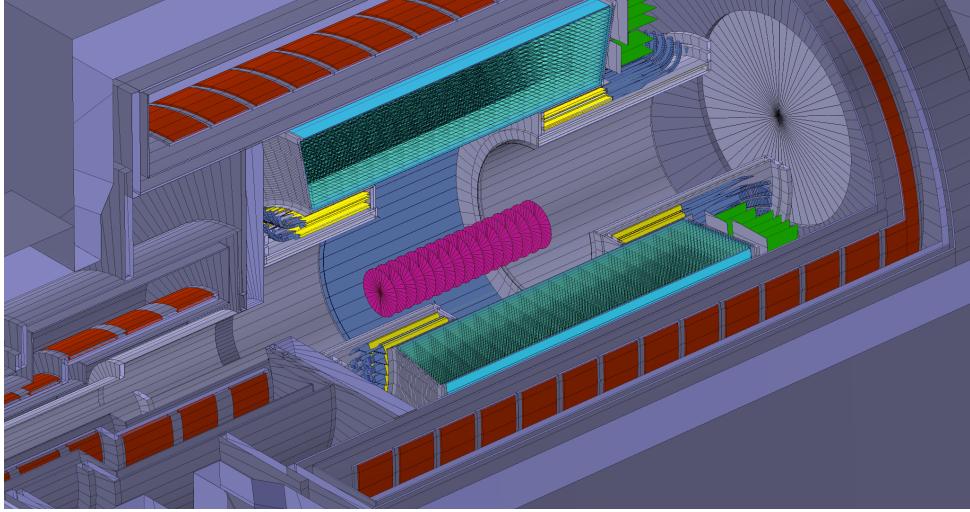


Figure 2.8: Cutaway view of the Cylindrical Detector, composed of the Cylindrical Drift Chamber (in teal) and Cylindrical Trigger Hodoscope (in yellow). The muon stopping target disks, in purple, sit in the centre of the detector system.

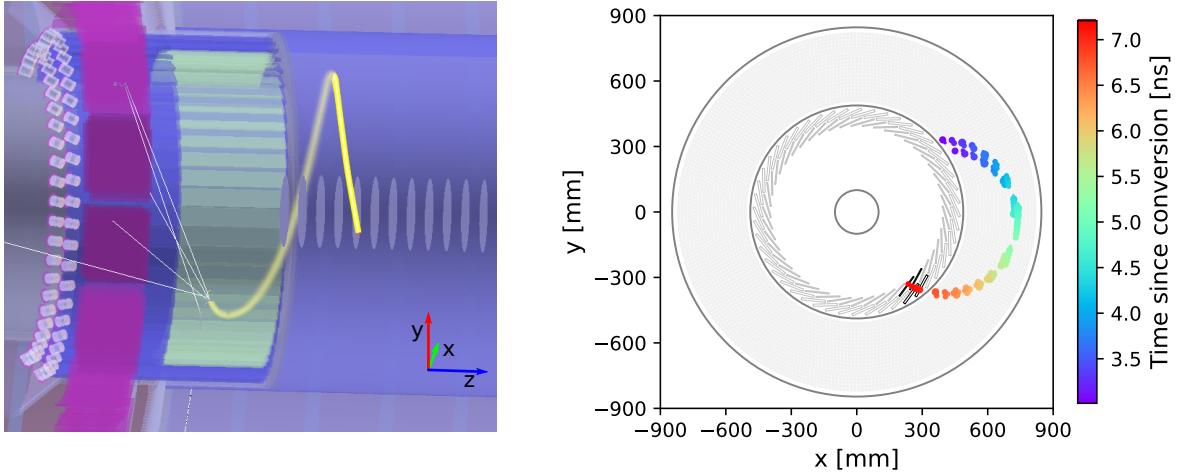
lutetium-yttrium oxyorthosilicate (LYSO) as its scintillating crystals and avalanche photodiodes to collect the emitted photons.

2.5.2 CyDet

The Cylindrical Detector (CyDet) consists of a Cylindrical Drift Chamber (CDC) for tracking and a Cylindrical Trigger Hodoscope (CTH) for triggering on specific event signatures. The CyDet surrounds the muon stopping target and sits within the detector solenoid which generates a 1 T longitudinal magnetic field. This configuration, shown in Figure 2.8, is designed to eliminate backgrounds from the beam itself as well as low-momentum products of the collision while maximising the acceptance of conversion electrons. Figure 2.9 additionally shows the signature of a simulated conversion electron inside the CyDet system.

Cylindrical Drift Chamber

The CDC is a drift chamber used to track charged particles emanating from the muon stopping target. In order to suppress hits from beam particles, the CDC is wrapped around the beam pipe and has an inner radius of 50 cm. Combined with the 1 T longitudinal magnetic field, this prevents charged particles with a transverse momentum less than $60 \text{ MeV}/c$ from reaching the chamber. In order to achieve the sensitivity goal



(a) Event shown in DisplayCore, the 3D event display of ICEDUST (see Chapter 3). (b) CyDet event display showing the timing of hits since conversion.

Figure 2.9: Conversion electron trajectory as observed by the Cylindrical Detector. The effect of stereo angles is visible at the start of the trajectory where hits have alternating azimuthal positions around the actual electron track. Upon reaching the CTH, the electron produces consecutive hits in four adjacent counters, satisfying the fourfold coincidence trigger criterion.

for Phase-I, the CDC has a momentum resolution better than $200 \text{ keV}/c$ in order to differentiate between conversion electrons and electrons from the high-energy tails of the decay-in-orbit and radiative muon capture spectra.

The CDC contains 4986 *sense wires* strung out parallel to its axis in 20 concentric layers. The wires in each layer are rotated slightly off from the longitudinal axis, and the rotation is alternatively clockwise and anti-clockwise from one layer to the next. This special property, called the *stereo angle*, allows the drift chamber to stereoscopically reconstruct, within 3 mm, the longitudinal position of a particle [29].

Each sense wire is held at a potential of up to 1900 V and is surrounded by 8 grounded *field wires* to generate an inward electric field. When a charged particle ionises the gas, the field accelerates freed electrons toward the sense wire. These electrons can gain enough energy to further ionise the gas, leading to avalanche multiplication (see e.g. [32, Chapter 6]). The avalanche produces a pulse on the sense wire, which is acquired by the readout electronics.

Cylindrical Trigger Hodoscope

The CTH consists of two *modules* that line the inner wall of the CDC, one upstream and one downstream of the muon stopping target. Each module contains two layers of 48 partially-overlapping scintillation counters. Each counter has a time resolution of 1 ns. The main purpose of the CTH is to reject background events coming from the beam itself and from products of the muon beam collision while complementing the CDC in identifying conversion electrons.

The overlap between counters allows the CTH to reject a large fraction of background hits, e.g. from photons produced in the muon beam collision. This is done by only accepting fourfold-coincident events, where four neighbouring counters (two in the inner layer, two in the outer layer) are hit in a short 10 ns time span. The CTH thus provides an online triggering mechanism which significantly reduces the number of background events accepted by the CyDet system due to its proximity with the muon stopping target. Figure 2.9 shows how a conversion electron might produce such a fourfold coincidence.

2.5.3 Cosmic ray veto

The COMET experiment hall, being at sea level, is constantly irradiated by muons produced in the atmosphere by cosmic rays. The cosmic ray veto (CRV) is an additional active detector which will tightly enclose the CyDet and StrECAL detector systems. Its purpose is to identify events induced by atmospheric muons rather than by the COMET beam, and thus reduce the probability that such events will be mistaken for a conversion signal.

2.6 Conversion signature

Conversion electrons are mono-energetic at $E = 104.97 \text{ MeV}$ and emitted isotropically by muons stopped in the target disks. In the magnetic field of the detector solenoid, they have a helical trajectory which can be reconstructed by the CDC or Straw Tracker. Figure 2.9 shows a simulated conversion electron going through the Cylindrical Detector.

Background source	Intrinsic	Stray protons	Antiprotons	Cosmics	Total
Estimated events	0.014	0.007	0.001	0.010	0.032

Table 2.1: Expected number of background events from each potential source [20]. The total count is much smaller than one, such that the observation of a single signal electron could be evidence that charged lepton flavour is violated.

The expected timing of conversion electrons is directly related to the time distribution of muons bound inside the stopping target, shown in Figure 2.2. In Phase-I, in order to suppress prompt backgrounds from the beam flash, the detector is tuned to only start triggering on events that occur at least $t = 700\text{ ns}$ after each proton collision. This acceptance criterion retains only 30 % of all signal events because the majority of bound muons, having an average lifetime of 864 ns , will have undergone decay or capture before the start of the window.

2.7 Experimental backgrounds

Intrinsic backgrounds from muon decay-in-orbit and nuclear muon capture can mimic the conversion signal, as discussed in Section 1.4.1. The COMET experiment is also affected by experimental backgrounds caused by the intense beam as well as cosmic rays.

Beam-induced backgrounds include delayed events caused by slow particles and products of stray protons arriving between two bunches. Antiprotons are the main source of delayed backgrounds due to their slow speed relative to pions and muons of the same momentum. They are negatively charged, therefore they cannot be efficiently filtered out by the beamline. Hence, they can move slowly toward the muon stopping target and produce signal-like secondary electrons during the trigger window.

Stray protons, i.e. protons from the proton beam that arrive at the pion-production target in the interval between bunches, can produce secondary particles that reach the detector during the trigger window. To reduce the background rate from this type of event, COMET requires a strict extinction factor $R_{\text{extinction}} = 10^{-10}$ from the J-PARC beam.

Finally, cosmic rays are a major source of background events. The background rate depends heavily on the amount of shielding and material above the COMET detector

system, as well as the efficiency of the cosmic ray veto. The topic of rate estimation for cosmic ray-induced backgrounds is discussed more thoroughly in Chapter 5.

Table 2.1 shows the rates estimated in the COMET Phase-I technical design report (TDR) [20, Section 10.6] for each source of background in the $\mu-e$ conversion search. The total number of background events over the data acquisition run time is predicted to be $N_{\text{background}} = 0.032$ given an extinction factor $R_{\text{extinction}} = 3 \times 10^{-11}$. Since $N_{\text{background}} \ll 1$, the observation of a single signal electron would suggest that there is a CLFV process at play, which would further motivate this search through COMET Phase-II and beyond.

2.8 Sensitivity and run time

The sensitivity of the COMET experiment is conventionally expressed as a *single event sensitivity* (SES), which is defined as the value of the $\mu-e$ conversion branching ratio for which COMET expects to observe one event (smaller is better). SES takes into account the net acceptance of signal events by the detector system, however it does not say anything about the experimental backgrounds. When using SES as a figure of merit, a study of potential background sources is usually necessary to show that the expected number of background events is less than one.

The total number of coherent $\mu-e$ conversions produced by a population of N_μ bound muons can be expressed as

$$N_{\text{conversion}} = N_\mu \cdot \mathcal{B}_{\text{conversion}} \cdot \mathcal{B}_{\text{capture}} \cdot f_{\text{coherent}}, \quad (2.3)$$

where $\mathcal{B}_{\text{conversion}}$ is the conversion branching ratio normalised to the branching ratio of nuclear muon capture $\mathcal{B}_{\text{capture}}$, and f_{coherent} is the fraction of conversions estimated to occur coherently and leave the nucleus in its ground state. The number of conversion electrons that will be observed by the detector is then $N_{\text{obs}} = N_{\text{conversion}} A_{\mu-e}$, where $A_{\mu-e}$ is the detector's net signal acceptance. If we require $N_{\text{obs}} = 1$ and rearrange to find the corresponding value of $\mathcal{B}_{\text{conversion}}$, we obtain the single event sensitivity:

$$\text{SES} \equiv \mathcal{B}_{\text{conversion}}^{N_{\text{obs}}=1} = \frac{1}{N_\mu A_{\mu-e} \mathcal{B}_{\text{capture}} f_{\text{coherent}}}. \quad (2.4)$$

Factor Efficiency	Geometrical 26 %	Hardware 81 %	Track-finding 99 %	Cuts 70 %	Momentum 93 %	Timing 30 %	Net 4.1 %
----------------------	---------------------	------------------	-----------------------	--------------	------------------	----------------	--------------

Table 2.2: Efficiency factors used to determine the signal acceptance in the COMET Phase-I TDR [20]. The net acceptance is the product of all efficiency factors. “Cuts” refers to track quality cuts, used to reject events with irregular tracks that cannot be accurately fitted and reconstructed.

In COMET Phase-I, the signal acceptance can be broken down into seven efficiency factors: geometrical acceptance, hardware (trigger and data acquisition), track finding, track reconstruction quality cuts, momentum window and trigger time window. Table 2.2 lists these factors as they were estimated in the COMET Phase-I TDR [20, Section 10.1].

From the net signal acceptance $A_{\mu-e} = 4.1\%$, we can now estimate the total data acquisition run time required to reach the sensitivity goal of the experiment, $\text{SES} = 3 \times 10^{-15}$, using Equation 2.4. The required number of bound muons is thus $N_\mu = 1.5 \times 10^{16}$. This quantity can be related to the total run time T via the proton beam current $I_p = 0.4 \mu\text{A}$ and the yield of stopped muons per collision $R_{\mu/p} = 4.7 \times 10^{-4}$:

$$N_\mu = T \cdot \frac{I_p}{e} \cdot R_{\mu/p}. \quad (2.5)$$

This equation is rearranged for T to yield $T = 146$ days of data acquisition for a Phase-I SES of 3×10^{-15} .

Chapter 3

Software and Simulation

Experiments in high energy physics are commonly accompanied by the development of a set of software tools to help in manipulating and analysing experimental data. Simulations are used extensively to optimise experimental designs and prepare for the collection of real data while the physical instruments — detectors, magnets, readout electronics and data acquisition systems — are being manufactured and assembled.

3.1 The ICEDUST Software Suite

The COMET collaboration develops a suite of software packages named ICEDUST (Integrated COMET Experimental Data User Software Toolkit) to satisfy the offline data-processing requirements of the experiment. Originally forked in 2013 from the software written for T2K’s near detector system ND280, ICEDUST supports the needs of the COMET experiment with regard to simulation, on-disk data format, calibration, reconstruction and analysis. Figure 3.1 shows the flow of real and simulated data within the ICEDUST framework, and lists the names of the key packages for each step.

3.1.1 Data format

A key design principle of ICEDUST is to define a common format adopted by both real and simulated data, such that the pipeline of calibration, reconstruction and analysis can be thoroughly tested with only simulation data, in advance of the data acquisition stage. This offline data format is called `oaEvent`. Based on the ROOT [33] serialisation system, `oaEvent` defines a file format where data is laid out as a sequence of events,

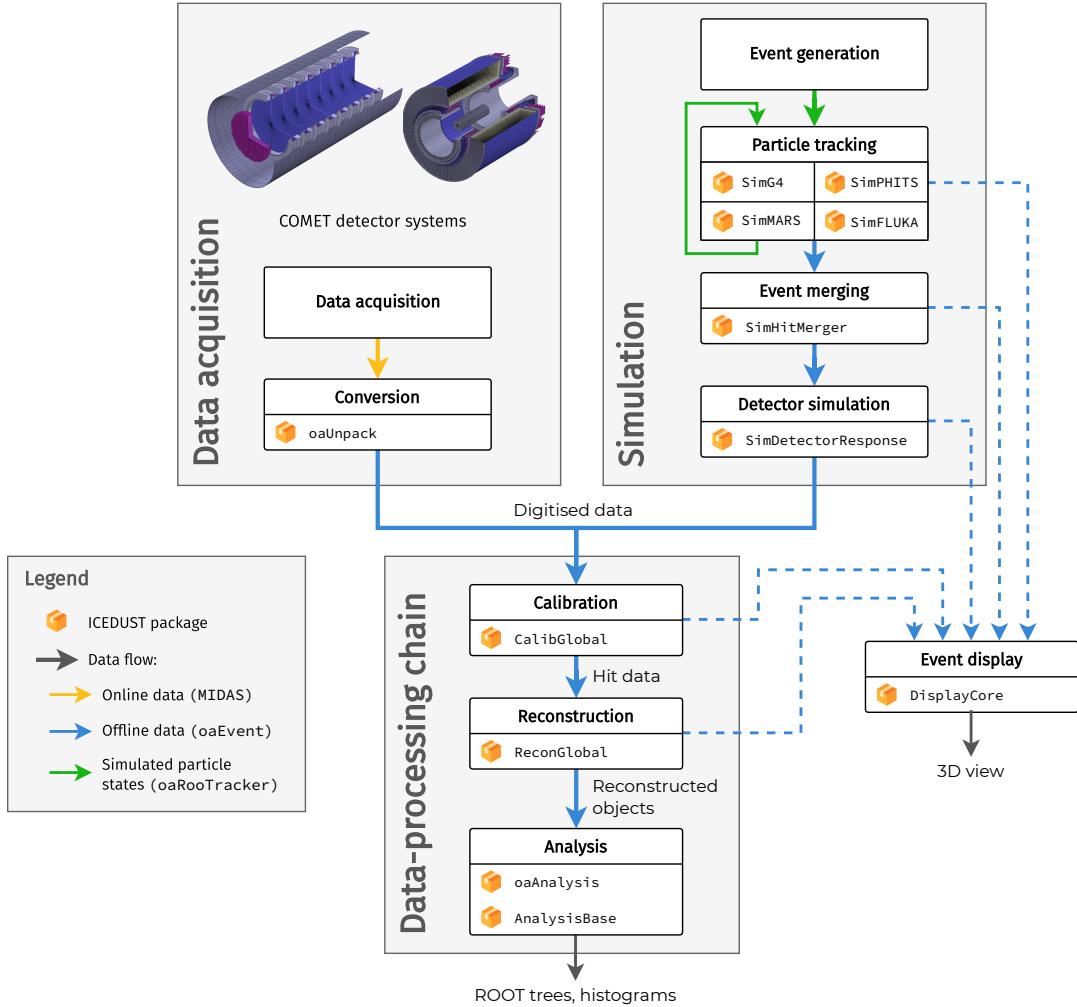


Figure 3.1: Data flow in the ICEDUST framework. The colour of arrows represents the data format. By design, simulated detector data and real data share a common format such that they can be processed identically by the calibration, reconstruction and analysis stages.

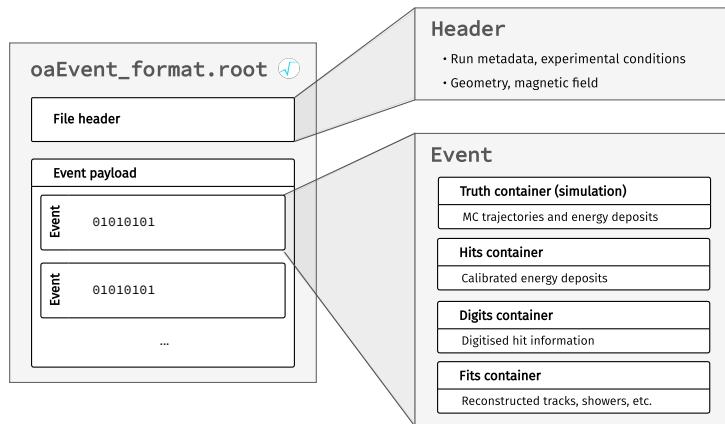


Figure 3.2: Layout of oaEvent files which contain MC or real data. Blue arrows in Figure 3.1 indicate steps of the data flow where data is stored in this format on disk.

shown schematically in Figure 3.2. Each event object contains an arborescence of data containers, each of which holds an array of some given data type (e.g. detector hits, calibrated energy deposits, reconstructed tracks, etc.).

From Phase-I onward, data will be collected using the MIDAS data acquisition system, hence the online on-disk format is based on MIDAS data banks. ICEDUST provides a direct translation from this format to oaEvent via the oaRawEvent and oaUnpack packages. Once translated, the data can flow through the calibration, reconstruction and analysis stages.

3.1.2 Simulation pipeline

Simulations of the COMET experiment can be split into four steps:

1. **Event generation** takes place at the level of individual protons. A proton beam profile is drawn ahead of time using a specialised Turtle [34] simulation of the beamline. The position and momentum of the protons are histogrammed at a short distance ahead of the pion-production target. Protons are sampled from the resulting distribution to generate events, and the outcome from each primary proton is referred to as a proton-on-target (POT) event.
2. **Particle tracking** is the step-by-step propagation of particles through the geometry and electromagnetic fields of the experimental setup. The primary proton will usually produce many secondary particles in the collision, some of which might propagate to the detector region. Simulated energy deposits in the active detector volumes are recorded for the detector response simulation stage.
3. **Event bunching** allows us to model the simultaneous arrival of a proton bunch into the setup. In the real beamline, millions of protons are bunched into a short 100 ns time interval, and two consecutive bunches are separated by about 1.2 μ s. In order to simulate this time structure, a time offset is applied to each POT event, after which the events are merged into bunches. Following this, individual bunch events can be further merged into *bunch trains* to obtain an entire sequence and account for potential pileup.
4. **Detector response simulation** translates the true energy deposits simulated in the particle tracking stage into detector-specific hits. It takes into account the way

in which charge or light is collected by the detector, and can apply smearing to the observed energy and timing. This step effectively reverses the calibration process by transforming energy deposits into uncalibrated, digitised hits and waveforms.

Once all four stages of the simulation pipeline are complete, the result should closely resemble real data post-translation from MIDAS to oaEvent. This implies that it should be able to naturally flow through the calibration, reconstruction and analysis stages. At each stage, the information may be compared with the truth information available from the particle tracking stage in order to refine each process. This exercise should also outline the performance of the offline data-processing pipeline and hence provide an estimate of the computational requirements of the experiment.

3.1.3 Intermediate simulation file format

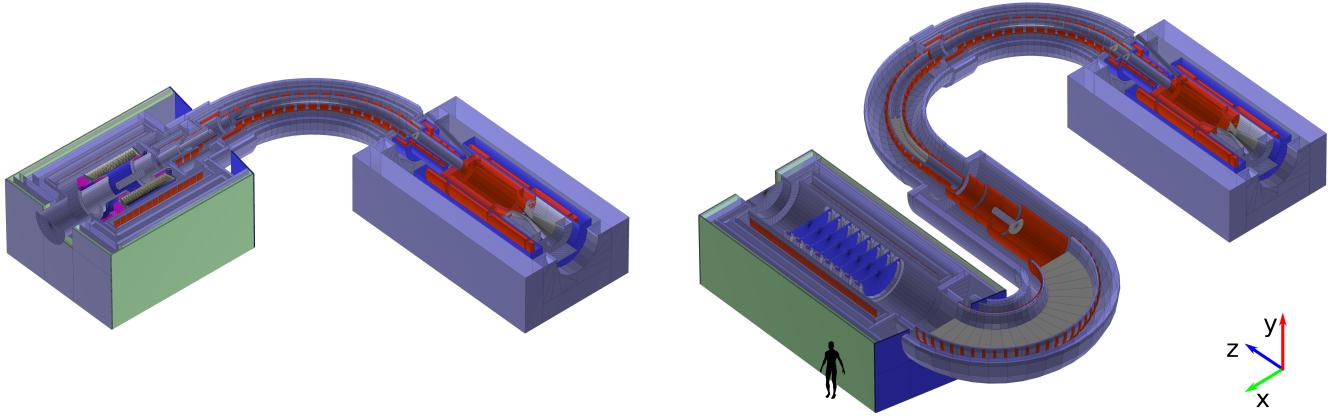
The oaRooTracker package defines another ROOT-based file format used by the simulation to save and retrieve particle states. The typical use case is for a simulation to record particles that have reached a key geometrical element (e.g. the detector region) to build up a sample of interesting events. The saved particle states can then be used as input to subsequent simulations, easing the need for a full simulation run every time a change is made to the geometry or physics models.

In addition, from a large enough sample of particles recorded at some key volume boundary, the flux distribution of inbound particles can be estimated. Events can then be sampled from this distribution to increase the statistics inside the volume of interest.

3.2 Monte Carlo simulation in ICEDUST

Monte Carlo (MC) simulation is the process through which hypothetical particles are realistically propagated through an experimental setup, with the aim of evaluating the experiment's performance and to prepare for its realisation. Currently, in COMET, MC simulations are heavily used to further optimise the experiment's design and to develop reconstruction and analysis algorithms.

Monte Carlo simulation in high energy physics can be described as the stepwise propagation (tracking) of particles through matter and electromagnetic fields. Starting from initial conditions of position and momentum, a particle advances in space and



(a) Phase-I in the CyDet configuration.

(b) Phase-II.

Figure 3.3: Cutaway views of the simulation geometries implemented in SimG4 and visualised with DisplayCore. The experiment hall is also modelled in the simulation but was hidden here for clarity.

time according to the laws of relativistic kinematics. It can change velocity or direction, and produce secondary particles via interactions with the local medium and field. The tracking of a particle ends when it decays, gets absorbed by some material, or exits the simulation space.

Interactions and decays occur stochastically according to measured cross-sections and lifetimes. Depending on the type of interaction, secondary particles may be emitted. In this case, the position and momentum of secondaries are stored in memory until the tracking of the primary particle is finished. The list of secondaries is then iterated through, propagating each one in turn. Since secondaries may produce their own secondaries, this process continues recursively until all products of the original primary particle have been accounted for.

In ICEDUST, particle tracking, event bunching and detector response simulation are handled in separate packages. Particle tracking can be performed using four different engines: Geant4 [35], FLUKA [36], MARS15 [37] and PHITS [38]. The corresponding ICEDUST packages are SimG4, SimFLUKA, SimMARS, and SimPHITS, respectively. Event bunching is handled by the SimHitMerger package, and SimDetectorResponse performs detector response simulation.

3.2.1 Geometry

The SimG4 package contains the most detailed and up-to-date definitions of the COMET simulation geometries. The various components that make up a simulation world are defined with constructive solid geometry, using the standard GEANT4 geometry classes. In addition, SimG4 extends the GEANT4 macro language to allow geometrical parameters to be defined in macro files. The position, dimensions, and material of every volume can thus be specified at run-time rather than compile-time. A custom parser attached to every volume allows interdependencies between components and supports looping, for instance to position identical elements in a regular pattern.

Once the simulation geometry is assembled, it is also translated into the ROOT TGeo representation, which simplifies its visualisation and its storage on disk as metadata inside oaEvent files. This representation can furthermore be interpreted by the MARS15 tracking software, which simplifies the process of running comparative simulations.

The SimG4 package includes geometries for the Phase-I (CyDet and StrECAL configurations) and Phase-II setups. Over time, these simulation worlds are refined by developers to better reflect the exact configuration of the experiment. Figure 3.3 shows the simulation geometries for the two conversion-searching COMET designs.

3.2.2 Physics

The COMET experiment involves many physical processes, from the initial proton collision to energy depositions in the detector. The MC simulation must faithfully account for any process which could lead to backgrounds in order to realistically estimate the experiment’s performance. Hence, the SimG4 simulation associates standard GEANT4 hadronic and electromagnetic physics models with custom code for nuclear muon capture and muon decay-in-orbit.

To simulate collisions between the proton beam and the pion-production target, SimG4 uses the QGSP_BERT_HP reference physics list, which handles hadronic interactions by modelling elastic, inelastic, capture and fission processes [39]. Electromagnetic processes are handled by the standard GEANT4 electromagnetic physics model.

Muons at rest receive a special treatment due to their important contribution toward the background rate in COMET. The default energy spectrum of electrons from muon decay-in-orbit is replaced by the 2011 numerical evaluation of Czarnecki et al. [25],

which includes the effect of nuclear recoil and thus allows electron energies up to 104.973 MeV. In addition, the nuclear muon capture (NMC) model is replaced to adhere to the results of the AlCap experiment [26] which measured the energy spectrum of protons emitted after NMC in aluminium.

3.2.3 Signal simulation

By default, the list of physical processes considered by SimG4 includes only SM-allowed interactions and decays, and hence does not include $\mu-e$ conversion. In order to simulate signal events, one must manually produce conversion electrons out of the muons stopped inside the stopping target.

From the normal beam simulation, one can estimate the spatial and temporal distributions of muons being stopped in the stopping target, and subsequently sample conversion electrons from them. In sensitivity studies, signal events are commonly overlaid onto a pure background sample in order to evaluate the signal acceptance and background rejection efficiencies of the hit filtering and track finding routines.

Figure 2.9 shows a simulated conversion electron emerging from the stopping target, depositing energy inside the CDC gas, to eventually trigger the CTH by hitting four adjacent counters.

3.2.4 Representation of simulated CDC hits

As a particle passes through a material, it tends to lose energy to the medium, e.g. through inelastic scattering or ionisation. In MC simulations, simulated energy deposits must be recorded inside active detector elements in order for us to determine the response of the detector and readout systems to the passage of the particle.

Detector elements in a GEANT4 simulation are defined as “sensitive volumes”, and energy deposits of incoming particles are accumulated and recorded as “hits”. The way in which hits are instantiated is typically dependent on the type of detector, because of differing granularities between e.g. a plastic scintillator and a drift chamber. A higher-granularity detector type requires finer-detailed information, hence more hit instances along the trajectory.

In SimG4, the data type associated with CDC hits is called `IG4HitGas`. When GEANT4 reports an energy deposit inside the CDC, an instance of that class is created. If a

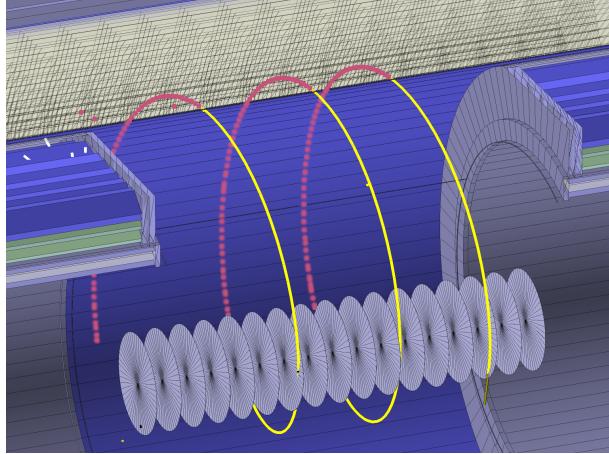


Figure 3.4: Hits, shown as red dots, produced by the CDC sensitive detector class when a particle deposits charge inside the gas. One instance of IG4HitGas is created every time the particle traverses a CDC cell, i.e. roughly every 16 mm.

particle makes multiple steps in a CDC cell, the deposit from each step is accumulated into the same hit instance, such that only one instance per particle per cell may exist. If a particle enters the same cell multiple times, one hit instance is created per entry. This is shown in Figure 3.4, where the position and granularity of IG4HitGas instances is drawn along an electron’s trajectory in the CDC.

This hit representation is designed toward gaseous detectors, hence it also applies to the Straw-Tube Tracker of Phase-I and Phase-II. Because of the relatively low density of hit instances along a trajectory, important details of the true energy deposition pattern may be lost. Therefore, SimG4 implements a run-time option to store *auxiliary points* inside each hit instance, which provide a more fine-grained description of the particle’s steps inside the cell.

3.3 Large-scale simulation: MC5

The 5th large-scale production of simulation data, MC5, was run in 2020 using computing facilities at the French National Institute of Nuclear and Particle Physics Computing Centre (CC-IN2P3) in Lyon, France. Using 2 000 concurrent machines over the course of a few weeks, the outcomes of 1 billion proton-on-target (POT) collisions were simulated with SimG4 in the CyDet configuration of COMET Phase-I.

Software

Leading up to the MC5 production, several aspects of the software were changed or updated in comparison with the previous large-scale production:

- The CMake build configuration system was integrated by Andreas Jansen to replace the legacy CMT system.
- I led the upgrade of the ROOT software, upon which the oaEvent data format depends, from major version 5 to 6.
- The CyDet geometry received multiple updates to make it as faithful as possible to the design. Detailed elements such as readout boards were added, and I contributed in tracing back and fixing inaccurate positioning of the CDC wires.
- As discussed in Section 3.2.4, I participated in changing the treatment of energy deposits in the CDC to the current IG4HitGas representation and thoroughly tested the new behaviour.
- I identified and resolved all memory-related issues in the SimG4 code using valgrind to bring the simulation software into a production-ready state.

Run configuration

The simulation was split into two stages by dividing the world along a boundary which effectively separates the pion-production section and transport solenoid (*upstream*) from the detector region (*downstream*), as shown in Figure 3.5. The boundary is set up to record the position and momentum of particles which enter the detector region. Most particles will enter via the beamline, but a small fraction (mostly neutrons) also penetrates through the wall, floor and ceiling. In the upstream run, only particles that enter the detector region are saved to disk, along with their ancestors. Upon doing so, their position and momentum are recorded to an oaRooTracker file (see Section 3.1.3) which is used as input for the downstream simulation.

This run configuration implies that the upstream simulation is unaffected by the disposition of the detector region, hence one can perform the upstream run once and use the results to seed multiple downstream configurations, changing the detector geometry (e.g. between CyDet and the StrECAL) or magnetic field. Since most of the

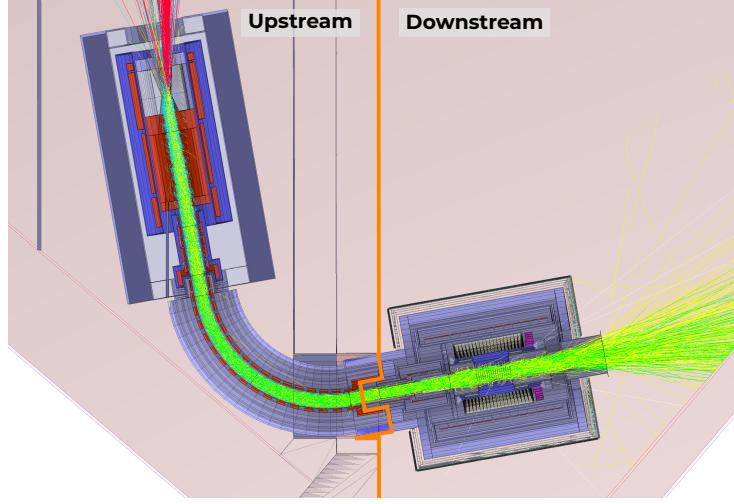


Figure 3.5: Top-down cutaway view of the running configuration for MC5, showing 1.6×10^5 overlaid events (1% of a bunch). The orange line shows the sampling boundary where particles that cross into the detector region via the beamline or wall are recorded. Only charged-particle trajectories are shown for clarity.

simulation time is spent in the pion-production section, this gives a flexible way to produce high-statistics MC datasets in multiple experimental configurations.

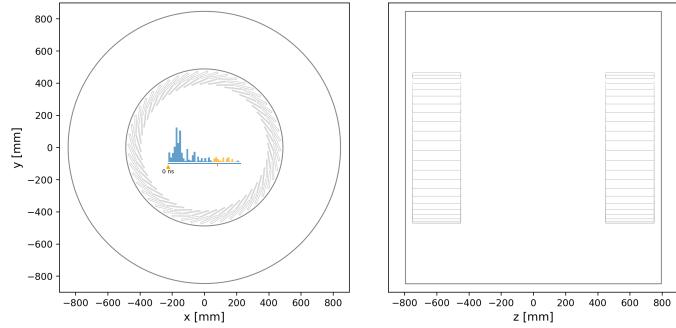
Outcome

The data produced with MC5 has a total disk size of 13 TB and is archived on the tape storage system at CC-IN2P3. The sample size of 10^9 POT events represents 62 unique beam bunches when merged, or the equivalent of about 0.1 ms of data acquisition in Phase-I.

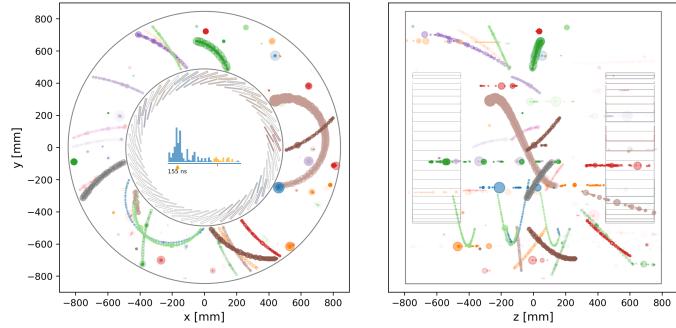
3.4 Animated CyDet event display

In order to visualise events in the CyDet system, I developed a tool to display hit data from MC simulations as animations, visually similar to a slow-motion online monitor of the detector. The software used to produce these animations has been committed to the CC-IN2P3 GitLab instance under the name `cydet_animations` [40]. Figure 3.6 shows a series of frames from one such animation involving a $\mu-e$ conversion electron.

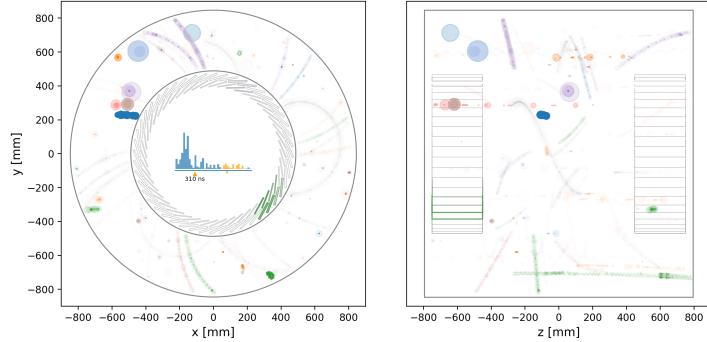
The produced animations show the CyDet system under two projections such that particle trajectories can be visualised in all three dimensions, over time. The left-hand side of the display shows the projection in the readout plane, with the CDC on



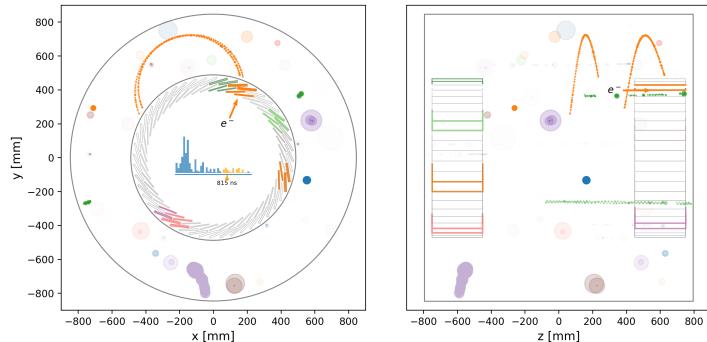
(a) $t = 0$ ns. Ignoring pileup and cosmics, the detector is clear of hits before the POT collision.



(b) $t = 155$ ns. Many tracks occupy the detector soon after the beam flash.



(c) $t = 310$ ns. The hit rate decreases significantly once the beam flash has ended.



(d) $t = 815$ ns. A conversion electron appears in the CDC and triggers the CTH.

Figure 3.6: Still frames of an animation rendered by the visualisation tool. The event shown outlines how a conversion electron would be seen by the CyDet system among background hits.

the outside and the CTH counters on the inside, while the right-hand side shows an orthogonal view with the same vertical axis. In the centre of the left-hand pane, a histogram shows the rate of hits over the event’s duration and a cursor indicates the current time in the simulation.

Each animation shows one bunch event unfolding over one cycle, i.e. 1170 ns from the collision of one bunch until the arrival of the next one. Time is slowed down by a factor of around 10^{-7} such that the event unfolds over 10 real seconds.

In the CDC, the true position of every hit is drawn as a circle with a radius proportional to the amount of energy deposited. Colour indicates hits produced by the same particle such that tracks can be disambiguated.

The CTH counters flash only in the case of a fourfold coincidence, whereby four adjacent counters are struck within a 10 ns window. Fourfold coincidences are most often caused accidentally by multiple particles, but occasionally a single track will hit four counters. In these occurrences, the track is emphasised in the animation and the particle type is displayed, as shown in Figure 3.6d.

Hits in the CDC and CTH fade out over time such that the display does not become too cluttered, but the fading rate is not representative of the actual time resolution of the sub-detectors.

The animation rendering tool is written in Python. The code includes an algorithm for finding fourfold coincidences among CTH hits. To draw the frames, it relies on the `matplotlib` package. Once the individual frames have been rendered, a bash script runs `ffmpeg` to assemble them into a video format, such as `webm`, `gif` or `mp4`.

3.5 ICEDUST development

3.5.1 Version control

The source code of the ICEDUST software project is version-controlled using Git. A shared repository is hosted on the GitLab instance of the IN2P3, where developers collaborate on building up and improving the code base. The repository contains a full history of the code, an issue tracker, and a set of wiki pages documenting the software.

Additions and changes are submitted to the central repository through merge requests from the developers. When submitting a merge request, changes to the code

are typically reviewed by another developer or maintainer who verifies that no new bugs are introduced into the main branch. To further reduce the likelihood of new issues appearing, the ICEDUST repository makes use of GitLab's continuous integration system.

3.5.2 Continuous integration and deployment

Every time new code enters the main branch, the whole code base goes through a three-step pipeline. The first step compiles the code and builds a new version of every binary. The second step runs custom unit tests and validations using this new build. Each unit test typically verifies that a single functionality works as intended in an isolated environment, while validations can run multiple pieces of the software and ensure that the results are consistent between one revision of ICEDUST and the next.

If the building, unit testing and validation stages all pass, the pipeline moves to the final deployment step where the new binaries are assembled into a Docker image which is published to the repository's container registry. Users who wish to use the compiled software as-is can download this image and run it via Docker or Singularity. If any of the pipeline steps reports a failure, the developer submitting the merge request is notified and full logs are provided to identify the issue. A merge request is usually only accepted once the code has been reviewed and if the pipeline finishes successfully.

3.5.3 Documentation

ICEDUST uses Doxygen to generate documentation of the source code. Doxygen parses the C++ code and comments to produce a set of static HTML pages containing types, functions and inheritance diagrams that can be navigated in a web browser.

In order to automate this process as the software evolves over time, the generation of the documentation was added as a step in the deployment stage of the continuous integration pipeline. Whenever a merge request is accepted, Doxygen is run on the updated code base to generate a new set of web pages containing the source code reference. This content is then pushed to the built-in *Pages* of the ICEDUST repository, a static web server hosted on the GitLab instance alongside the Git repository and wiki pages. This then provides a central, always up-to-date version of the documentation that all developers and users can access and use as reference.

Chapter 4

Data Augmentation with Generative Adversarial Networks

When performing a full simulation of COMET Phase-I, most of the activity takes place in the initial collision between the proton beam and the graphite target. The many hadronic interactions caused by the proton beam collision represent 99.7 % of the computational cost of the Monte Carlo simulation. In addition, because of the large distance between the pion production section and the detector area, only about one POT collision in a thousand will produce observable hits in the detector system.

Hence, despite being the most physically accurate means to synthesise data in the detector, simulating each proton individually is also very computationally inefficient. Using this brute-force method with the infrastructure used for MC5, it is only possible to produce the equivalent of 100 to 1 000 beam bunches, which is far from the 2×10^{12} expected bunch collisions of COMET Phase-I.

Efficient sampling methods

In order to produce detector-level data more efficiently, the outcomes of a full MC simulation can be re-used to generate events. Generally, any sampling mechanism which enables us to forego the proton collision will make the simulation more efficient at the cost of more uncertainty in the outcomes.

In the MC5 simulation described in Section 3.3, particles entering the detector region (either via the beam pipe or through the walls) are frozen and saved to disk to be processed later. These saved states can be reused by propagating them with multiple

distinct random seeds, thus affecting the outcome of every interaction and augmenting our sample.

Another way of generating events closer to our region of interest is to histogram the kinematics of particles at a given boundary and then sample events from the estimated distribution. In ICEDUST, the oaRooTracker format used to store particle states (see Section 3.1.3) can be aggregated into histograms, which are then used as input by the simulation.

To optimise the efficiency of the simulation even further, one could sample hits directly inside the detector system. In this chapter, we consider a fast generator of hit data to supplement Monte Carlo simulation. This generator should replicate hit patterns produced by simulated particles in the detector, without relying on a full particle-by-particle tracking approach.

The approach shown here is based on current state-of-the-art machine learning methods which were designed to be applied mainly in other domains, such as image and audio generation. Our hit generator is built using the principles underlying these methods as well as the technologies that were developed to implement them. In this chapter, we describe from “first principles” how the generative model was designed so that similar applications may build upon our work in the future.

4.1 Generative Adversarial Networks

Generative Adversarial Networks (GANs) are a class of generative models that can learn the underlying distribution of a dataset in order to synthesise new samples [41]. Training a GAN requires two neural networks to compete in a zero-sum game where one network (the *generator*) generates samples and the other (the *discriminator*) tries to discriminate between real and generated data. Figure 4.1 schematically illustrates a generic GAN training procedure.

As the conventional example, let us consider the case where both the discriminator and the generator are implemented as multi-layer perceptrons, denoted as D and G respectively. The discriminator D is designed as a binary classifier whose goal is to tell real samples (from the *training* dataset) apart from fake (generated) ones. Given a sample x , it outputs a score between 0 and 1, i.e. $D(x) \in [0, 1]$. Similar to a classification

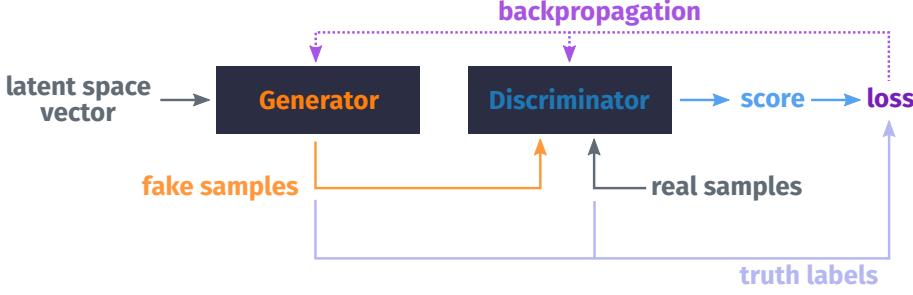


Figure 4.1: Schematic diagram of the data flow in the GAN training procedure. Real samples from the training dataset and fake samples from the generator are iteratively evaluated by the discriminator to compute a score. The loss for both networks is estimated by comparing the score to the sample's associated truth label (real or fake). By *backpropagation*, the loss is then numerically differentiated with respect to the network weights. The weights are then adjusted in the direction of decreasing loss.

task, we define the loss function of D as the cross entropy between its prediction and the true label (conventionally, 0 for a fake sample and 1 for a real sample):

$$\mathcal{L}_D = -\mathbb{E}_{x \sim p}[\log D(x)] - \mathbb{E}_{\tilde{x} \sim g}[\log(1 - D(\tilde{x}))], \quad (4.1)$$

where \mathbb{E} denotes the expected value or mean, p is the distribution of real samples and g is the distribution of samples generated by G . Minimising this function with respect to D implies that the discriminator will tend to assign a high score to samples drawn from the training dataset and a low score to generated samples.

The generator produces fake samples by mapping vectors from a latent space into data space. Latent-space vectors z are sampled according to a prior $p_z(z)$, typically a multivariate normal distribution for simplicity and speed. The objective of G is to generate samples $\tilde{x} = G(z \sim p_z)$ such that $D(\tilde{x}) \rightarrow 1$. In other words, the generator aims to maximise the second term in Equation 4.1, and its loss function is

$$\begin{aligned} \mathcal{L}_G &= \mathbb{E}_{\tilde{x} \sim g}[\log(1 - D(\tilde{x}))] \\ &= \mathbb{E}_{z \sim p_z}[\log(1 - D(G(z)))] \end{aligned} \quad (4.2)$$

Combining these minimisation tasks, we obtain the mathematical formulation of the adversarial training objective:

$$\min_G \max_D \mathbb{E}_{x \sim p}[\log D(x)] + \mathbb{E}_{z \sim p_z}[\log(1 - D(G(z)))] \quad (4.3)$$

Although it is not a formal requirement, using neural networks for D and G allows the above minimax solution to be approximated via backpropagation and stochastic gradient descent. At every iteration, we evaluate the gradient of each loss with respect to the internal weights of its respective network. Every weight is then adjusted toward the direction of steepest decrease in the loss.

4.1.1 Wasserstein GAN

The original formulation of GANs is notoriously difficult to train due to either non-convergence, instability or mode collapse¹. Training a GAN model is highly sensitive to the choice of hyperparameters: learning rate, optimisation algorithm, network architecture.

The Wasserstein GAN (WGAN) formulation is an attempt to address the stability issues of the original GAN [42]. The authors argue that solving Equation 4.3, which implicitly minimises the Jensen-Shannon divergence between p and g , leads to vanishing gradients when the discriminator is too powerful, and thus to unstable training. Instead, they propose that the Wasserstein-1 distance² between p and g should be minimised, because of the superior continuity and differentiability properties of that metric. They go on to show that if the discriminator is 1-Lipschitz continuous³, then the adversarial training problem can be formulated as:

$$\min_G \max_D \quad \mathbb{E}_{x \sim p} [D(x)] - \mathbb{E}_{z \sim p_z} [D(G(z))]. \quad (4.4)$$

In practice, aside from a change to the loss functions, this method requires that the discriminator be replaced by a “critic”, so-called because its output is not bounded to $[0, 1]$ and can be better interpreted as a score. In order for D to satisfy the Lipschitz continuity constraint, the authors propose that the magnitude of its weights should be restricted to a small range, e.g. $[-0.01, 0.01]$.

In addition to demonstrating the superior training stability of this WGAN method,

¹Mode collapse is the situation where G maps every point in the latent space onto the same output, leading to low diversity in the generated samples.

²Also called earth mover’s distance, Wasserstein distance is a distance function between probability distributions. If each distribution is a pile of soil, it corresponds to the minimum cost of turning one pile into the other, where cost is defined as mass times distance. Wasserstein-1 denotes the first order Wasserstein distance.

³A function is 1-Lipschitz continuous if its gradient has a norm less or equal to 1 everywhere.

it was observed empirically that this formulation does not lead to mode collapse and improves the robustness of the GAN with respect to changes in the network architectures.

Gradient penalty

The same year, another method to enforce the Lipschitz constraint on D was discussed by Gulrajani et al. [43], which outlines the shortcomings of weight clipping and instead proposes that the discriminator's gradient should be constrained. Since this is difficult to achieve in practice, the authors suggest a soft constraint on the gradient norm of D using an explicit term in its loss function:

$$\mathcal{L}_D = -\mathbb{E}_{x \sim p} [D(x)] + \mathbb{E}_{z \sim p_z} [D(G(z))] + \underbrace{\lambda_{\text{GP}} \mathbb{E}_{\hat{x} \sim p_{\hat{x}}} \left[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2 \right]}_{\text{Gradient penalty}}, \quad (4.5)$$

where the third term is the gradient penalty (GP) added to the WGAN loss, and λ_{GP} is the gradient penalty constant, a new hyperparameter. In this term, samples \hat{x} are drawn from $p_{\hat{x}}$ by sampling uniformly along straight lines between pairs of points from p and g .

The generator loss is, again, defined so as to increase the likelihood that a fake sample will fool the discriminator:

$$\mathcal{L}_G = -\mathbb{E}_{z \sim p_z} [D(G(z))], \quad (4.6)$$

which is the second term of Equation 4.5 with a minus sign, similarly to the relationship between Equations 4.3 and 4.2. Note that more generally, the generator loss should be defined as $\mathcal{L}_G \equiv -\mathcal{L}_D$ to achieve adversarial training. However, in the case of neural networks trained by gradient descent, only the derivative of \mathcal{L}_G with respect to G 's weights is relevant. Since the other terms in \mathcal{L}_D do not depend on G 's weights, we can discard them in \mathcal{L}_G to achieve the same result.

In our experiments, we found that the above WGAN-GP formulation provides the most stable training procedure for a variety of network architectures. In addition, it has the significant advantage of making the critic loss more interpretable and overfitting noticeable: as the authors demonstrate, the critic loss tends to converge to a maximum

value over training iterations, and in the case of overfit the critic losses evaluated on training and test samples diverge during training.

4.2 GANs in High Energy Physics

In HEP, GANs (both in their original and subsequent formulations) have been proposed in a variety of experiments to supplement traditional Monte Carlo simulation. Their usage typically falls in one of two categories: event generation, and generation of hit data at the detector level.

4.2.1 Event generation

Event-generating GANs typically focus on synthesising kinematic properties of outgoing particles in specific processes. In high-energy and hadronic collisions especially, a GAN generator has the potential to greatly reduce the computational cost of generating events via traditional MC simulations only.

In collider experiments, GANs have been used to simulate final-state kinematics for Z- and top-producing events [44, 45]. In anticipation for the High-Luminosity LHC, a GAN model was also trained to generate dimuon final states from Z decay, toward the production of large analysis-specific datasets [46]. In the context of the SHiP experiment, a GAN model is used to sample the position and momentum of muons produced by the collision of a 400 GeV/c proton beam with a fixed target, thereby reducing the need to simulate every hadronic interaction [47].

4.2.2 Detector data generation

Generation of detector data means building a model of the possible particle signatures inside a specific detector system. The hit patterns depend on the type of detector, detector geometry, particle types and energies, as well as the timescales considered.

In experiments where hadronic jets are commonly observed in the detector system, such as ATLAS and CMS, deep convolutional GANs were proposed as an alternative to MC simulations to generate *jet images*, a 2D representation of the energy deposition patterns from jets [48].

Subsequently, the CALOGAN model [49] used parts of the jet-image GAN as building blocks in order to generate realistic 3D electromagnetic showers. As a noteworthy addition, the generator was also conditioned to produce appropriate showers given the energy of the inbound particle. This implies that conservation of total energy could be implicitly learned by the model along with the distribution of shower patterns, leading to more physically consistent samples.

In addition to the original GAN concept, the WGAN-GP formulation is also used in the HEP domain to generate simulated detector data, for instance to produce cosmic ray-induced showers in a water-Cherenkov detector array such as the Pierre Auger Cosmic Ray Observatory [50], or showers caused by an electron beam hitting the CMS High Granularity Calorimeter [51].

In the case of COMET, both event generation and detector hit generation could be applied, as the main computational bottleneck occurs in simulating the proton beam collision. An event-generating GAN could for instance be trained to sample backward-going particles from the collisions. However, because of the large separation between the proton target and the detector, generating hits directly is a more efficient approach. The rest of this chapter focuses on hit generation inside the CDC of COMET Phase-I.

4.3 The CDC hit generator

Our approach draws inspiration from recent GAN implementations in other high energy physics experiments and in other areas, such as image and text synthesis. We design a GAN model whose purpose is to efficiently and realistically generate hits in the CDC, with the aim of augmenting MC datasets for COMET Phase-I.

In order to build a generative model to mimic the patterns of hits in the detector, it is necessary to consider these hits as physical energy deposits created by simulating energy loss processes of particles inside the detector gas volume. The true process of simulating hits thus obeys strict rules which stem from the physical models at play. In contrast, our GAN model is not required to be aware of the physics, which implies it can be much faster at producing hits, but it can also lead to inaccuracies in the resulting generated data which must be evaluated.

This section describes how the simulated hits are selected and arranged, and how

	Energy deposit	Time	Dist. of closest approach	Wire index
Unit	MeV	ns	mm	—
Label	edep	t	doca	wire
Data type	float	float	float	integer

Table 4.1: Physical features which make up each hit in the training dataset.

the networks are consequently built in order to learn a representation of the data. Multiple aspects of the problem must be considered carefully in order for the GAN to produce a faithful model, such as how to select, lay out and pre-process the hit data.

4.3.1 Hit data structure

The GEANT4 simulation tracks particles one by one. Secondaries produced during the tracking of a particle are only propagated once the parent has ended its journey. This implies that simulated hits occur in a specific order (which is not strictly the time order), where hits produced by the same particle are laid out contiguously. During training, data is presented to the GAN in this particular order such that it learns to model entire tracks rather than isolated hits.

Each hit is composed of four variables, each of which is necessary to later simulate the ionisation in the gas and the response of the readout electronics. These variables, or *features*, are: the amount of energy deposited, the absolute time since the start of the event, the distance of closest approach (DCA) from the track to the wire, and the position of the hit inside the CDC. This position is represented as the unique index of the cell, from 0 to 4985, where the energy loss occurred. These features are summarised in Table 4.1.

The CDC gas simulation dictates that each particle should produce one hit in every cell it traverses. The GAN ought to be able to learn how hits are related within a track, thus it should be able to process more than one hit at a time. The hit data is arranged into sequences, where series of hits produced by multiple particles are present. Figure 4.2 illustrates the arrangement of hit data as it is generated by the generator and fed to the discriminator.

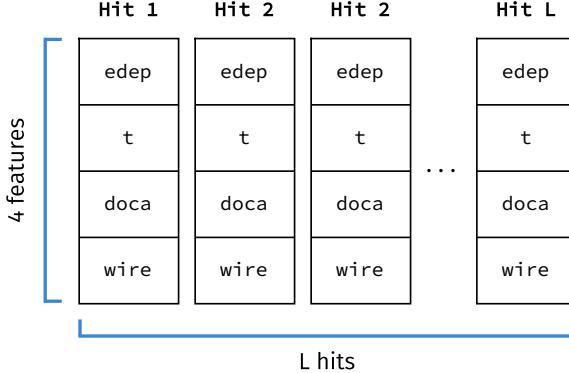


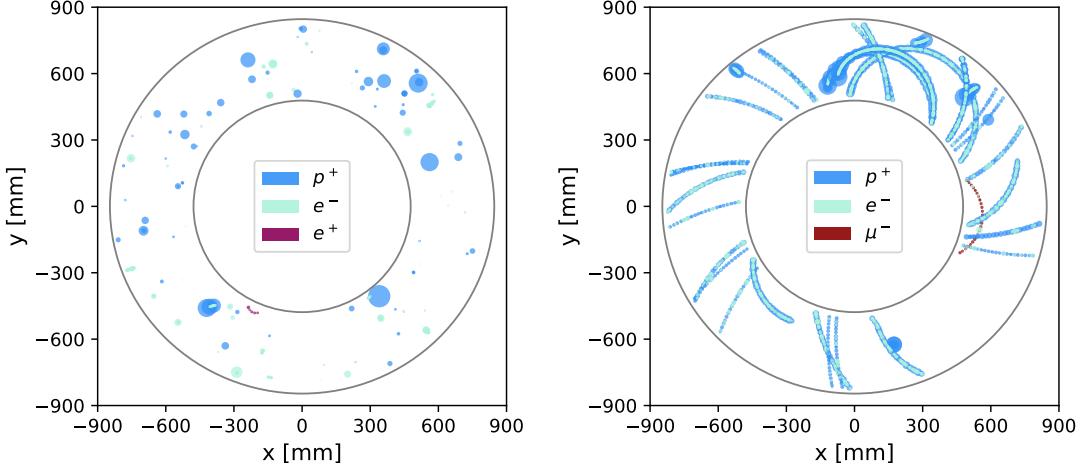
Figure 4.2: Structure of one training sample. Hits are arranged into fixed-length sequences in the GEANT4 order, i.e. whereby hits from the same track are contiguous.

4.3.2 Event selection

Not all MC events are fed to the GAN for training. In particular, potential background sources, i.e. particles with a momentum around $105 \text{ MeV}/c$, produce characteristic series of hits in the CDC which are likely to be reconstructed by a track fitting algorithm. Such a particle's trajectory is firmly dictated by the Lorentz force, which is a constraint difficult to impose on a GAN model. In addition, a GAN might generate too large or too small a proportion of reconstructible events, leading to more uncertainty in the background rates estimated from samples containing synthetic hits. Unlike MC data, fake tracks can yield no information concerning their origin, as the GAN-generated hits come completely unlabelled.

These reasons motivate a splitting of the dataset into *reconstructible* events, which only the MC simulation may simulate, and *noise-like* events, where no reconstructible track occurs, and the data is deemed “uninteresting” enough that it can be augmented by the GAN without biasing the physical results.

In each MC event (corresponding to the outcomes of one proton-on-target collision), we determine if at least one particle with $p > 50 \text{ MeV}/c$ has entered the CDC and produced at least 4 hits. In this case, all hits from the event are given the “reconstructible” label and will not be used to train the GAN. If no reconstructible track has occurred during the event, all hits are considered noise-like and are added to the training dataset. Figure 4.3 compares noise-like and reconstructible hits and outlines the difference in structure between the two categories.



(a) Noise-like hits from 100 MC events. (b) Reconstructible hits from 10 MC events.

Figure 4.3: Comparison of hit patterns in the CDC depending on whether or not a reconstructible track occurred. The area of each hit is proportional to the amount of energy deposited, and the colour denotes particle type.

The MC5 dataset contains 10^9 events of which 51 137 have hits in the CDC. Of those, 2 036 events have at least one reconstructible track, and 49 101 events do not. Inside this latter category, 634 266 hits are present, and these make up the GAN training dataset.

4.3.3 Pre-processing

The three continuous hit features (energy deposit, time and DCA) are physical quantities with different units and scales. Their distributions are shown in Figure 4.4. In order to maximise gradient flow in the discriminator, it is common practice for all features to be re-scaled into a fixed range, e.g. [0, 1]. However, in the case of highly uneven and/or discontinuous distributions, the generator can struggle to model the data which results in a distribution of generated data which does not match the distribution of training data.

To alleviate this issue and define a common scale for all continuous features, each feature undergoes a non-linear *quantile transformation*, which corresponds to a mapping from each value x_i to $y_i = \Phi^{-1}(F(x_i))$, where F is the empirical cumulative distribution function (CDF) of the feature, and Φ is the Gaussian CDF. This transformation effectively yields normally distributed features, which the generator can more easily reproduce. Its task is not however made trivial, as the correlations between features remain, which it must learn to model.

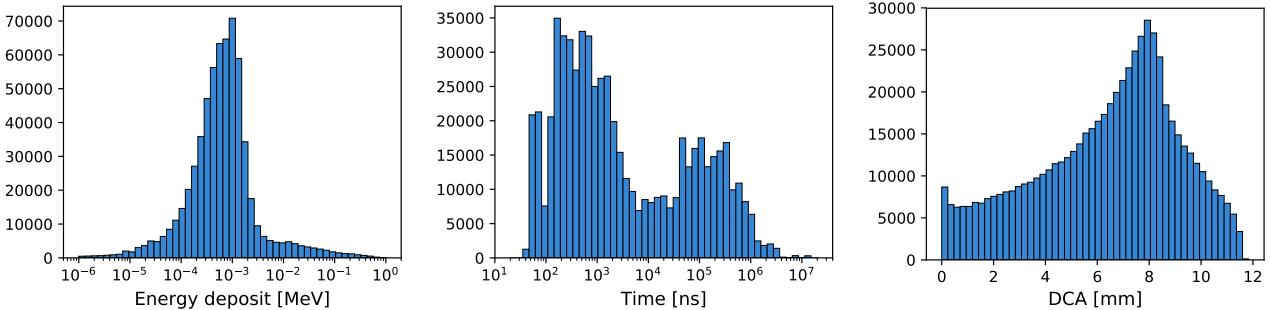


Figure 4.4: Distributions of the three continuous features used to train the GAN.

In addition to the quantile transformation, the values of each feature are re-scaled independently to lie within the $[-1, 1]$ range. To prevent the generator from producing values out of this range, which would give the discriminator a clear indication of a sample being fake, the hyperbolic tangent function is used as the generator’s final layer.

4.3.4 Hit sequence sampling and data augmentation

At each step of the training, the discriminator receives two mini-batches of hit sequences: one from the training data and one synthesised by the generator. The discriminator’s output is then used to compute the loss for G and D , according to Equations 4.5 and 4.6.

To allow parallel processing of the hit sequences by the networks, a fixed sequence length L is defined prior to training. Hence hits can be packed into dense tensors upon which vectorised operations can be applied. Typically, a sequence will contain concatenated sets of hits from multiple tracks and from multiple events.

Hits belonging to different events are completely unrelated, as they are not the product of the same POT collision. This provides the freedom to swap events and concatenate their hits in any order. The order of hits within the same event should however be preserved. Combining events in many ways enhances the diversity and size of the training dataset, which alleviates the possibility of overfit or mode collapse. Concatenating the 634 266 hits linearly into sequences of length $L = 512$ would yield only 1239 samples. By rearranging the events, we produce 500 000 unique samples which the GAN learns from. Although it is not the maximum number of event recombinations, this number of samples was found to be large enough to prevent overfit of our GAN model.

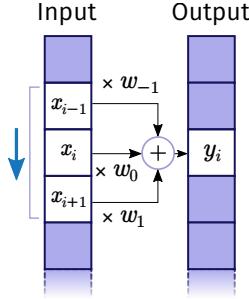


Figure 4.5: Temporal (1D) convolution operation with kernel size 3. For clarity, both the input and output sequences have one channel, i.e. $C_{\text{in}} = C_{\text{out}} = 1$. The blue arrow shows the direction along which the kernel slides across the input sequence.

4.3.5 Network components

Temporal convolutions

In order to process sequences of hits efficiently, both networks use a series of convolutional layers which (unintuitively) apply cross-correlation operations between the input x and the learned kernel w :

$$y_i = (w * x)_i = \sum_k w_k \cdot x_{i+k}, \quad (4.7)$$

where i is the position along the output sequence y , and k runs over the kernel weights. This operation is illustrated in Figure 4.5.

Unlike in the common case of raster images, the input to the discriminator is not a two-dimensional array of pixels, but a one-dimensional sequence of hits. Hence, the kernels are also one dimensional. The “channels” of the input sequence correspond to features of the hits. Within a convolutional layer, Equation 4.7 is applied as many times as there are channels in the input sequence, and the result is the sum over channels. The operation can also be applied in parallel with many different kernels, in which case each kernel will produce one channel in the output sequence.

The fact that the kernel acts on adjacent elements implies that the network effectively processes multiple consecutive hits at a time. This allows the generator to synthesise coherent sequences and the discriminator to perceive whether a generated sequence has the same overarching structure as real sequences.

Optionally, convolution kernels can be applied over the sequence with a stride,

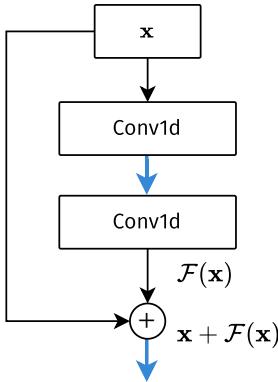


Figure 4.6: Residual blocks used in both networks to facilitate gradient flow. Blue arrows indicate an activation function.

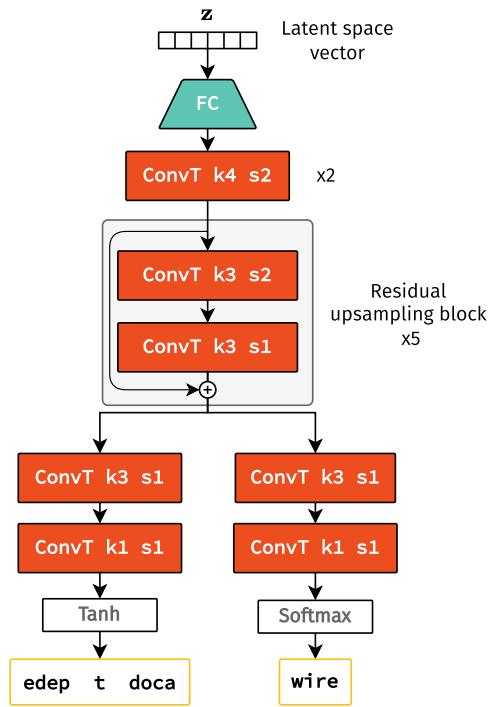
meaning that the kernel moves down by more than one element between two cross-correlation operations. This has the effect of producing an output sequence that is shorter than the input sequence. This is often used to allow the network to see the data over increasingly large scales and to keep the total number of operations reasonable in layers where the input has many channels.

Residual connections

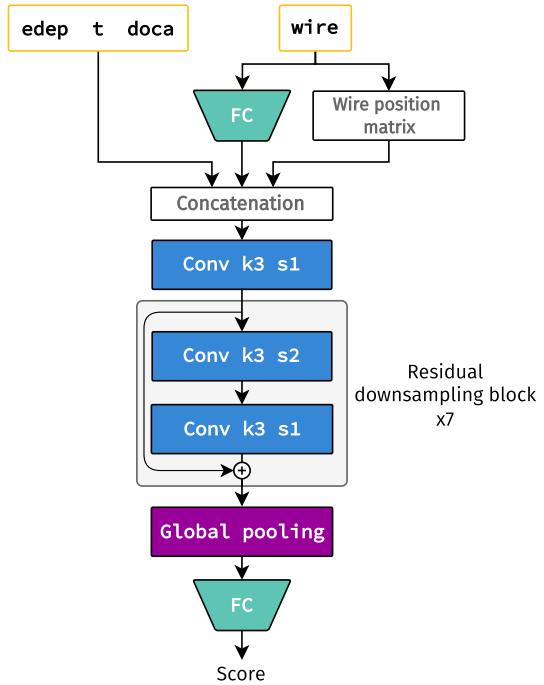
In order to facilitate gradient flow from the discriminator to the generator and thus speed up training, convolutional layers can be replaced by residual blocks [52] in both networks. A residual block combines two convolutional layers with a residual connection from the input to the output, allowing the gradients to propagate via two paths, as shown in Figure 4.6. The output y of the block can be written as $y = \mathcal{F}(x) + x$, where \mathcal{F} represents the operation of the two convolutional layers and x is the input tensor. When $\mathcal{F}(x)$ and x have different shapes, a dimension-matching linear transformation can be applied to x . In this case, the result is $y = \mathcal{F}(x) + Wx$, where the weights of the W tensor are adjusted during training.

4.3.6 Network architectures

Having introduced the building blocks, we can now assemble them into our discriminator and generator. The two networks, shown schematically in Figure 4.7, are composed of stacked residual blocks, which progressively increase the size of the sequence in the generator, or decrease it in the discriminator.



(a) Generator.



(b) Discriminator.

Figure 4.7: Network architectures. The layer types are: FC for fully-connected, Conv for a standard convolution and ConvT for a transposed (fractionally-strided) convolution. In convolutional layers, kernel size and stride are denoted by k and s , respectively.

Data flow

In the generator, the input vector, sampled from latent space, is first passed through a fully-connected layer, after which it is reshaped to have a length of 4. The proto-sequence is then up-sampled by fractionally-strided convolutional layers, adding finer and finer details to the sample. As the length increases, the number of channels is decreased to limit the total number of operations and the network’s complexity. Once the sample has reached a length of L , the information is projected to data space: to obtain the continuous features, a convolutional layer with three output channels is applied and its output passed through a tanh function, yielding values in the range $[-1, 1]$; while for the discrete wire index, we use a layer with 4986 output channels⁴ followed by a softmax⁵ function to determine the network’s preferred choice of wire for each hit.

The discriminator extracts information from the input hits through its convolutional layers, progressively down-sampling the sequence while increasing the number of channels. After 7 residual blocks, a global average pooling is performed, i.e. for each channel, the result is the average value over the sequence. The resulting tensor is passed through a final fully-connected layer, combining the features into a single scalar which is the score for that sample.

Gradient flow for discrete features

We would like the discriminator to have some knowledge of where, in the detector, hits are located, so that it can use this knowledge as a criterion in its judgement. A straightforward way to give this information to the discriminator would be to use the wire index to query a lookup table of wire positions. However, indexing an array is not a differentiable operation, hence it does not allow gradients to flow from the output score to the generator weights, and this prevents the generator from learning to accurately position hits.

To allow gradient flow, we can represent the wire index as a *one-hot encoded* vector, whose size is the total number of wires, and where only the entry corresponding to the index is 1, and the others are 0. This vector representation can be used in differentiable

⁴As many as there are wires.

⁵Defined as $\text{Softmax}(x)_i = \frac{\exp(-x_i)}{\sum_j \exp(-x_j)}$.

Optimiser	Adam	Sequence length	512
Learning rate	10^{-4}	Loss	WGAN-GP
(β_1, β_2)	(0.9, 0.999)	λ_{GP}	10

Table 4.2: Training hyperparameters.

operations to query a particular entry in a table: if we represent the table of wire positions as a matrix W with dimensions 2×4986 , and the one-hot encoded wire vector as a 4986-dimensional vector v , the product $W \cdot v$ is a differentiable operation and its output is the wire’s position since the one-hot vector picks out the correct entry from the position matrix.

In the case of generated hits, the wire vector is not exactly one-hot: instead, the output of the softmax operation is a vector of real-valued probabilities for each wire. The multiplication of the above wire position matrix with such a vector can be interpreted as a weighted sum of all wire positions with the weight corresponding to the softmax probability. Hence, the result is not an exact wire location but an approximate region of the detector in which the generator envisions the hit to be.

As discussed in the original WGAN-GP article [43], one might suspect that the discriminator can learn to reject the output of the generator because it does not look like a one-hot vector. However, the Wasserstein distance between the real and generated data remains well-defined and continuous even in this particular case of a discrete variable, hence using the WGAN loss does allow the networks to learn.

4.3.7 Training

The networks are trained via gradient descent of the WGAN-GP loss, using the Adam algorithm [53]. The PyTorch framework with which the GAN is implemented allows all tensor operations to be performed on a Graphics Processing Unit (GPU), which speeds up training iterations roughly tenfold. Hyperparameters, such as learning rate and λ_{GP} , were initially set to a default value from common practice guidelines and then adjusted if needed to ensure stable loss curves. Table 4.2 summarises the values of various hyperparameters, and Figure 4.8 shows the loss curves for G and D over 200 training epochs.

The loss functions for D and G evaluated at each training step are recorded to show

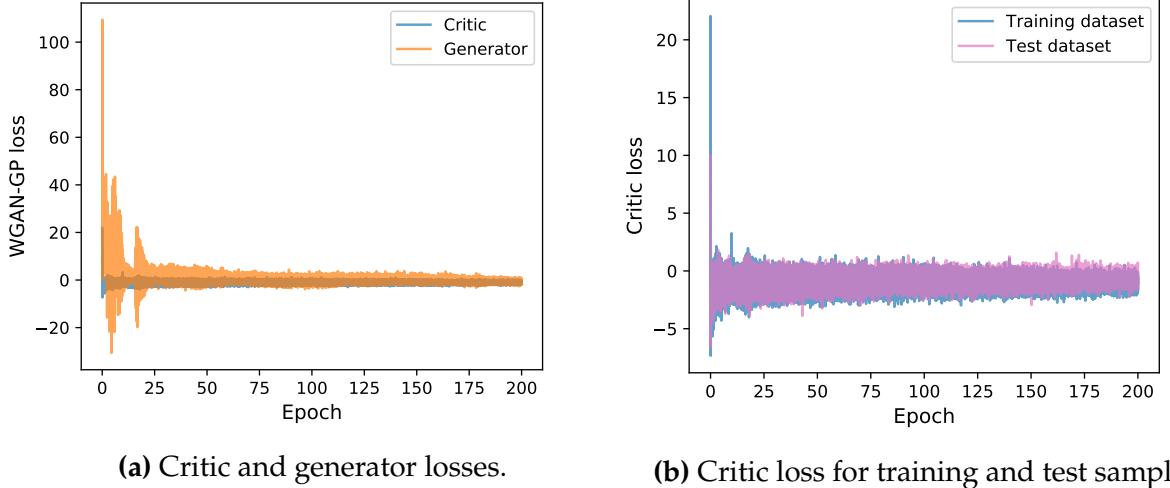


Figure 4.8: Loss curves as a function of training epochs.

the progression and ensure that the networks are learning. In addition, every 100 mini-batch iterations, the critic loss is evaluated on test samples which are absent from the training dataset. This allows us to determine whether D is overfitting to the training data, in which case the loss calculated using training and test samples (Figure 4.8b) would diverge [43].

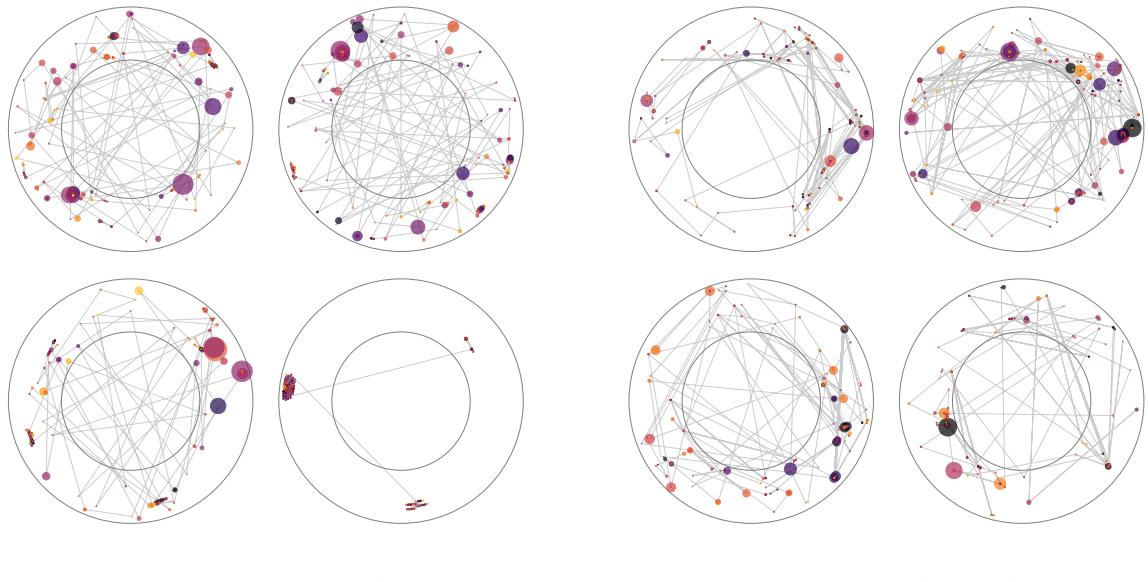
4.3.8 Evaluation

Comparing the raw rate of hit production in the CDC, the GAN, once trained, generates hits six orders of magnitude faster than the downstream simulation used to produce MC5. Using MC5 as reference, the GAN model can produce as many noise-like hits as there are in the downstream dataset in six seconds on a single GPU machine, versus a day of running on 2000 cores for the Monte Carlo simulation. The increase in speed is indeed the main reason for using a generative model, however it is also crucial to determine how faithful the generated samples are to the original dataset.

Evaluation of GAN-generated samples is notoriously difficult in all domains because of the lack of a well-defined, unique metric for quality⁶. The score returned by the discriminator network is dependent upon its exact architecture, the dataset and all hyperparameters, hence it cannot be used as a reliable metric.

Unlike natural images which can be examined visually to determine whether they depict something real, sequences of hits in a detector are difficult to evaluate percep-

⁶Unlike in supervised learning, where an obvious quality metric is usually the loss function itself.



(a) Training samples.

(b) Generated samples.

Figure 4.9: Comparison of individual samples, comprised of 512 hits. Hits are represented by a circle whose area is proportional to the energy deposit and whose colour shows distance of closest approach. Lines connect consecutive hits in the sequence.

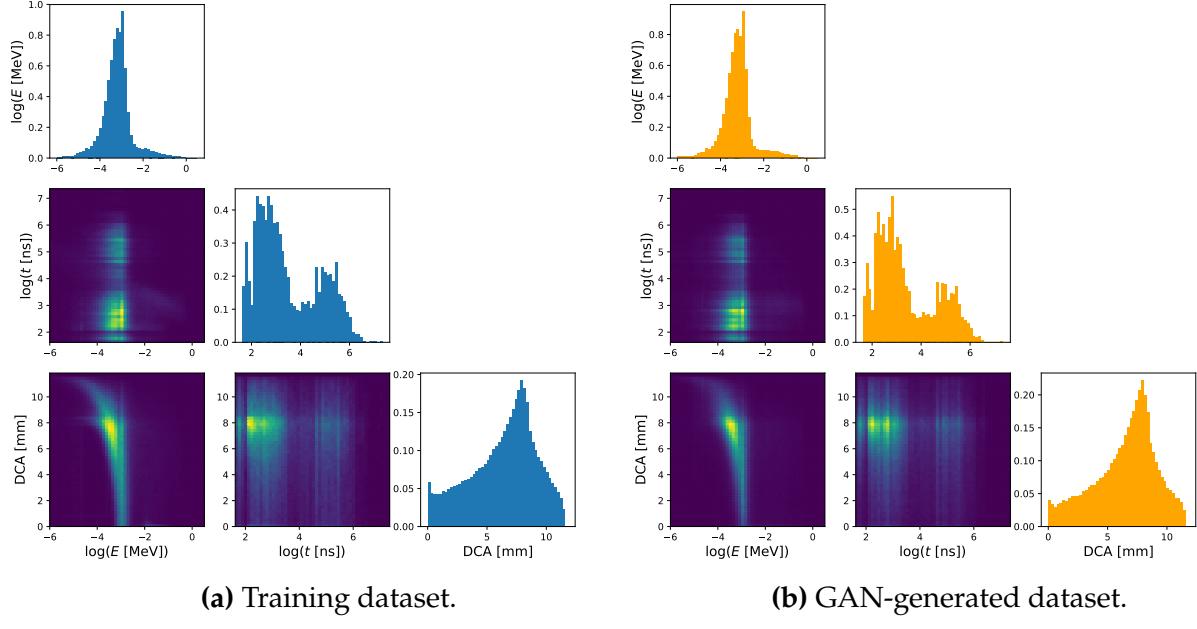
tually. Figure 4.9 shows a comparison between uncurated individual samples, four from the training dataset and four generated by G . In this type of evaluation, we try to emphasise any differences in the features and sequence structure through colours and visual hints such as lines connecting consecutive hits.

On a larger scale, we also compare the distributions of hit features across many samples in order to verify that the GAN can model the dataset properly. Drawing two-dimensional histograms of each pair of features, as in Figure 4.10, also reveals whether feature correlations are faithfully modelled by G .

In order to summarise the similarity between the distribution of real and generated samples into one quantity, a commonly used metric is the Kullback-Leibler (KL) divergence, which can be interpreted as the amount of information lost when approximating the real dataset with a generated one [54]. KL divergence is defined on a probability space \mathcal{X} as

$$D_{\text{KL}}(P \parallel Q) = \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)}, \quad (4.8)$$

where in our case P denotes the distribution of real hits and Q that of generated hits. To compute D_{KL} , the three continuous features of every hit are binned into a three-dimensional histogram to approximate the probability density functions P and Q and



(a) Training dataset.

(b) GAN-generated dataset.

Figure 4.10: Comparison of feature distributions and correlations when generating a dataset of the same size as the training dataset.

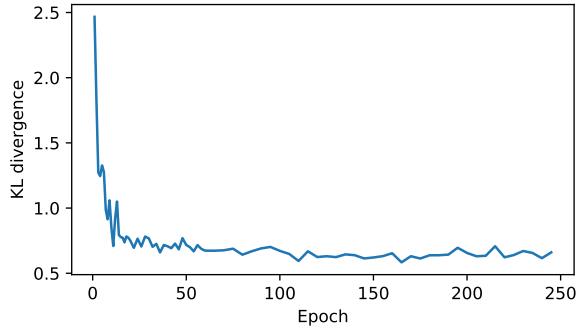
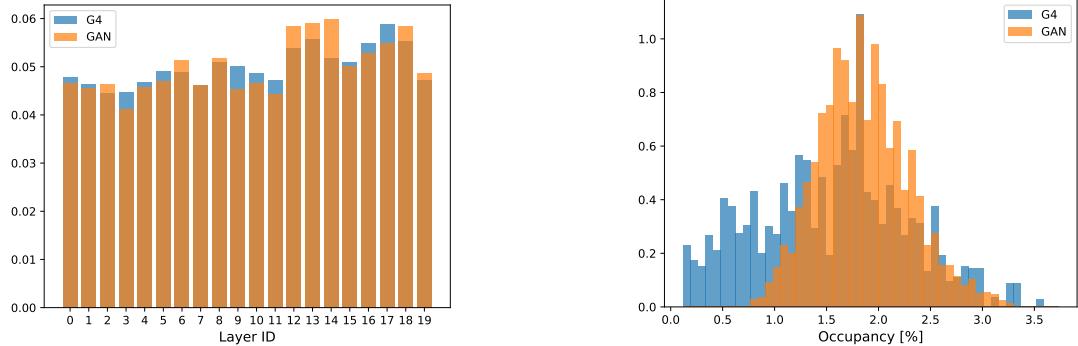


Figure 4.11: Kullback-Leibler divergence between the distributions of real and generated hits as a function of training epochs.

the sum in Equation 4.8 is a sum over all (non-zero) bins. The edep and t features are log-transformed such that the resulting histogram is smoother. Figure 4.11 shows the Kullback-Leibler divergence between the continuous distributions of real and generated hits over training iterations. For this specific model, the KL divergence converges after around 100 epochs.

In addition to comparing the features that are given as input to the GAN for training, we also examine quantities which are only present implicitly in the training samples. Figure 4.12 shows a comparison of the relative amounts of hits in each CDC layer and of the “occupancy” per sample, i.e. the fraction of wires on which a hit occurs. This allows us to determine whether or not the model sees beyond the features provided



(a) Layer index, which is related to wire index and radial position.

(b) Occupancy, i.e. the number of activated wires in a given sequence.

Figure 4.12: Comparison of quantities not explicitly learned by the GAN model.

explicitly for training.

4.4 Quality metrics

The Kullback-Leibler divergence provides a metric to measure how close a distribution of generated samples is to that of training samples. However, it does not provide an indication of how closely individual generated samples resemble real samples.

For that, one method which is commonly used in the field of GANs is an external classifier that can extract features from the generated and real samples. These extracted features can then be used to understand the differences between the two and why they arise.

The task of investigating the quality of generated hit samples from the CDC GAN was tackled by a pair of Master's students during a six-month project under my supervision [55, 56]. The following methods and results are the product of their work and were presented at the 21st International Workshop on Advanced Computing and Analysis Techniques [57].

4.4.1 Inception score

In image generation tasks, the *Inception score* was proposed as a quality metric to compare images generated by different GAN models [58]. An Inception model [59], trained independently on the ImageNet dataset to classify images, is applied to GAN-

generated images to obtain label probabilities for each sample. The metric is then formulated in terms of the label distribution to favour generated images which depict a specific object rather than some average of classes, while disfavouring a generator whose images all resemble the same class.

4.4.2 Fréchet Inception Distance

The Fréchet Inception Distance (FID) is an evolution of the Inception score which additionally uses statistics extracted from the training dataset to estimate a distance between real and generated samples [60]. Conventionally, a pre-trained Inception-v3 model [61] is applied to real and generated samples. Activations from an intermediate layer are extracted and their mean μ and covariance Σ are estimated, from which the Fréchet (or Wasserstein-2) distance is calculated as

$$d^2 = |\mu_G - \mu_R|^2 + \text{Tr}(\Sigma_G + \Sigma_R - 2(\Sigma_G \Sigma_R)^{1/2}) \quad (4.9)$$

where the G and R subscripts respectively denote generated and real statistics.

Applying these methods to domains other than image generation is not straightforward because the Inception network used to compute the metric is specific to raster images with three colour channels. This has prompted the development of domain-specific metrics such as Fréchet Audio Distance [62] and Fréchet Video Distance [63], as well as the Graph Fréchet Distance used to evaluate particle jets generated by a graph-based GAN for LHC experiments [64].

4.4.3 Evaluating the CDC GAN performance

Pre-trained Inception-v3

As a baseline to evaluate the CDC GAN, the Inception-v3 model trained on the ImageNet dataset was used to compute FID between generated and real samples. The network expects images as input, i.e. tensors with two spatial dimensions and three channels, which means our hit data must be arranged to fit this shape.

GAN-generated hits have three continuous features and a position. The three features can simply be passed as analogues to the three colour channels, but the hit

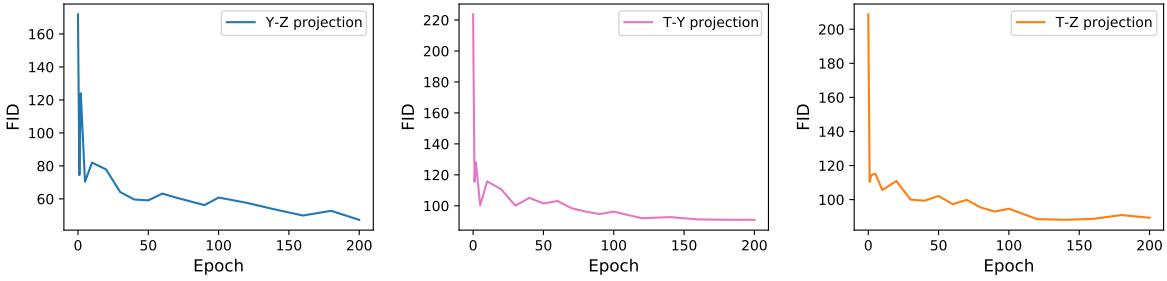


Figure 4.13: Fréchet Inception Distance computed by the pre-trained Inception-v3 network over training epochs of the CDC GAN. Since Inception requires a 2D image as input and our hit sequences have a third temporal dimension, the hit data is projected along its axes and a separate FID is calculated for each projection.

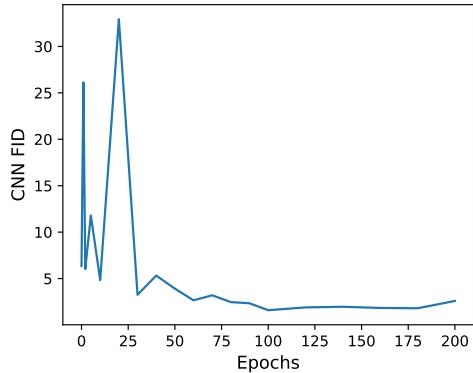
positions must be discretised in order to get an image-like object. In addition, the ordering of hits bears importance, so it ought to be considered as a third, temporal dimension. However, Inception cannot process 3D objects hence a solution is to project out one of the dimensions and compute three orthogonal scores, one for each projection. Figure 4.13 shows FID computed for each of the three projections as a function of training iterations.

In order to process whole hit sequences rather than projections, two alternative methods were considered: training a custom network which can take three-dimensional tensors as input, and adding layers to the Inception network so that it can more naturally process our hit data.

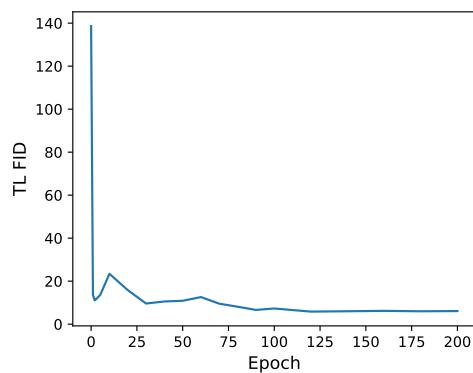
3D convolutional neural network

A custom neural network is developed to process whole hit sequences and compute a domain-specific FID. The architecture is composed of four convolutional layers that extract features from hit sequences, i.e. three-dimensional tensors with three channels. The convolution kernels are thus themselves 3D, and strides help the network extract features at different scales, as for the GAN critic.

The convolutional neural network (CNN) is trained to discriminate Monte Carlo *noise-like* hit sequences from *reconstructible* ones (as defined in Section 4.3.2). We expect such a classifier to rely on the positioning and clustering of hit patterns to discriminate samples. The intermediate activations which will be used to compute the Fréchet distance ought to be a representation of that information and hence help to define a



(a) Domain-specific 3D CNN.



(b) Transfer Learning Inception-v3.

Figure 4.14: Fréchet Inception Distance computed over training epochs of the CDC GAN by two external neural networks. Unlike in Figure 4.13, hit sequences can be processed at once without projecting them along an axis. The 3D CNN is trained on Monte Carlo hit data, in the same way as the front and back layers added to Inception-v3 for transfer learning.

sensible distance metric.

Once trained, the CNN reaches a classification accuracy of 98.7 %, and it can be used to compute FID on GAN-generated samples. Figure 4.14a shows FID over epochs of the CDC GAN training.

Transfer learning Inception-v3

Transfer learning (TL) refers to the addition and subsequent training of layers to a fixed pre-trained network in order to apply it to a new domain. By adding layers to the front and back of Inception-v3, it can be adapted to the hit sequence classification task. The weights in the new layers are adjusted to discriminate *noise-like* from *reconstructible* hit sequences, which allows domain-specific features to be extracted. In addition, because the input layer is modified, this new network is made to provide a single FID score, rather than three, for comparing GAN-generated samples to real ones. Figure 4.14b shows the resulting FID score computed by the TL Inception-v3 over epochs of the CDC GAN training.

4.4.4 Interpretation of FID

The Fréchet Inception Distance is a single quantity whose value indicates how similar generated samples are to real samples. It summarises the statistics of features extracted

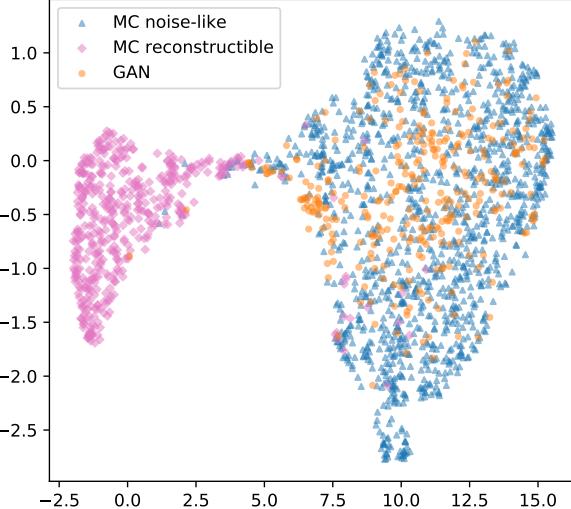


Figure 4.15: UMAP computed from feature vectors extracted by the TL Inception-v3 network, for reconstructible, noise-like and generated hit sequences. The number of neighbours in the UMAP transform was set to 50.

by an external neural network, but it does not directly show which features seem different.

We could also plot the feature vectors directly and determine where their distributions differ between real and fake samples, however the vectors have a high dimensionality and cannot be easily visualised. The solution that was adopted is to use the Uniform Manifold Approximation and Projection (UMAP) method to optimally project the feature vectors onto two dimensions [65]. Each sample can then be associated with a single point in a 2D space and thus one can more easily visualise the distribution of samples.

Figure 4.15 shows a UMAP plot of generated and real samples, including reconstructible hit sequences. The position of each point is a projection of the feature vector extracted from each sample by the transfer learning Inception-v3 network. Such a plot shows that reconstructible and noise-like sequences are adequately discriminated by the network, and that the generated and noise-like sequences overlap in large part. Hence, the GAN is clearly able to produce samples which resemble the training dataset. However, some portion of the space covered by noise-like MC samples is not occupied by any generated samples, which indicates that the GAN cannot produce these types of samples.

By tracing UMAP points back to the sample they were computed from, one can then

attempt to understand the type of sample which the GAN cannot faithfully produce. This yields insight into the GAN’s strategy and where exactly it fails to perform, which subsequently helps to adjust its design and architecture.

Chapter 5

Backward Monte Carlo Simulation of Atmospheric Muons

Cosmic ray-induced events represent a significant part of the expected backgrounds in the COMET experiment, as discussed in Section 2.7. Cosmic rays interact with Earth’s atmosphere to produce secondaries with a wide energy spectrum. Some fraction of these particles will be able to mimic the conversion signal by entering the COMET detector system. In this chapter, we discuss how backward Monte Carlo simulation helps to efficiently estimate the atmospheric background rate.

5.1 Backward Monte Carlo simulation

5.1.1 Principle

Estimating the rate at which atmospheric muons might produce signal-like events is computationally expensive through standard Monte Carlo simulation methods. This is due to the fact that the source of atmospheric muons, i.e. the atmosphere, has a much greater spatial extent than the active detector region. Hence, most cosmic events generated from an atmospheric source will miss the detector, thus they will not contribute in the estimation of the background rate, resulting in wasted computation time.

A backward (or adjoint) Monte Carlo simulation is one where the flow of particle transport is reversed, i.e. events are generated in the sensitive detector volume and

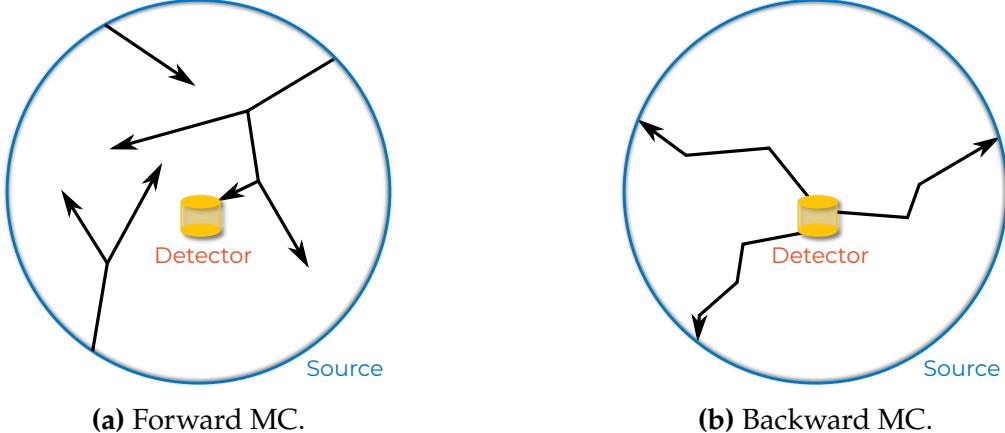


Figure 5.1: Comparison of the forward and backward (or adjoint) configurations of Monte Carlo simulation, in the case where a small detector volume is surrounded by a comparatively large particle source. Reversing the flow of particle transport means that sampled events are more likely to contribute toward the rate estimation.

propagated backward in time until they reach the source, as illustrated in Figure 5.1. This method was initially used in 1967 to estimate gamma-ray and neutron radiation doses in nuclear reactors [66, 67]. More recently, backward MC has also been integrated into Geant4 simulations for dosimetry in space, reversing the electromagnetic physics of electrons, protons and ions [68].

In the case of muons, a backward transport method was developed by Niess et al. [69] in the context of muon tomography. The software package handling the backward simulation, PUMAS [70], was later adapted to the COMET experiment in order to refine estimations of the background rate from cosmic rays. Appendix A describes how PUMAS was integrated into the ICEDUST code base to enable this and future studies of the atmospheric muon background in COMET.

5.1.2 Method

In a backward MC simulation, events are generated according to an arbitrary probability density function (PDF) around the detector volume. The overall sampling PDF is typically the composite of a position, direction and energy sampling distribution, i.e.

$$p(x, p, E) = p(x) \times p(p) \times p(E), \quad (5.1)$$

where p denotes an arbitrary PDF whose integral is 1, x denotes position, \mathbf{p} direction and E energy. If more than one particle type is considered, the discrete distribution of particles can also enter Equation 5.1. In our case, both atmospheric muons and anti-muons are relevant, so they are sampled identically and in equal proportions.

Each event i is propagated backward using adjoint transport and interaction kernels such that the particle undergoes reverse continuous energy loss and reverse discrete processes, while moving backward toward the source. At each step j of the propagation, a multiplicative weight w_j^i is computed given the path taken and the interactions along it. When the particle reaches the source, the directional flux at the source ϕ_s is used as a normalisation factor to yield a weighted event flux

$$\phi(\mathbf{x}'_i, \mathbf{p}'_i, E'_i) = \prod_j w_j^i \times \phi_s(\mathbf{x}'_i, \mathbf{p}'_i, E'_i), \quad (5.2)$$

where \mathbf{x}' , \mathbf{p}' and E' denote respectively the position, direction and energy of the particle upon reaching the source.

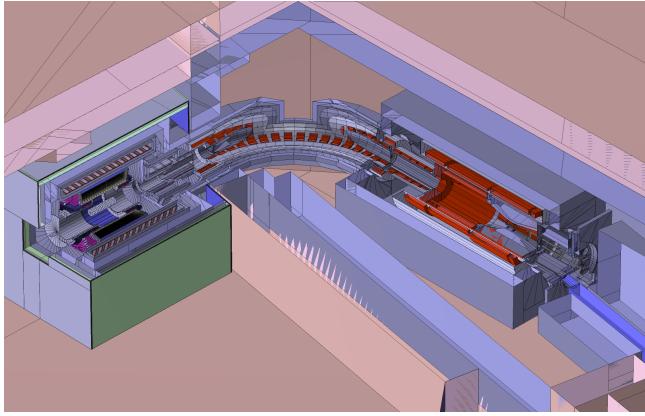
The weighted event flux ϕ is related to the event rate R by the inverse of the sampling PDF of Equation 5.1 used to generate the event initially:

$$R(\mathbf{x}_i, \mathbf{p}_i, E_i, \mathbf{x}'_i, \mathbf{p}'_i, E'_i) = \frac{\phi(\mathbf{x}'_i, \mathbf{p}'_i, E'_i)}{p(\mathbf{x}_i, \mathbf{p}_i, E_i)}. \quad (5.3)$$

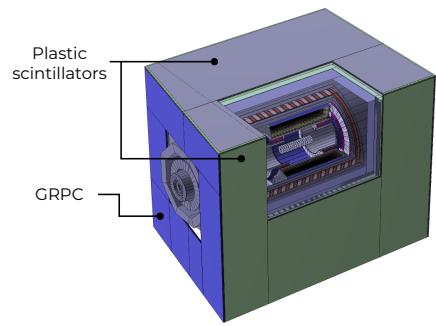
Since the rate depends on the position, direction and energy of the particle at the source, it is susceptible to variation across backward runs because of the stochastic aspect of Monte Carlo transport. Hence, backward transport is repeated N times for each event to find the average rate

$$\begin{aligned} \bar{R}(\mathbf{x}_i, \mathbf{p}_i, E_i) &= \sum_{k=1}^N R(\mathbf{x}_i, \mathbf{p}_i, E_i, \mathbf{x}'_k, \mathbf{p}'_k, E'_k) \\ &= \frac{1}{p(\mathbf{x}_i, \mathbf{p}_i, E_i)} \times \frac{1}{N} \sum_{k=1}^N \phi(\mathbf{x}'_k, \mathbf{p}'_k, E'_k), \end{aligned} \quad (5.4)$$

smearing out the dependence on \mathbf{x}' , \mathbf{p}' and E' . Given a large enough N , i.e. a significant enough sample of backward transports, the sum of weighted fluxes becomes equivalent to an integration over the possible transport paths and interactions. Results obtained



(a) Cutaway view of the simulation world.



(b) Detail of the CRV geometry.

Figure 5.2: COMET Phase-I simulation geometry used in the cosmic background study.

by following this procedure were demonstrated to be consistent with forward Monte Carlo [68]¹.

5.2 Application to COMET Phase-I

Backward MC simulation was used to estimate the cosmic background rate in COMET Phase-I. This section describes the simulation setup, validations of the method using experimental data, and the results of the study.

5.2.1 Geometry

The simulation geometry used in the simulation is shown in Figure 5.2. In the COMET Phase-I experiment, the cosmic ray veto (CRV) described in Section 2.5.3 will enclose the Cylindrical Detector in order to help identify events caused by atmospheric muons. The CRV uses a glass resistive plate chamber (GRPC) as the active detector on the upstream-facing side, and plastic scintillator counters on the other faces. Two holes are present in the CRV, one in the upstream face and one in the downstream face. These holes allow atmospheric muons to sneak into the CyDet unnoticed and potentially produce signal-like tracks, contributing to the background rate.

¹The article by Desorgher, Lei, and Santin [68] also provides a more thorough discussion of adjoint cross-sections, weights, weight corrections and source normalisation.

5.2.2 Event sampling

CRV envelope sample

As discussed in Section 5.1.2, events must be generated around our volume of interest, the Cylindrical Detector. We define a sampling envelope around the CRV, where muons (and anti-muons) are generated and backward-propagated. Events are sampled uniformly over this envelope, i.e. the position PDF $p(x)$ is independent of x :

$$p(x) = \frac{1}{N_x}, \quad (5.5)$$

where N_x is the normalisation term satisfying $\int p(x) dx = 1$.

The direction is sampled according to Lambert's cosine law, which favours events with a large vertical momentum component:

$$p(p) = \frac{1}{N_p} (-\hat{p}) \cdot \hat{y} = \frac{1}{N_p} \cos(\theta_z), \quad (5.6)$$

where θ_z is the angle between $-\hat{p}$ and the vertical axis \hat{y} (or zenith angle), and N_p is the normalisation factor.

The energy is sampled according to an inverse law to favour low-energy events, since they are more likely on average:

$$p(E) = \frac{1}{N_E} \frac{1}{E}, \quad (5.7)$$

where N_E is the normalisation factor.

The arbitrary PDFs selected for event sampling do not affect the results so long as we ensure that no region of the phase-space which could be responsible for backgrounds is omitted. Here, the direction and energy distributions are optimised to favour events that are generally more likely to occur given the configuration of our source and detector (i.e. small zenith angles and low energies). When computing the rate for an event, any bias originating from the sampling PDF should be cancelled out when dividing by $p(x, p, E)$ as in Equation 5.4.

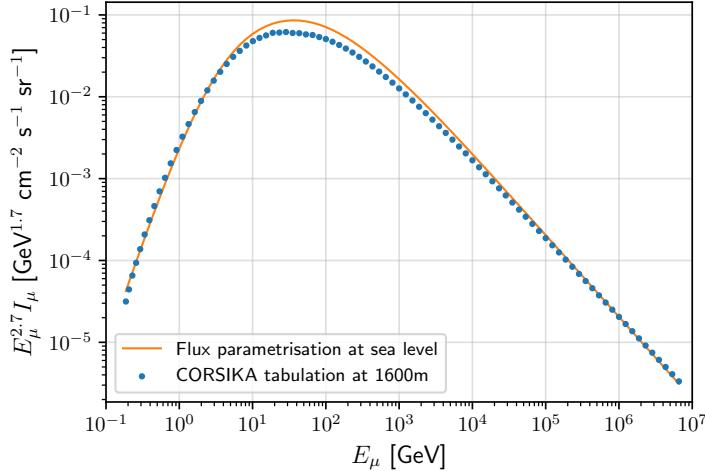


Figure 5.3: Spectrum of the atmospheric muon flux (multiplied by $E_\mu^{2.7}$) calculated at an altitude of 1600m using a CORSIKA simulation. This spectrum, tabulated in both energy and zenith angle, is used as the flux source in the COMET Phase-I backward MC. Here, it is compared with the parametrisation of the muon flux at sea level from Guan et al. [71].

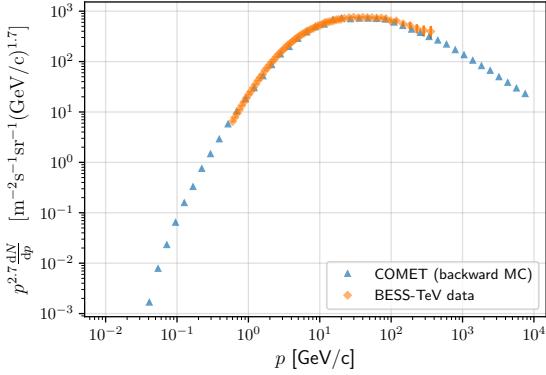
CRV openings sample

The backward MC method gives freedom to generate a specialised sample to increase statistics in any region of phase-space. In our case, in order to focus on the events that might sneak into the detector system, a sample is produced with events generated only in the holes on either side of the CRV. The energy and direction are sampled identically to the envelope sample.

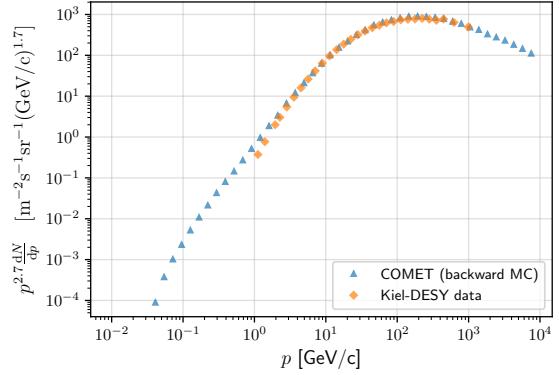
Since the position PDF in this sample is concentrated around the holes, we need to simulate fewer events in order to find a sneaking signal-like track than with the envelope sample. This allows us to obtain reasonable statistics in the background rate study while saving on computational resources and storage space.

5.2.3 Atmospheric muon flux

Estimating rates requires knowledge of ϕ_s , the directional flux at the source. We use tabulated flux data from a CORSIKA [72] simulation as an efficient way to query the flux as a function of energy and direction. The flux data, shown as a function of energy in Figure 5.3, was estimated 1600 m above sea level for energies between 10 MeV and 10 TeV, separately for muons and anti-muons. In the backward simulation, any event which is unable to travel back to this infinite source plane is assumed to have a rate of



(a) Comparison to BESS-TeV spectrometer data of the atmospheric muon flux at a near-vertical zenith angle in Manitoba, Canada [73].



(b) Comparison to the Kiel-DESY spectrometer observation of the atmospheric muon flux at a zenith angle of $75^\circ \pm 7^\circ$ [74] in Hamburg, Germany.

Figure 5.4: Validations of the backward MC-estimated flux with experimental observations of the atmospheric muon flux by two independent experiments. Events from the backward MC simulation are selected to reflect the angular acceptance of the corresponding experiment.

zero.

5.2.4 Validations

In order to ensure that the backward MC simulation yields accurate results, the flux of atmospheric muons estimated around the CRV is compared to two independent measurements of the flux at sea level.

BESS-TeV

The atmospheric muon flux in Manitoba, Canada was measured using the BESS-TeV spectrometer in 2004 [73]. The apparatus only observed muons incident at a near-vertical zenith angle θ_z . In order to compare our simulation results with the experimental data, we select simulated events based on the condition that $\cos \theta_z > 0.98$ to replicate the acceptance of the spectrometer. The comparison of the differential flux as a function of momentum is shown in Figure 5.4a.

Kiel-DESY

The Kiel-DESY spectrometer was used to measure the atmospheric muon flux in Hamburg, Germany in 1975 [74]. This detector measures particles within a zenith angle

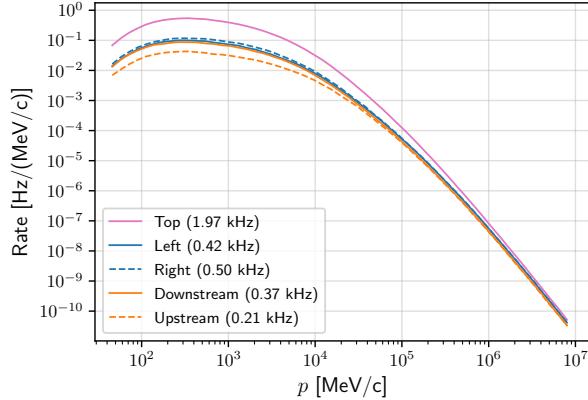


Figure 5.5: Average rate of inbound atmospheric muons over each face of the CRV as a function of momentum. Figures in parentheses in the legend are the integrated rate averaged over each face.

$\theta_z = 75^\circ \pm 7^\circ$ and an azimuthal angle $\phi = 288^\circ \pm 20^\circ$. Similarly to the previous comparison, simulated events are selected to replicate these acceptances such that the flux can be compared. Figure 5.4b shows the comparison of the flux spectra.

The fact that both comparisons are in reasonable agreement with the backward simulation results increases our confidence in the method. Here, we have shown that the backward MC method can effectively compute the flux of events at sea level using knowledge of the flux at an altitude of 1600 m. Note that our simulation setup differs from either experiment by the amount of material present between the detector and the atmospheric source. Hence, an imperfect agreement is expected and the discrepancies visible in either comparison were not investigated in detail.

5.2.5 Absolute rate estimation

From the estimated flux, one can compute the absolute rate using Equation 5.4. Figure 5.5 shows the atmospheric muon rate as a function of momentum on each face of the CRV. The top face is the most exposed, with an integrated rate of 1.97 kHz. Over the whole surface, the estimated event rate is 3.47 kHz. This estimation of the rate accounts for all events inbound on the CRV, regardless of whether or not they enter the CDC or CTH.

5.2.6 μ - e conversion background rate

In order to determine how frequently signal-like tracks are produced by atmospheric muons, we combine forward and backward MC simulations. The forward MC allows us to select events based on detector acceptance criteria, such as trigger conditions. The backward MC is then used to estimate the rate of selected events.

Track reconstruction

After the forward MC step, we also simulate the reconstruction of tracks in the CDC. The algorithm is a helix fit through the positions of CDC hits. In this step, the true hit positions given by the SimG4 Monte Carlo are used. The fit also returns the momentum of the reconstructed trajectory, which is used in the event selection.

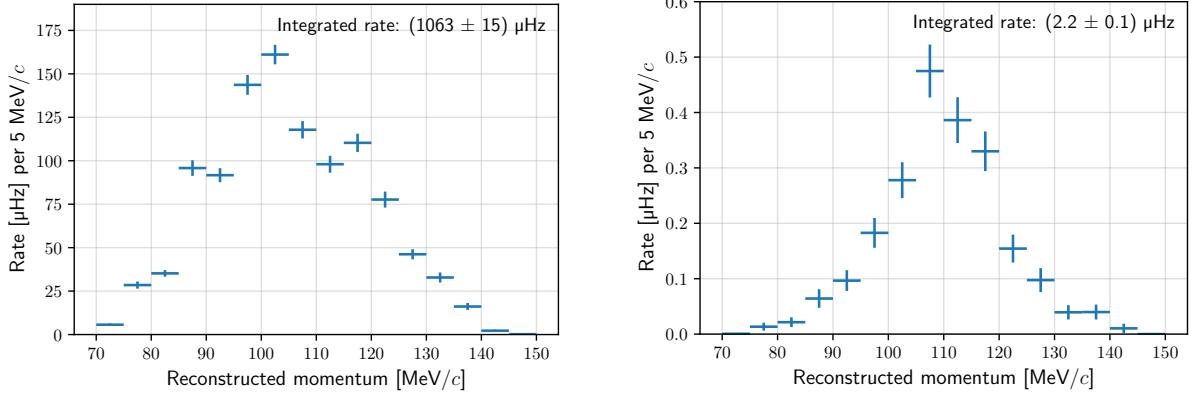
Event selection

The event selection criteria are as follows:

- Fourfold coincidence: four neighbour CTH counters must be hit within a 10 ns window,
- CDC layers: the track must reach up to the 5th layer of CDC wires and no hit should occur on the outermost layer,
- Muon stopping target intersection: the fitted track must intersect the muon stopping target placed in the centre of the CyDet,
- Momentum window: the reconstructed momentum must lie between 55 MeV/ c and 155 MeV/ c .

These criteria reduce the number of events whose flux will be sampled by the backward MC, which is crucial to limit the usage of computational resources. In addition, we obtain a reasonably-sized sample from which we can study the shape and properties of background events.

Other acceptance criteria and inefficiencies of the CyDet system are taken into account by applying multiplicative factors to the estimated background rates. Specifically, the hardware efficiency of Table 2.2 is used, reducing the overall background rate by a factor 0.81. To take into account the fact that the Phase-I trigger is only active during



(a) Without a veto.

(b) Including the effect of the Cosmic Ray Veto detector, assuming an efficiency of 99.99 %.

Figure 5.6: $\mu-e$ conversion background rates from atmospheric muons estimated via backward MC simulation. The error bars show the statistical uncertainty on the sampled flux over backward sampling trials. In the presence of the CRV, the background rate is dominated by events where an atmospheric muon sneaks through the openings.

the $t \in [700, 1170]$ ns window after each proton pulse, we weight the rate by a factor $\epsilon_{\text{timing}} = \frac{8}{9} \frac{1170-700}{1170} = 36\%$, where the term $\frac{8}{9}$ arises from the bunch structure of the J-PARC main ring.

Backward sampling

The flux is sampled by backward MC simulation for all events that pass the above selection criteria. The backward propagation and flux sampling is performed $N = 5000$ times for each event. The flux for each event is the average over these N trials, as discussed in Section 5.1.2. The variation in the sampled flux over these N trials is interpreted as the statistical uncertainty on the average flux for each event. This uncertainty is shown in the resulting plots and figures.

Results

Figure 5.6 shows the estimated background rates as a function of the reconstructed momentum of the signal-like track. We compare the rate in the absence and in the presence of the CRV to demonstrate its importance. Without a veto mechanism, the rate over the entire momentum range is around 1 mHz, which is lowered to 2 μ Hz with the CRV. The spectrum peaks around 105 MeV, the conversion energy, because we have selected events that pass the CTH and CDC trigger criteria. Since the CyDet is heavily

biased toward observing conversion electrons, selected background events tend to have a signal-like signature.

In the next chapter, a complete sensitivity and background study for COMET Phase-I, including the results discussed here, will be presented. Despite the CRV being expected to reject a large fraction of atmospheric muon backgrounds, they remain the most important source of backgrounds in the μ - e conversion search.

Chapter 6

COMET Phase-I Sensitivity and Background Estimates

With the framework to simulate backgrounds from atmospheric muons in place, we can now combine the results into a complete sensitivity and background simulation study for COMET Phase-I. In this chapter, we discuss our simulation samples consisting of signal, decay in orbit (DIO), and atmospheric events, and present our resulting expectations of the performance of Phase-I. All data samples simulated for this study were produced on the CC-IN2P3 cluster.

6.1 Data samples

Three simulation samples are produced for this study: a μ - e conversion sample, a DIO sample and an atmospheric muon sample. All three are simulated in the geometrical world shown in Figure 5.2. This geometry differs from the older TDR design [20] by the number of counters in the CTH. Here, each layer of the CTH has 64 counters instead of the original 48, which mainly affects the detector's geometrical acceptance. In addition, we use the first-stage layout of the CTH, where both layers are composed entirely of plastic scintillation counters. In contrast, the final stage will have Cherenkov counters in the outer layer. As we shall see, this has a considerable impact on the detector's ability to discriminate atmospheric muons from conversion electrons.

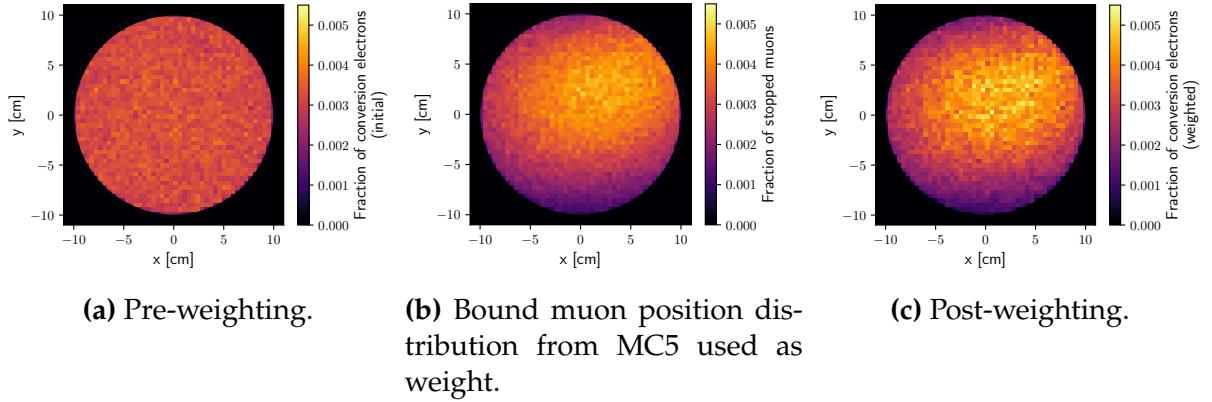


Figure 6.1: Initial position distribution of signal electrons before and after weighting them by the likelihood of a muon being bound in each bin.

6.1.1 $\mu-e$ conversion

Sample

The signal sample is the most straightforward to produce. We initially generate primary electrons with energy $E = 104.97$ MeV uniformly inside the stopping target disks. Their direction is isotropically distributed, as would be the case in the conversion process. A uniform position distribution in each disk is not realistic, because the actual distribution depends on where in the stopping target the muons in the beam come at rest. To account for this, the events are weighted according to the stopping positions of muons recorded in the MC5 simulation. The weighting factor is determined from the relative probability for a muon to stop at the sampled position, which is estimated by histogramming the stopping positions in each disk. Figure 6.1 shows the initially-sampled uniform position distribution, the muon stopping position distribution from MC5, and the result of event weighting. In total, $N_{\text{signal}} = 2 \times 10^6$ events are simulated to compose the signal sample.

Selection

Signal events are selected based on detector acceptance criteria. We first require fourfold coincidence in the CTH. Figure 2.9 shows an example of a conversion event which passes this trigger criterion. The fraction of events remaining defines the geometrical acceptance

$$A_{\text{geom}} \equiv \frac{N_{\text{4-fold}}}{N_{\text{signal}}}. \quad (6.1)$$

In this simulation, the estimated geometrical acceptance is $A_{\text{geom}} = 21\%$. In comparison, the TDR cites $A_{\text{geom}} = 26\%$ with the previous design of the CTH.

We do not fully simulate the reconstruction of signal electron trajectories in the CDC. In order to approximate the effect of reconstruction uncertainties, a smearing is applied to the true momentum of each track. The reconstructed momentum is estimated as $p_r = p_t + x$, where p_t is the true momentum of the electron as it enters the CDC, $x \sim \mathcal{N}(0, \sigma)$, and $\sigma = 200 \text{ keV}/c$ is the expected momentum resolution of the CDC. Since tracks are not properly reconstructed, we do not apply any track quality cuts to select events but instead weight the signal sample by the associated acceptance factor from Table 2.2 to account for the rejection of some events.

In this simulation sample, the initial time for each event does not correspond to the realistic time distribution of the $\mu-e$ conversion process. Hence, events are not selected based on whether they reach the detector within the trigger time window, but we weight the sample by the trigger time window efficiency factor instead. Similarly, the sample is weighted by the trigger and data acquisition efficiency factors to account for the loss of acceptance from hardware effects. Table 2.2 lists the values of these efficiency factors.

6.1.2 Muon decay in orbit

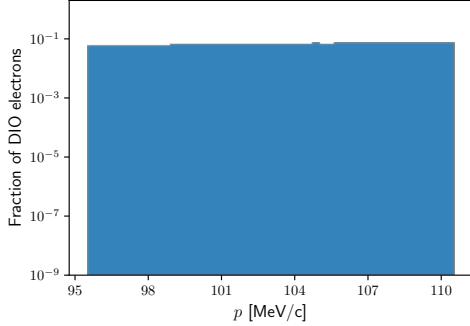
Sample

The DIO sample is similar to the signal sample in that the initial position of signal and DIO electrons is identically distributed. Hence, we also sample uniformly in the stopping target disks and then weight the events according to the MC5 stopping position distribution. Similarly, the direction of DIO electrons is sampled isotropically. The energy distribution is thus the only difference between the two samples.

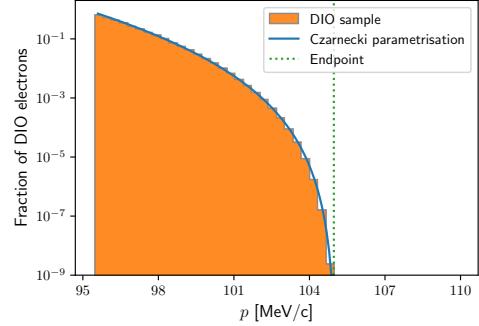
The theoretical energy spectrum of DIO electrons produced in aluminium muonic atoms was investigated by Czarnecki et al. [25]. Here, we use their proposed parametrisation of the energy spectrum around the DIO energy endpoint:

$$P(E_e) = a_5 \delta^5 + a_6 \delta^6 + a_7 \delta^7 + a_8 \delta^8, \quad (6.2)$$

where $\delta \equiv E_\mu - E_e - \frac{E_e^2}{2m_{\text{Al}}}$, $E_\mu = 105.194 \text{ MeV}$, $m_{\text{Al}} = 25133 \text{ MeV}$, with the values of the

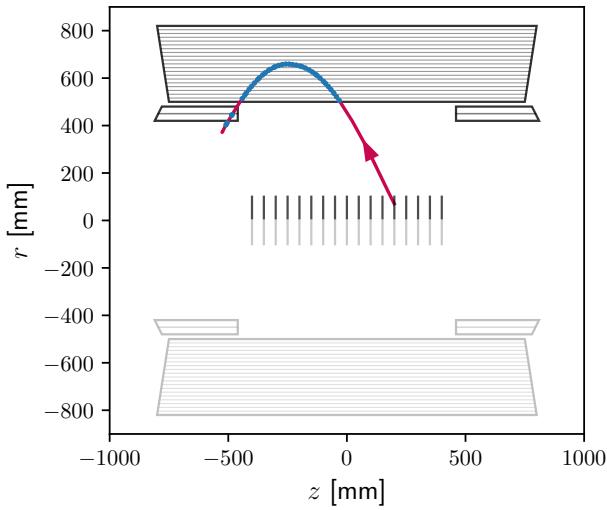


(a) Initial sampling.

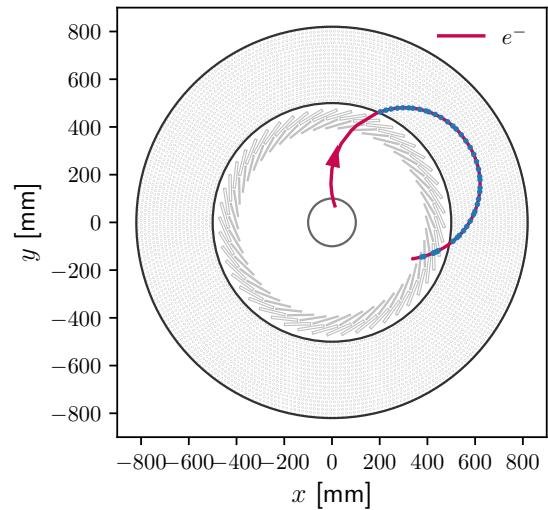


(b) Weighted with the Czarnecki et al. parametrisation [25].

Figure 6.2: Momentum spectrum of electrons in the DIO sample.



(a) $z-r$ projection, where $r = \sqrt{x^2 + y^2}$.



(b) $x-y$ projection.

Figure 6.3: Potential DIO-induced background with momentum $p = 103.7 \text{ MeV}/c$.

four a coefficients provided by the authors. This approximation is valid in the region $E_e > 85 \text{ MeV}$, which is sufficient to cover our whole DIO sample. We use this spectrum parametrisation to weight DIO events based on their energy. Similarly to the position weighting procedure, DIO events are first generated with a uniform energy distribution in the range $E \in [95, 110] \text{ MeV}/c$, and then weighted according to the probability for a DIO electron to have the sampled energy. Figure 6.2 shows the momentum spectrum of DIO electrons before and after weighting.

The DIO sample is composed of $N_{\text{DIO}} = 10^7$ events in total. Figure 6.3 shows an example of a potential background event from DIO where an electron with an energy close to the conversion energy flies into the CyDet system and activates the trigger.

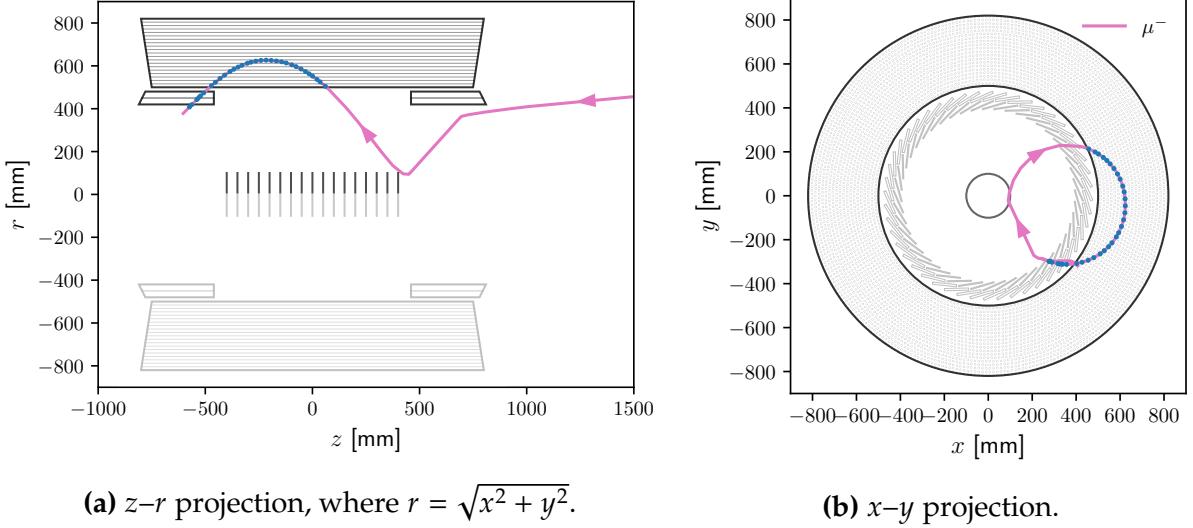


Figure 6.4: Sneaking atmospheric μ^- background with a reconstructed momentum $p = 103.5 \text{ MeV}/c$. The muon enters through the downstream opening of the CRV and scatters in the aluminium CTH support structure. Then, it produces a track in the CDC and a fourfold coincidence in the CTH. Importantly, the trajectory crosses the muon stopping target, making it more difficult to distinguish from the conversion signal.

Selection

The selection of DIO events is identical to the selection of conversion events. We require a fourfold coincidence in the CTH, and then apply a Gaussian smear to the true momentum of the electron to approximate track reconstruction. The same efficiency and acceptance factors as for signal events are applied to weight the sample and determine the absolute background contribution from the DIO process.

6.1.3 Atmospheric muons

Sample

Atmospheric muon events are simulated as discussed in Section 5.2.2. Two samples are produced, one around the entire surface of the CRV and one more densely concentrated on its upstream and downstream openings. Because of how rarely a sampled atmospheric event produces signal-like features, the atmospheric dataset is the largest of the three, with 1×10^9 events generated on the envelope and 1.4×10^9 on the openings.

Selection

As for signal and DIO events, atmospheric events must at least pass the fourfold coincidence criterion. Unlike signal and DIO events, hit information from the CDC is also used to select track candidates. Two criteria apply: the particle must reach at least the 5th innermost layer of the CDC, but it should not hit the outermost layer. Hence, we select tracks with sufficient transverse momentum to be reconstructed, but reject tracks that either have too large a momentum to be a signal event, or appear to enter the CDC from the outside.

In addition, the hit positions are used to reconstruct the track using a helix fit. The fitted trajectory should intersect the muon stopping target to pass the selection. Figure 6.4 shows an example of such an atmospheric event. The helix fit estimates the momentum of the particle, which is used in the analysis stage. Hence, for this sample, we do not apply a Gaussian smear to the true momentum but use the estimated momentum from the fit instead.

Rate estimation

As discussed in Section 5.2.6, the rate for each selected event is estimated by backward MC simulation. The backward propagation and flux sampling is repeated 5000 times for every event, which yields an average flux and a statistical error shown in the results.

6.1.4 Sample weighting

Because they originate from different processes, the three MC samples must be individually weighted in order to determine the absolute contribution from each. The conversion and DIO processes both originate from the muons bound in the stopping target, hence we can express the total number of expected conversion and DIO electrons in terms of the total number of stopped muons N_μ . The number of signal electrons, as a function of the conversion branching ratio $\mathcal{B}_{\text{conversion}}$, is:

$$N_{\text{conversion}} = N_\mu \mathcal{B}_{\text{conversion}} \mathcal{B}_{\text{capture}} f_{\text{coherent}}, \quad (6.3)$$

where $\mathcal{B}_{\text{capture}} = 0.61$ is the branching ratio of nuclear muon capture¹ and $f_{\text{coherent}} = 0.9$ is the fraction of conversions that are coherent.

Similarly, the total number of DIO electrons can be expressed as:

$$N_{\text{DIO}} = N_\mu \mathcal{B}_{\text{DIO}}, \quad (6.4)$$

where $\mathcal{B}_{\text{DIO}} = 1 - \mathcal{B}_{\text{capture}} = 0.39$ is the branching ratio of DIO. Our simulation sample, however, only covers the part of the DIO spectrum where $E_e > 95 \text{ MeV}$, so we cannot simply use N_{DIO} as the sample weight. The proper weight is given by

$$\begin{aligned} N_{\text{DIO}}^{p>95 \text{ MeV}} &= N_{\text{DIO}} P(E_e > 95 \text{ MeV}) \\ &= N_{\text{DIO}} \int_{95 \text{ MeV}}^{E_{\text{endpoint}}} P(E_e) dE_e, \end{aligned} \quad (6.5)$$

where $E_{\text{endpoint}} = 104.973 \text{ MeV}$ is the energy above which no DIO electron can be produced. The integral term in this equation is estimated by numerically integrating the Czarnecki parametrisation of Equation 6.2.

In the case of atmospheric muons, the backward MC procedure yields an absolute rate R for each event. This rate can simply be integrated over the total data acquisition time of the experiment T_{DAQ} to obtain the expected event count:

$$N_{\text{atmospheric}} = \int_0^{T_{\text{DAQ}}} R dt = R \times T_{\text{DAQ}}. \quad (6.6)$$

In the TDR, COMET Phase-I was estimated to reach its sensitivity goal for $T_{\text{DAQ}} = 146$ days. In this study, we keep this value fixed to be able to compare our sensitivity and background estimates with the original study.

Trigger time window efficiencies

In order to take into account the trigger timing window, we assume that atmospheric muons irradiate the detector with a uniform time distribution. Hence, the time window

¹The nuclear muon capture branching ratio appears in this expression because the branching ratio of $\mu-e$ conversion is conventionally normalised to $\mathcal{B}_{\text{capture}}$.

efficiency factor is the average fraction of time when the trigger is active:

$$\epsilon_{\text{time window}}^{\text{atmospheric}} = \frac{8}{9} \frac{1170 - 700}{1170} = 36 \% \quad (6.7)$$

assuming the trigger window is between 700 and 1170 ns, and where the factor $\frac{8}{9}$ arises from the bunch structure of the J-PARC main ring. In contrast, the efficiency factor for conversion and DIO electrons $\epsilon_{\text{time window}}^{\text{conversion} | \text{DIO}} = 30 \%$ is smaller because the time distribution of bound muons is not uniform but peaks around 300 ns, before the trigger becomes active (see Figure 2.2).

6.2 Single event sensitivity

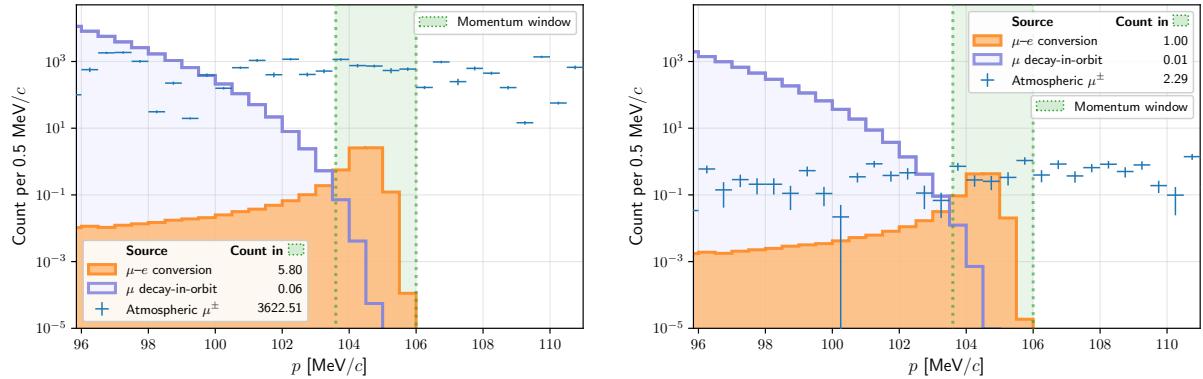
As discussed in Section 2.8, the single event sensitivity (SES) is defined as the value of the $\mu-e$ conversion branching ratio required for the experiment to observe one signal event. It can be expressed in terms of the experimental acceptance $A_{\mu-e}$ and the total number of muons stopped in the stopping target N_μ :

$$\text{SES} = \frac{1}{N_\mu A_{\mu-e} \mathcal{B}_{\text{capture}} f_{\text{coherent}}}, \quad (6.8)$$

where $\mathcal{B}_{\text{capture}} = 0.61$ and $f_{\text{coherent}} = 0.9$.

In this simulation study, we found the geometrical acceptance A_{geom} of the CyDet with the new CTH layout to be reduced from 26 % to 21 %. Hence, the net signal acceptance $A_{\mu-e}$ decreases from 4.1 % to 3.3 %. On the other hand, the yield of stopped muons per proton collision was estimated from the MC5 dataset to be $R_{\mu/p} = 4.86 \times 10^{-4}$, which is slightly higher than the TDR value. Overall, although the total number of bound muons is increased for the same run time, it is outweighed by the decrease in acceptance. Keeping T_{DAQ} fixed at 146 days, we obtain $N_\mu = 1.53 \times 10^{16}$, hence our estimation of the COMET Phase-I sensitivity is

$$\text{SES} = 3.61 \times 10^{-15}. \quad (6.9)$$



(a) Without μ^\pm identification by the CRV, timing window cuts or track quality cuts.

(b) With 99.99 % μ^\pm rejection by the CRV, timing and track quality cuts.

Figure 6.5: Momentum spectrum of conversion, DIO and atmospheric events around the conversion energy, integrated over the Phase-I run time $T_{\text{DAQ}} = 146$ days, assuming $\mathcal{B}_{\text{conversion}} = 3.61 \times 10^{-15}$. Error bars on the atmospheric event bins show the statistical uncertainty from the backward MC flux estimation.

6.3 Signal and background event counts

After applying the efficiency factors and weighting each sample by the total number of expected events as per Section 6.1.4, we are able to determine the absolute contribution from each source over a given data acquisition period.

6.3.1 Momentum spectrum

The absolute event counts from conversion, DIO and atmospheric muon events are plotted as a function of the reconstructed track momentum in Figure 6.5. Note again that for conversion and DIO, p is the smeared momentum of the electron as it enters the CDC, whereas we use the momentum reconstructed via a helix fit for atmospheric muon events.

In order to determine the effect of the CRV on the atmospheric background, as in Figure 6.5b, we apply a weight to each event based on whether it passed through the active material of the CRV. For events that produce hits in the CRV, the weight is the assumed inefficiency of the CRV: $1 - (99.99\%) = 10^{-4}$. Events that sneak into the detector are given a weight of 1 since the CRV cannot veto them. Similarly, to simulate the absence of the CRV, as in Figure 6.5a, all atmospheric muon events have a weight of 1, corresponding to a 0%-efficient CRV.

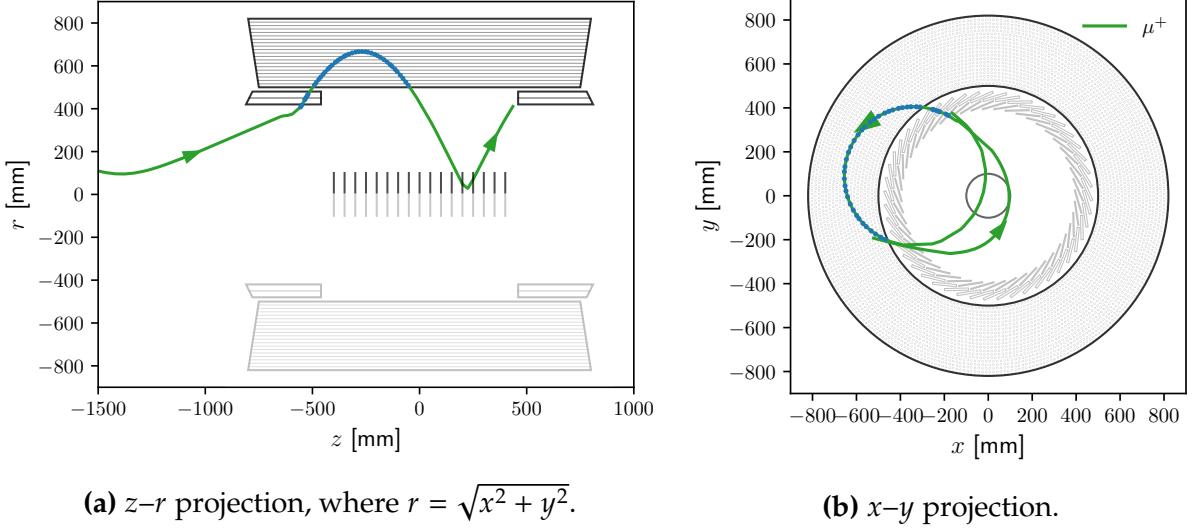


Figure 6.6: Sneaking atmospheric μ^+ background. It enters through the upstream opening of the CRV and scatters in the aluminium CTH support structure before inducing a fourfold coincidence in the CTH, producing a signal-like track in the CDC, and flying through the muon stopping target. The anti-muon thus appears as a time-reversed signal-like event.

The momentum of the candidate track is one of the most important criteria in distinguishing signal electrons from DIO or RMC² electrons, because the energy spectra of the latter two fall off sharply close to the conversion energy. In Phase-I, the momentum window within which an event is counted is $p \in [103.6, 106.0]$ MeV/ c . This cut eliminates the majority of DIO and RMC backgrounds. However, atmospheric background events appear to be uniformly distributed in this range, making the cut ineffective in rejecting them.

Figure 6.5a shows that in the absence of the CRV, atmospheric backgrounds overwhelm the detector and outnumber conversion events by three orders of magnitude in the momentum window. In the presence of the CRV, and assuming an efficiency of 99.99 % (Figure 6.5b), atmospheric backgrounds are suppressed by a factor 10^3 . However, their integrated count still exceeds one, making the $\mu-e$ conversion search impossible unless there is an additional way to identify them.

6.3.2 Particle identification

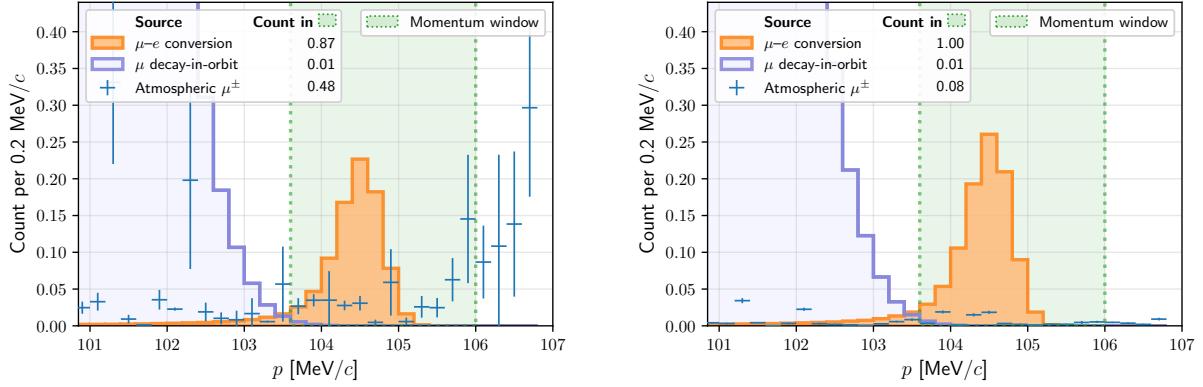
So far, we have assumed that the CyDet system has no way of discriminating between electrons, muons, and anti-muons. However, this is not accurate. At a momentum of $105 \text{ MeV}/c$, an electron has a velocity close to the speed of light, whereas a muon travels only at around $0.7c$. This allows the two to be distinguished when they pass through a Cherenkov detector with the right index of refraction. The final stage of the CTH will have Cherenkov counters on the outer layer, allowing electrons to be discriminated from muons. However, the first stage of the CTH will be composed of only scintillation counters. Hence, initially, the CTH by itself will not suffice in identifying atmospheric muon backgrounds. We now investigate the signatures of atmospheric muons that sneak into the CyDet system without being vetoed by the CRV, and try to determine how impactful identifying the particle type is in lowering the background rate.

Negative atmospheric muons typically produce signal-like tracks in the CyDet by scattering off a dense supporting element, and then mimicking the path of a conversion electron emanating from the muon stopping target, as illustrated in Figure 6.4. Without a Cherenkov-counter layer in the CTH, it is very difficult to identify this type of background. However, our results indicate that these occur 5 to 10 times less frequently than backgrounds from positive atmospheric muons.

Positive atmospheric muons can induce backgrounds by following the reverse path of a conversion electron. An anti-muon in the longitudinal magnetic field would usually gyrate in the opposite way to a negative particle, and hence the track fitting algorithm should easily determine that it has the wrong charge. However, if the anti-muon travels from the CTH to the muon stopping target instead of the opposite, it then follows a time-reversed signal-like trajectory which seems to gyrate identically to a negative particle produced in the disks, as illustrated in Figure 6.6. The only difference between this trajectory and that of a negative particle is the order in which it flies from the CDC to the CTH. Therefore, in the absence of a particle-identifying CTH, one can only rely on timing information to identify this kind of background event.

In order to lower the atmospheric background rate, the identification of atmospheric μ^+ -induced backgrounds via time-of-flight information was studied for COMET

²We chose not to show the radiative muon capture spectrum in our plots because the contribution from RMC is smaller than that of DIO by a factor of 5 [20].



(a) With track direction identification: 89 % of μ^+ -induced backgrounds are rejected and 87 % of signal events are preserved.

(b) With particle type identification by the Cherenkov layer of the CTH, assuming a μ^\pm rejection efficiency of 99 %.

Figure 6.7: Signal and background event counts versus reconstructed track momentum, over $T_{\text{DAQ}} = 146$ days and assuming $\mathcal{B}_{\text{conversion}} = 3.61 \times 10^{-15}$.

Phase-I. The principle is to determine the direction of the track using timing information from the CDC and CTH. Depending on the direction of the particle, hits in the CDC will either occur earlier or later than hits in the CTH. Each track is fitted under two hypotheses: either the particle is a negative electron flying from the muon stopping target to the CTH via the CDC, or it is a positive muon doing the same in reverse. The quality of the resulting fit depends on the hypothesis, and thus the difference in quality between the two fits can be used to distinguish electrons from anti-muons. An early simulation study of this method indicates that 89 % of μ^+ -induced backgrounds can be rejected using timing information, while preserving 87 % of signal events [75].

In Figure 6.7a, we use this result to show the effect of μ^+ identification on the background spectrum. The integrated background event count from atmospheric muons is reduced to 0.48 in the momentum window. Figure 6.7b shows the effect of a particle-identifying Cherenkov layer in the CTH, assuming an efficiency of 99 % and no negative effect on signal acceptance. The background count falls to 0.08 for one signal event.

6.3.3 CRV efficiency

The CRV efficiency determines the fraction of atmospheric events which can enter the CyDet through the active parts of the CRV while not being vetoed. Although we have assumed so far that the CRV successfully rejects 99.99 % of atmospheric muon events,

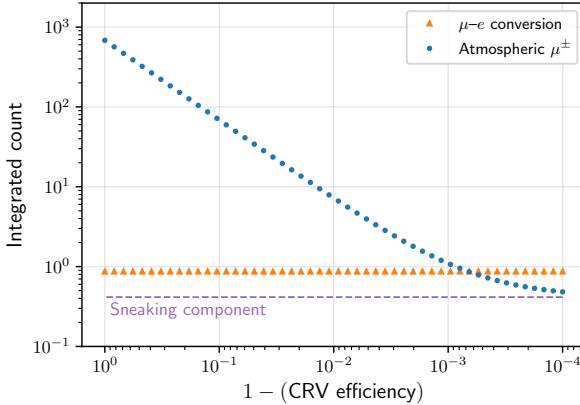


Figure 6.8: Integrated atmospheric background count versus CRV inefficiency (note the reversed horizontal axis), assuming no Cherenkov CTH layer and 89 % μ^+ rejection rate by direction identification. The lower bound arises from sneaking events, which are not detected by the CRV.

PID \ ϵ_{CRV}	99.9 %	99.99 %
None	3.22	2.29
89 % μ^+ (direction)	1.10	0.48
99 % μ^\pm (Cherenkov)	0.57	0.08

Table 6.1: Summary of atmospheric background counts for $T_{\text{DAQ}} = 146$ days under different scenarios for particle identification (PID) and CRV efficiency (ϵ_{CRV}).

this will not necessarily be realised in practice. Figure 6.8 shows the background event count as a function of the CRV efficiency. An efficient CRV helps to reject the vast majority of atmospheric backgrounds. An efficiency of at least 99.9 % seems necessary to suppress the background counts to fewer than one. Beyond 99.99 %, the contribution from sneaking events becomes significant, and reducing the background rate is better tackled by means of particle identification, such as with the Cherenkov CTH layer.

6.3.4 Summary

Table 6.1 summarises the atmospheric background event counts for $T_{\text{DAQ}} = 146$ days under various run conditions. Obviously, one major aspect of identifying atmospheric muons is the CRV, and the background rate greatly depends on its efficiency.

The problem caused by sneaking muons is more subtle. However, it can still plague the conversion search by producing on the order of one signal-like event during data

acquisition. The most frequent sneaking backgrounds are produced by anti-muons that disguise as negative particles by flying from the detector to the muon stopping target. These events can, however, be distinguished from conversion electrons via a direction identification technique using timing information from the CDC and CTH. Being able to identify 89 % of μ^+ -induced events in this manner reduces the background counts to 1.10 with a 99.9 %-efficient CRV or 0.48 with a 99.99 %-efficient CRV.

Eventually, when the CTH reaches its final stage with an outer Cherenkov layer, muons will be more effectively distinguished from electrons. If the CTH allows us to reject 99 % of all μ^\pm events, the background count could be as low as 0.08, with a 99.99 %-efficient CRV.

Chapter 7

Discussion

This thesis presents two main original contributions: we designed a fake Monte Carlo data generator for the CDC using machine learning, and we estimated the rate of atmospheric muon backgrounds in the COMET Phase-I $\mu\text{-}e$ conversion search using a backward MC simulation technique. In this chapter, we discuss our main findings, the limitations we encountered and ideas toward future studies on either topic.

7.1 Data augmentation with machine learning

7.1.1 Summary

The CDC GAN, discussed in Chapter 4, provides a way to produce synthetic Monte Carlo data in the Cylindrical Drift Chamber. The design of this generative algorithm is based on Generative Adversarial Networks, a technique which is more and more widely used in HEP and in other domains to produce fake samples by learning from an existing dataset of real samples. The CDC GAN is trained on MC-simulated hits using the WGAN-GP algorithm. Its discriminator and generator are convolutional neural networks that can process ordered sequences of hits, which allows the GAN to learn both from per-hit features, and from the relations between multiple consecutive hits. Once trained, the GAN is evaluated in multiple ways. Firstly, the distributions of the synthetic data are compared to those of the training dataset. In addition, we investigated quality metrics, which typically involve an external performance criterion implemented by a third neural network trained independently. These evaluation methods allow us to determine which aspects of the training data are adequately modelled by the GAN,

and where there are discrepancies. This then helps to guide the design of the model to increase the quality of generated samples.

After training, the CDC GAN provides a way to produce original hit data at least six orders of magnitude more quickly than by means of traditional MC simulation, using GPUs. We are therefore able to generate large samples of background events very efficiently. This can be useful, for instance, in studies which require a pure signal sample to be overlaid onto an ambient background, e.g. to optimise a hit filtering algorithm. Alternatively, GAN-generated data can serve to compose a mock dataset, combining signal events with all sources of backgrounds into one large-scale simulation sample used to test the readiness of our data-processing chain prior to the data acquisition period.

The CDC GAN design and architecture are generalisable to other detector systems that have discrete elements. For instance, the CTH is somewhat analogous to the CDC if we treat the scintillation counters like the cells of the CDC. One difference is that distance of closest approach (doca) has meaning for CDC hits but not for CTH hits. Although it is not discussed in detail in this thesis, a CTH GAN prototype was developed based on this analogy, with the `wire` feature acting as the counter index, and the `doca` feature removed. Based on early evaluations, we observed similar performance and sample quality as exhibited by the CDC GAN.

7.1.2 Limitations and recommendations

Training data preselection

The CDC GAN is only trained on a restricted part of the CDC hit data. Reconstructible tracks, defined by their momentum $p > 50 \text{ MeV}/c$, are removed from the training dataset in order to prevent the model from generating patterns that are too similar to the conversion signal without the proper associated truth information. Unlike high-momentum tracks whose source (conversion, DIO, RMC, etc.) matters in studies of the experimental outcomes, the origin of low-momentum tracks is typically irrelevant, hence those hits can be grouped into one large “noise” category. The GAN then learns only from hits in that category, and all GAN-generated data can safely be labelled as noise as well. However, training on a subset of hit data prevents us from using the GAN to generate an entire background dataset: the GAN must be complemented by

real MC hits of reconstructible tracks, which is computationally expensive to produce.

Training a conditional GAN [76] with the ability to generate labelled hits was considered. However, the concern that the model would be unable to grasp the physical laws at play in the hit patterns of high-momentum particles was too great, hence this solution was discarded. In a future study, a physics-informed conditional GAN, where the dynamics of charged particle transport in a magnetic field are encoded into the discriminator in order to regulate the generator, could be considered to build a universal background hit generator.

Feature representation

One of the difficulties in designing the GAN was in determining the best way to handle the discrete feature, wire index. Multiple methods were considered, such as using actual or transformed wire positions as a real-valued feature, or using a trainable embedding matrix to make the GAN learn its preferred representation of the geometry. Eventually, we chose to represent the wire index as a one-hot encoded vector and use a matrix of wire positions to assist the model in locating wires, since that yielded the most faithful results. However, this aspect of the GAN design appears to be one of the major quality-limiting factors because of the large dimensionality of this one-hot vector compared to the other features.

A better design would be achieved by representing the wire index as a low-dimensional quantity while informing both networks of the location of hits, however a way to fulfil both requirements at once was not found in our investigations.

Hit sequence arrangement

In order to train the GAN as efficiently as possible, we arrange hits from multiple particles together into fixed-length sequences that can be processed quickly convolutional layers. This arrangement is not physical: unrelated hits from different particles become neighbours in the sequence. When processed by a convolution kernel, consecutive hits will be treated as if there was some relationship between them (as for neighbour pixels on an image), but this is sometimes false. This leads the GAN to learn patterns that are artificial and only the product of our arrangement.

Keeping hits from different tracks separate would most likely yield a better model

of the underlying tracks. However, it would also prevent the networks from being built with convolutional layers. Techniques developed and used in natural language processing, such as transformers [77], could be applied here to build a model that can process variable-length hit sequences (in analogy with sentences) while remaining efficient to train.

Quality evaluation

Evaluating samples generated by the CDC GAN is inherently difficult. Unlike visual or audio data, one cannot precisely estimate the sample quality perceptually. Hence, in Section 4.3.8, we compare feature distributions and use KL divergence and more complex metrics to define a quantitative measure of sample quality. Although these methods help in refining the GAN architecture and in tuning training parameters, no metric can affirm that the generated data is physical enough to be used to study the experiment and its outcomes. This is in part due to the fact that the GAN generates unlabelled hits from the broad category of “noise”, hence one can only hope to determine if generated hits belong in that category, or not.

One option could be to condition the GAN [76] on particle type and momentum. If the model is able to generate a sequence of hits based on a prompt, such as “electron with $p = 15 \text{ MeV}/c$ ”, then it becomes easier to verify whether the hit sequence is faithful to the prompt by comparing to real samples of a similar description *and* by checking that the physical rules are respected (e.g. track curvature, total energy deposit, etc.). Of course, conditioning comes at the cost of a more complex network design and training procedure.

7.2 Atmospheric muon backgrounds

7.2.1 Summary

In Chapter 5, we presented the backward Monte Carlo simulation principle, method, validations, and results when applied to the COMET Phase-I setup. Backward MC consists in the time-reversed transport of particles from a volume of interest to a source where the flux is known. Since events are generated arbitrarily near the volume of interest, this provides a way to focus computational resources on important events. In

this study, we used backward MC transport of atmospheric muons in order to determine the rate of cosmic ray-induced backgrounds in the μ - e conversion search. Events are generated from the surface of the Cosmic Ray Veto in Phase-I and transported to a 1600 m-high atmospheric plane where a pre-computed flux model is sampled. The flux estimated at the surface of the CRV is then compared with experimental data from the BESS-TeV and Kiel-DESY spectrometers in order to validate that the backward sampling yields realistic results.

Then, in Chapter 6, we combine the BMC method with a μ - e conversion sample and a DIO sample in order to determine how each process contributes in the event counts recorded during the Phase-I data acquisition run. The analysis of signal events yields our estimate of the single event sensitivity of the COMET Phase-I experiment. Integrating over the entire run time of Phase-I, we find that atmospheric muon backgrounds outnumber the next largest source of background, DIO, by an order of magnitude or more depending on the exact conditions. The main factor in eliminating atmospheric backgrounds is the efficiency of the CRV, followed by the CyDet system's ability to distinguish between electrons and muons. In the absence of a particle-identifying Cherenkov layer in the CTH, our results summarised in Table 6.1 suggest that the background count is at least 0.48 over 146 days, versus one signal event assuming $\mathcal{B}_{\text{conversion}} = 3.61 \times 10^{-15}$. With the final-stage CTH layout, the background count falls to 0.08 assuming a 99 % μ^\pm identification rate.

Backward MC simulation is a powerful tool which enabled us to determine that extremely rare conversion-like events can be produced by atmospheric muons sneaking into the COMET detector system. These events can be an important obstacle for the COMET Phase-I conversion search if no additional way to reject them, such as particle type identification, is put in place. Using a standard MC simulation where events are generated far from the detector system, those rare events that get past the CRV and produce signal-like tracks would most likely not appear at all in the sample. Hence, the backward MC method is absolutely key in this study and should continue to prove useful in COMET toward Phase-I and Phase-II. More generally, any experiment searching for rare processes and which is susceptible to cosmic ray-induced backgrounds will also benefit from using backward MC to identify the kinds of atmospheric events that are likely to mimic their signal, and estimate their frequency.

7.2.2 Limitations and recommendations

Backward transport

In our study, we used backward MC simulation to transport muons from the surface of the CRV to the atmospheric plane. The assumption associated with this procedure is that the muon is able to travel all the way from the atmospheric source to the surface of the CRV. In reality, this assumption does not necessarily hold and there is expected to be some fraction of atmospheric muons which will interact close to the CRV and whose products will induce a background event. These events do not contribute toward our results. Therefore, we are underestimating the background rate from cosmic rays, but by how much is difficult to determine.

This bias could be addressed by sampling the muons not at the surface of the CRV, but a little farther away. However, doing so would lead to more events missing the detector, as is the case in forward MC, and therefore to wasted computation. Already, in our study, very few sampled events pass the CyDet selection criteria: around 1 in 10^5 in the envelope sample, 1 in 10^7 for sneaking events. Pushing the sampling surface outward would decrease these ratios and reduce our statistics further. Nevertheless, it can be done in the future in order to estimate the extent of the bias.

Sampling distributions

Atmospheric muons are sampled around the CRV with arbitrary energy and direction distributions. Our selection shows that only one event in 100 000 produces a fourfold coincidence and a track in the CyDet system. In a future iteration of this study, one could determine if there is some region of the sampling phase space which never contributes to the selected events: perhaps very high or very low energy events, or events which are incident upon the CRV within a specific range of azimuthal or zenith angle. Eliminating this region of phase space from the sampling distribution function would increase the fraction of simulated events which pass the selection, thus increasing the statistics on events of interest for the same computational cost.

Event selection

To identify background candidates, events are selected that have a fourfold coincidence in the CTH and an associated track in the CDC. When reconstructing a trajectory, the track fitting algorithm ignores hits produced by other particles in the CDC during the same event. Instead of using a track finding algorithm, we use truth information from the MC simulation in order to select hits from the right particle and provide them to the track fitting function. This has caused at least one known inaccuracy in our study: events where a high energy muon produces a shower in the CDC sometimes induce a fourfold coincidence, thus they pass our selection criteria and may be counted as background events. In realistic conditions, events that suddenly saturate a large fraction of CDC channels, such as a high-energy shower, will most likely be discarded because tracks cannot be reconstructed from the event. Therefore, these should not have been counted as part of the integrated background, for instance in Figures 6.5a and 6.8. These high energy muons typically fly straight through the CyDet system and hit the CRV along the way. Hence, although this issue in our selection is likely to have caused a slight overestimation of the total background rate, we expect the sneaking component to be unaffected.

Reconstruction and detector response

Various aspects of our study are not faithful to the actual conditions of Phase-I. On the reconstruction side, for the atmospheric muon sample, we use a helix fit to reconstruct the trajectory of each event where a fourfold coincidence is produced. During data acquisition, a more complex fitting algorithm, based on Kalman filtering, will be used. Similarly, for the conversion and DIO samples, we do not perform reconstruction but naively smear the true momentum of the track using the design momentum resolution of the CDC. Because of those simplifications, we expect the resulting momentum spectra to differ slightly in the real conditions of Phase-I.

Although ICEDUST has the capability to fully simulate each sub-detector's response to incoming energy deposits, digitisation and calibration were not simulated in this study. Instead, we simplified those processes by applying a hardware efficiency factor to all samples. Similarly, the timing of events is also disregarded in our simulation. Instead of assigning a realistic time of production to conversion, DIO and atmospheric

events, and then selecting events that occur in the trigger time window, we apply an efficiency factor to each sample based on their expected time distribution and the start and end time of the trigger window.

Hence, our study makes use of estimates based on other studies within the collaboration, rather than conducting fully realistic calculations in the detector response and reconstruction aspects. In future studies, and as the Phase-I data acquisition period approaches, it will be especially important to rigorously simulate these effects to assess the performance of the experiment as accurately as possible.

Chapter 8

Conclusion

Since the discovery of the muon in 1937, our understanding of its properties has been steadily evolving. Today, although it fits nicely into the Standard Model of particle physics, some questions remain standing, and recent tensions observed between theory and experiment are suggesting that the SM does not give a complete picture of the muon’s underlying nature. Charged lepton flavour violation is one of the ways in which Nature might deviate from the SM. If CLFV were observed, it would be a clear indication of new physics. Neutrino-less muon-to-electron conversion is one of the most sensitive channels to search for CLFV as its signature is a mono-energetic electron with few sources of potential backgrounds. According to theoretical constraints, this process ought to be suppressed beyond experimental sensitivities. Therefore, any observation of the $\mu-e$ conversion signal would yield clear evidence for CLFV and hence guide the field toward a better understanding of physics beyond the Standard Model.

The COMET Phase-I experiment will soon start its data acquisition run toward the search for muon-to-electron conversion in aluminium. It is expected to be a hundred times more sensitive to $\mu-e$ conversion than the previous best measurement by SINDRUM II. To achieve this sensitivity, COMET Phase-I will produce an intense pulsed muon beam directed toward an aluminium target to create muonic aluminium atoms, from which conversion electrons may emerge. The Cylindrical Detector, which surrounds the muon stopping target, is designed to clearly identify conversion electrons while rejecting as many experimental backgrounds as possible.

In order to cover the needs of the experiment in terms of simulation, calibration, reconstruction, data formats and data analysis, the COMET collaboration develops a

comprehensive set of software utilities named ICEDUST. The work presented throughout this thesis relies heavily on Monte Carlo (MC) data produced by ICEDUST’s GEANT4 simulations of Phase-I. MC simulations allow us to study the experiment outcomes and to optimise its design before assembly. However, they are computationally expensive and typically only allow us to simulate a small fraction of the amount of data expected to be collected. Additionally, traditional MC sampling is particularly inefficient when the source of events is far from the detector system, as is the case when considering cosmic ray-induced backgrounds. This thesis aims to address these two limitations by partially circumventing the brute force MC method.

We investigate a novel approach to the mass-production of simulation data based on Generative Adversarial Networks. We propose a neural network generator of hits for the Cylindrical Drift Chamber, which can produce synthetic energy deposits in the detector at a rate 10^6 times higher than the ICEDUST simulation. The machine learning model is trained on a sample of simulated hits and learns from their features and relationships. The trained model allows us to generate only a subset of the hits produced by MC simulation, but far more efficiently than was previously achievable. Our ambition for this work and future developments is to enable the production of entire mock datasets, which can be used by the collaboration in anticipation of data acquisition runs.

Additionally, we present a study of the backgrounds caused by cosmic ray-induced atmospheric muons in the COMET Phase-I muon-to-electron conversion search. We use a backward Monte Carlo simulation method to efficiently estimate the flux of atmospheric muons near the detector system. We then perform a full analysis of the event count contributions from $\mu-e$ conversion, muon decay-in-orbit, and atmospheric muons. Our findings suggest that an efficient rejection of atmospheric backgrounds by the Cosmic Ray Veto, combined with particle-type identification by the Cylindrical Detector, are necessary for the Phase-I search to succeed. Under optimistic conditions, namely with a 99.99 %-efficient CRV and 99 % muon identification rate by the CyDet, we expect 0.08 atmospheric background events during the 146 days of data acquisition. According to our analysis, during this run time, COMET Phase-I will be able to achieve a single event sensitivity of 3.61×10^{-15} .

Our most important contributions have been integrated into the main ICEDUST

software repository and documented for others to use and build upon. As the COMET experiment goes into the Phase-I measurement period, we hope that these techniques will help to refine our understanding of the data and that they can be applied in preparation for Phase-II and beyond.

Appendix A

Integration of backward Monte Carlo software into ICEDUST

A.1 Original implementation

The backward Monte Carlo simulation software was originally written and applied to the COMET Phase-I experiment by Valentin Niess, author of [69, 70]. Figure A.1 shows the layout of the original set of software layers used to run backward MC simulations in the Phase-I world.

At the core of the backward Monte Carlo simulation is the backward transport engine, PUMAS [70]. PUMAS is responsible for the propagation of particles through matter and the computation of MC weights based on the path taken and the interactions undergone, as discussed in Section 5.1.2. Because it is designed exclusively for this purpose, PUMAS does not provide utilities for navigating through complex geometries or sampling the flux from pre-computed tables. Another package, GOUPIL, takes on this role as an interface to PUMAS, and allows the user to specify a geometry, a source flux, as well as the local topography and geomagnetic field around the experiment’s location. An additional interface between GOUPIL and GEANT4, goupil-geant4, was developed to allow the user to control backward MC runs with the GEANT4 macro user interface and to load and navigate the geometry from a GDML file. The final layer, resample-simg4, provides a command-line interface and enables events generated by SimG4 to be read as input for the backward MC simulation to estimate their flux.

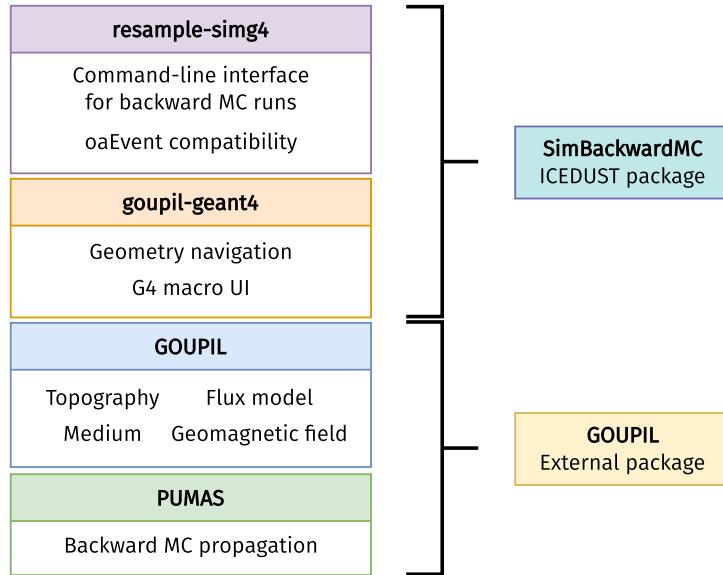


Figure A.1: Software packages that compose the original backward MC simulation stack for the Phase-I study, and organisation of their integrated ICEDUST counterparts.

A.2 Integration

Figure A.1 also shows how the original software was arranged to become part of the ICEDUST framework. Since the PUMAS and GOUPIL packages are unlikely to undergo quick iterative changes over time, they are integrated into an external package. The functionalities of the upper two layers of the stack are combined into a new ICEDUST package called SimBackwardMC.

The way in which the layers are organised is not the only difference between the original and current software. Three major changes were implemented during the integration in order to bring the original software closer to the ICEDUST workflow.

- **Geometry format:** originally read from GDML files, the geometry is now created from the same classes and macro files that define the simulation world in forward MC runs by the SimG4 package. This additionally ensures that the geometry used in forward and backward MC simulations is identical.
- **Entry point:** the main `resample-simg4` entry point was previously a Lua script executed via LuajIT which defined the command-line interface, configured the backward simulation run, and processed the output. SimBackwardMC uses a standard ICEDUST event loop executable as its entry point to handle command-line

arguments, read events from input oaEvent files and output the results of the backward simulation.

- **Output format:** resample-simg4 wrote the results of backward sampling to a text file in a format similar to CSV. SimBackwardMC writes to a RooTracker file instead, since that is a standard ICEDUST format and a good fit to store the kinematics of processed events along with their estimated flux.

The integrated backward MC software presented here was used to obtain the results of Chapters 5 and 6. It is also currently being used in the work of others, for instance to estimate the atmospheric muon flux at various locations around the Phase-I detector system and validate the results with real measurements. Through our standardisation of some parts of the original software and the addition of specific documentation, we hope that the SimBackwardMC package will remain a readily accessible tool for future cosmic ray studies in COMET.

Bibliography

- [1] Seth H. Neddermeyer and Carl D. Anderson. "Note on the Nature of Cosmic-Ray Particles". In: *Phys. Rev.* 51 (10 May 1937), pp. 884–886. doi: [10.1103/PhysRev.51.884](https://doi.org/10.1103/PhysRev.51.884). URL: <https://link.aps.org/doi/10.1103/PhysRev.51.884>.
- [2] Y. Nishina, M. Takeuchi, and T. Ichimiya. "On the Nature of Cosmic-Ray Particles". In: *Phys. Rev.* 52 (11 Dec. 1937), pp. 1198–1199. doi: [10.1103/PhysRev.52.1198](https://doi.org/10.1103/PhysRev.52.1198). URL: <https://link.aps.org/doi/10.1103/PhysRev.52.1198>.
- [3] J. C. Street and E. C. Stevenson. "New Evidence for the Existence of a Particle of Mass Intermediate Between the Proton and Electron". In: *Phys. Rev.* 52 (9 Nov. 1937), pp. 1003–1004. doi: [10.1103/PhysRev.52.1003](https://doi.org/10.1103/PhysRev.52.1003). URL: <https://link.aps.org/doi/10.1103/PhysRev.52.1003>.
- [4] Hideki Yukawa. "On the Interaction of Elementary Particles. I". In: *Progress of Theoretical Physics Supplement* 1 (Jan. 1955), pp. 1–10. ISSN: 0375-9687. doi: [10.1143/PTPS.1.1](https://doi.org/10.1143/PTPS.1.1). eprint: <https://academic.oup.com/ptps/article-pdf/doi/10.1143/PTPS.1.1/5310694/1-1.pdf>. URL: <https://doi.org/10.1143/PTPS.1.1>.
- [5] C. M. G. Lattes et al. "Processes involving charged mesons". In: *Nature* 159.4047 (May 1947), pp. 694–697. ISSN: 1476-4687. doi: [10.1038/159694a0](https://doi.org/10.1038/159694a0). URL: <https://doi.org/10.1038/159694a0>.
- [6] Cássio Leite Vieira and Antonio Augusto Passos Videira. "Carried by History: Cesar Lattes, Nuclear Emulsions, and the Discovery of the Pi-meson". en. In: *Physics in Perspective* 16.1 (Mar. 2014), pp. 3–36. ISSN: 1422-6960. doi: [10.1007/s00016-014-0128-6](https://doi.org/10.1007/s00016-014-0128-6). URL: <https://doi.org/10.1007/s00016-014-0128-6> (visited on 11/29/2022).
- [7] Toshinori Mori. "Final results of the MEG experiment". In: *arXiv preprint arXiv:1606.08168* (2016).

- [8] B. Abi et al. "Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm". In: *Phys. Rev. Lett.* 126 (14 Apr. 2021), p. 141801. doi: [10.1103/PhysRevLett.126.141801](https://doi.org/10.1103/PhysRevLett.126.141801). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.126.141801>.
- [9] Sz. Borsanyi et al. "Leading hadronic contribution to the muon magnetic moment from lattice QCD". In: *Nature* 593.7857 (May 2021), pp. 51–55. issn: 1476-4687. doi: [10.1038/s41586-021-03418-1](https://doi.org/10.1038/s41586-021-03418-1). URL: <https://doi.org/10.1038/s41586-021-03418-1>.
- [10] R. Aaij et al. "Test of lepton universality in beauty-quark decays". In: *Nature Physics* 18.3 (Mar. 2022), pp. 277–282. issn: 1745-2481. doi: [10.1038/s41567-021-01478-8](https://doi.org/10.1038/s41567-021-01478-8). URL: <https://doi.org/10.1038/s41567-021-01478-8>.
- [11] André de Gouvêa and Petr Vogel. "Lepton flavor and number conservation, and physics beyond the standard model". In: *Progress in Particle and Nuclear Physics* 71 (2013). Fundamental Symmetries in the Era of the LHC, pp. 75–92. issn: 0146-6410. doi: <https://doi.org/10.1016/j.ppnp.2013.03.006>. URL: <https://www.sciencedirect.com/science/article/pii/S0146641013000252>.
- [12] A. E. Cárcamo Hernández et al. "Is it possible to explain the muon and electron $g - 2$ in a Z' model?" In: *Phys. Rev. D* 101 (11 June 2020), p. 115016. doi: [10.1103/PhysRevD.101.115016](https://doi.org/10.1103/PhysRevD.101.115016). URL: <https://link.aps.org/doi/10.1103/PhysRevD.101.115016>.
- [13] Basabendu Barman et al. "Correlating the anomalous results in $b \rightarrow s$ decays with inert Higgs doublet dark matter and the muon $g - 2$ ". In: *Phys. Rev. D* 100 (11 Dec. 2019), p. 115010. doi: [10.1103/PhysRevD.100.115010](https://doi.org/10.1103/PhysRevD.100.115010). URL: <https://link.aps.org/doi/10.1103/PhysRevD.100.115010>.
- [14] Y. Fukuda et al. "Evidence for Oscillation of Atmospheric Neutrinos". In: *Phys. Rev. Lett.* 81 (8 Aug. 1998), pp. 1562–1567. doi: [10.1103/PhysRevLett.81.1562](https://doi.org/10.1103/PhysRevLett.81.1562). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.81.1562>.
- [15] S. L. Glashow, J. Iliopoulos, and L. Maiani. "Weak Interactions with Lepton-Hadron Symmetry". In: *Phys. Rev. D* 2 (7 Oct. 1970), pp. 1285–1292. doi: [10.1103/PhysRevD.2.1285](https://doi.org/10.1103/PhysRevD.2.1285). URL: <https://link.aps.org/doi/10.1103/PhysRevD.2.1285>.

- [16] R.H. Bernstein and Peter S. Cooper. “Charged lepton flavor violation: An experimenters guide”. In: *Physics Reports* 532.2 (2013). Charged Lepton Flavor Violation: An Experimenter’s Guide, pp. 27–64. issn: 0370-1573. doi: <https://doi.org/10.1016/j.physrep.2013.07.002>. URL: <https://www.sciencedirect.com/science/article/pii/S0370157313002688>.
- [17] Andrzej Czarnecki, William J. Marciano, and Kirill Melnikov. “Coherent muon-electron conversion in muonic atoms”. In: *AIP Conference Proceedings* 435.1 (1998), pp. 409–418. doi: <10.1063/1.56214>. eprint: <https://aip.scitation.org/doi/pdf/10.1063/1.56214>. URL: <https://aip.scitation.org/doi/abs/10.1063/1.56214>.
- [18] A. M. Baldini et al. “The design of the MEG II experiment”. In: *The European Physical Journal C* 78.5 (May 2018), p. 380. issn: 1434-6052. doi: <10.1140/epjc/s10052-018-5845-6>. URL: <https://doi.org/10.1140/epjc/s10052-018-5845-6>.
- [19] K. Arndt et al. “Technical design of the phase I Mu3e experiment”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1014 (2021), p. 165679. issn: 0168-9002. doi: <https://doi.org/10.1016/j.nima.2021.165679>. URL: <https://www.sciencedirect.com/science/article/pii/S0168900221006641>.
- [20] The COMET Collaboration et al. “COMET Phase-I technical design report”. In: *Progress of Theoretical and Experimental Physics* 2020.3 (Mar. 1, 2020), p. 033C01. issn: 2050-3911. doi: <10.1093/ptep/ptz125>. URL: <https://academic.oup.com/ptep/article/doi/10.1093/ptep/ptz125/5805094> (visited on 01/29/2021).
- [21] L. Bartoszek et al. *Mu2e Technical Design Report*. 2015. arXiv: <1501.05241> [physics.ins-det].
- [22] E. P. Hincks and B. Pontecorvo. “Search for Gamma-Radiation in the 2.2-Microsecond Meson Decay Process”. In: *Phys. Rev.* 73 (3 Feb. 1948), pp. 257–258. doi: <10.1103/PhysRev.73.257>. URL: <https://link.aps.org/doi/10.1103/PhysRev.73.257>.
- [23] A. Knecht, A. Skawran, and S. M. Vogiatzi. “Study of nuclear properties with muonic atoms”. In: *The European Physical Journal Plus* 135.10 (Oct. 2020), p. 777. issn: 2190-5444. doi: <10.1140/epjp/s13360-020-00777-y>. URL: <https://doi.org/10.1140/epjp/s13360-020-00777-y>.

- [24] H.C. Chiang et al. "Coherent and incoherent (e^- , e^-) conversion in nuclei". In: *Nuclear Physics A* 559.4 (1993), pp. 526–542. issn: 0375-9474. doi: [https://doi.org/10.1016/0375-9474\(93\)90259-Z](https://doi.org/10.1016/0375-9474(93)90259-Z). URL: <https://www.sciencedirect.com/science/article/pii/037594749390259Z>.
- [25] Andrzej Czarnecki, Xavier Garcia i Tormo, and William J. Marciano. "Muon decay in orbit: Spectrum of high-energy electrons". In: *Phys. Rev. D* 84 (1 July 2011), p. 013006. doi: <10.1103/PhysRevD.84.013006>. URL: <https://link.aps.org/doi/10.1103/PhysRevD.84.013006>.
- [26] Andrew Edmonds et al. "Measurement of proton, deuteron, triton, and α particle emission after nuclear muon capture on Al, Si, and Ti with the AlCap experiment". In: *Phys. Rev. C* 105 (3 Mar. 2022), p. 035501. doi: <10.1103/PhysRevC.105.035501>. URL: <https://link.aps.org/doi/10.1103/PhysRevC.105.035501>.
- [27] John F. Donoghue, Eugene Golowich, and Barry R. Holstein. *Dynamics of the Standard Model*. 2nd ed. Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology. Cambridge University Press, 2014. doi: <10.1017/CBO9780511803512>.
- [28] Ben Krikler. "Sensitivity and background estimates for Phase-II of the COMET experiment". PhD thesis. Imperial College London, Sept. 2016. doi: <https://doi.org/10.25560/45365>.
- [29] Ewen Gillies. "COMET Phase-I Track Reconstruction using Machine Learning and Computer Vision". PhD thesis. Imperial College London, Aug. 2018. doi: <https://doi.org/10.25560/68463>.
- [30] Wilhelm H. Bertl et al. "A Search for muon to electron conversion in muonic gold". In: *Eur. Phys. J. C* 47 (2006), pp. 337–346. doi: <10.1140/epjc/s2006-02582-x>.
- [31] T. Suzuki, D. F. Measday, and J. P. Roalsvig. "Total nuclear capture rates for negative muons". In: *Phys. Rev. C* 35 (6 June 1987), pp. 2212–2224. doi: <10.1103/PhysRevC.35.2212>. URL: <https://link.aps.org/doi/10.1103/PhysRevC.35.2212>.
- [32] William R Leo. *Techniques for nuclear and particle physics experiments: a how-to approach*; 2nd ed. Berlin: Springer, 1994. doi: <10.1007/978-3-642-57920-2>. URL: <https://link.springer.com/book/10.1007/978-3-642-57920-2>.

- [33] Rene Brun and Fons Rademakers. "ROOT An object oriented data analysis framework". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 389.1 (1997). New Computing Techniques in Physics Research V, pp. 81–86. ISSN: 0168-9002. doi: [https://doi.org/10.1016/S0168-9002\(97\)00048-X](https://doi.org/10.1016/S0168-9002(97)00048-X). URL: <http://www.sciencedirect.com/science/article/pii/S016890029700048X>.
- [34] David C. Carey, F. Christoph Iselin, and Karl Leslie Brown. "Decay TURTLE (Trace Unlimited Rays Through Lumped Elements): A Computer Program for Simulating Charged Particle Beam Transport Systems, Including Decay Calculations". In: 1974.
- [35] S. Agostinelli et al. "Geant4a simulation toolkit". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. doi: [https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8). URL: <http://www.sciencedirect.com/science/article/pii/S0168900203013688>.
- [36] Alfredo Ferrari et al. "FLUKA: A multi-particle transport code (Program version 2005)". In: (Oct. 2005). doi: <10.2172/877507>.
- [37] Nikolai Mokhov. *MARS15, Version 00*. July 2016. URL: <https://www.osti.gov/biblio/1282121>.
- [38] Tatsuhiko Sato et al. "Features of Particle and Heavy Ion Transport code System (PHITS) version 3.02". In: *Journal of Nuclear Science and Technology* 55.6 (2018), pp. 684–690. doi: <10.1080/00223131.2017.1419890>. eprint: <https://doi.org/10.1080/00223131.2017.1419890>. URL: <https://doi.org/10.1080/00223131.2017.1419890>.
- [39] J. Apostolakis et al. "Geometry and physics of the Geant4 toolkit for high and medium energy applications". In: *Radiation Physics and Chemistry* 78.10 (2009). Workshop on Use of Monte Carlo Techniques for Design and Analysis of Radiation Detectors, pp. 859–873. ISSN: 0969-806X. doi: <https://doi.org/10.1016/j.radphyschem.2009.04.026>. URL: <https://www.sciencedirect.com/science/article/pii/S0969806X09001650>.

- [40] Matthias Dubouchet. *cydet_animations, a set of software tools to produce animated event displays of the Cylindrical Detector*. Version v1.0. 2022. URL: https://gitlab.in2p3.fr/m.dubouchet18/cydet_animations.
- [41] Ian J. Goodfellow et al. “Generative Adversarial Networks”. In: *arXiv:1406.2661 [cs, stat]* (June 10, 2014). arXiv: [1406.2661](https://arxiv.org/abs/1406.2661). URL: <http://arxiv.org/abs/1406.2661> (visited on 01/25/2021).
- [42] Martin Arjovsky, Soumith Chintala, and Léon Bottou. *Wasserstein GAN*. 2017. arXiv: [1701.07875 \[stat.ML\]](https://arxiv.org/abs/1701.07875).
- [43] Ishaan Gulrajani et al. “Improved Training of Wasserstein GANs”. In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon et al. Vol. 30. Curran Associates, Inc., 2017. URL: <https://proceedings.neurips.cc/paper/2017/file/892c3b1c6dccc52936e27cbd0ff683d6-Paper.pdf>.
- [44] Anja Butter, Tilman Plehn, and Ramon Winterhalder. “How to GAN LHC Events”. In: *SciPost Phys.* 7 (6 2019), p. 75. doi: [10.21468/SciPostPhys.7.6.075](https://doi.org/10.21468/SciPostPhys.7.6.075). URL: <https://scipost.org/10.21468/SciPostPhys.7.6.075>.
- [45] Sydney Otten et al. “Event Generation and Statistical Sampling for Physics with Deep Generative Models and a Density Information Buffer”. In: *arXiv:1901.00875 [hep-ex, physics:hep-ph, physics:physics]* (Dec. 4, 2019). arXiv: [1901.00875](https://arxiv.org/abs/1901.00875). URL: <http://arxiv.org/abs/1901.00875> (visited on 01/13/2021).
- [46] Bobak Hashemi et al. *LHC analysis-specific datasets with Generative Adversarial Networks*. 2019. arXiv: [1901.05282 \[hep-ex\]](https://arxiv.org/abs/1901.05282).
- [47] C. Ahdida et al. “Fast simulation of muons produced at the SHiP experiment using Generative Adversarial Networks”. In: *Journal of Instrumentation* 14.11 (Nov. 27, 2019), P11028–P11028. ISSN: 1748-0221. doi: [10.1088/1748-0221/14/11/P11028](https://doi.org/10.1088/1748-0221/14/11/P11028). URL: <https://iopscience.iop.org/article/10.1088/1748-0221/14/11/P11028> (visited on 01/25/2021).
- [48] Luke de Oliveira, Michela Paganini, and Benjamin Nachman. “Learning Particle Physics by Example: Location-Aware Generative Adversarial Networks for Physics Synthesis”. In: *Computing and Software for Big Science* 1.1 (Sept. 2017). ISSN: 2510-2044. doi: [10.1007/s41781-017-0004-6](https://doi.org/10.1007/s41781-017-0004-6). URL: <http://dx.doi.org/10.1007/s41781-017-0004-6>.

- [49] Michela Paganini, Luke de Oliveira, and Benjamin Nachman. "CaloGAN: Simulating 3D High Energy Particle Showers in Multi-Layer Electromagnetic Calorimeters with Generative Adversarial Networks". In: *Physical Review D* 97.1 (Jan. 30, 2018), p. 014021. issn: 2470-0010, 2470-0029. doi: [10.1103/PhysRevD.97.014021](https://doi.org/10.1103/PhysRevD.97.014021). arXiv: [1712.10321](https://arxiv.org/abs/1712.10321). URL: <http://arxiv.org/abs/1712.10321> (visited on 01/13/2021).
- [50] Martin Erdmann et al. "Generating and Refining Particle Detector Simulations Using the Wasserstein Distance in Adversarial Networks". In: *Computing and Software for Big Science* 2.1 (July 2018), p. 4. issn: 2510-2044. doi: [10.1007/s41781-018-0008-x](https://doi.org/10.1007/s41781-018-0008-x). URL: <https://doi.org/10.1007/s41781-018-0008-x>.
- [51] Martin Erdmann, Jonas Glombitza, and Thorben Quast. "Precise Simulation of Electromagnetic Calorimeter Showers Using a Wasserstein Generative Adversarial Network". In: *Computing and Software for Big Science* 3.1 (Jan. 2019), p. 4. issn: 2510-2044. doi: [10.1007/s41781-018-0019-7](https://doi.org/10.1007/s41781-018-0019-7). URL: <https://doi.org/10.1007/s41781-018-0019-7>.
- [52] Kaiming He et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.
- [53] Diederik P. Kingma and Jimmy Ba. "Adam: A Method for Stochastic Optimization". In: *CoRR* abs/1412.6980 (2015).
- [54] S. Kullback and R. A. Leibler. "On Information and Sufficiency". In: *The Annals of Mathematical Statistics* 22.1 (1951), pp. 79–86. doi: [10.1214/aoms/1177729694](https://doi.org/10.1214/aoms/1177729694). URL: <https://doi.org/10.1214/aoms/1177729694>.
- [55] Noam Mouelle. "Evaluation of Generative Adversarial Networks for the COMET experiment". Master's thesis. Imperial College London, May 2022.
- [56] Irene Andreou. "Evaluating AI-generated particle hits for Beyond the Standard Model searches with COMET". Master's thesis. Imperial College London, May 2022.
- [57] Noam Mouelle and Irene Andreou. "Evaluating Generative Adversarial Networks for particle hit generation in a cylindrical drift chamber using Fréchet Inception Distance". In: *21st International Workshop on Advanced Computing and Analysis Techniques in Physics Research*. 2022 (proceedings forthcoming). URL: <https://indico.cern.ch/event/1106990/contributions/5096906/>.

- [58] Tim Salimans et al. "Improved Techniques for Training GANs". In: *arXiv:1606.03498 [cs]* (June 10, 2016). arXiv: 1606.03498. URL: <http://arxiv.org/abs/1606.03498> (visited on 01/12/2021).
- [59] Christian Szegedy et al. "Going deeper with convolutions". In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015, pp. 1–9. doi: [10.1109/CVPR.2015.7298594](https://doi.org/10.1109/CVPR.2015.7298594).
- [60] Martin Heusel et al. "GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium". In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. NIPS'17. Long Beach, California, USA: Curran Associates Inc., 2017, pp. 6629–6640. ISBN: 9781510860964.
- [61] Christian Szegedy et al. "Rethinking the Inception Architecture for Computer Vision". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 2818–2826. doi: [10.1109/CVPR.2016.308](https://doi.org/10.1109/CVPR.2016.308).
- [62] Kevin Kilgour et al. "Fréchet Audio Distance: A Reference-Free Metric for Evaluating Music Enhancement Algorithms". In: *Proc. Interspeech 2019*. 2019, pp. 2350–2354. doi: [10.21437/Interspeech.2019-2219](https://doi.org/10.21437/Interspeech.2019-2219).
- [63] Thomas Unterthiner et al. "Towards Accurate Generative Models of Video: A New Metric & Challenges". In: *ArXiv* abs/1812.01717 (2018).
- [64] Raghav Kansal et al. "Graph Generative Adversarial Networks for Sparse Data Generation in High Energy Physics". In: *arXiv preprint arXiv:2012.00173* (2020).
- [65] Leland McInnes, John Healy, and James Melville. *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*. 2018. doi: [10.48550/ARXIV.1802.03426](https://doi.org/10.48550/ARXIV.1802.03426). URL: <https://arxiv.org/abs/1802.03426>.
- [66] M. H. Kalos. "Monte Carlo Integration of the Adjoint Gamma-Ray Transport Equation". In: *Nuclear Science and Engineering* 33.3 (1968), pp. 284–290. doi: [10.13182/NSE68-A19235](https://doi.org/10.13182/NSE68-A19235). eprint: <https://doi.org/10.13182/NSE68-A19235>. URL: <https://doi.org/10.13182/NSE68-A19235>.
- [67] Bo Eriksson et al. "Monte Carlo Integration of the Adjoint Neutron Transport Equation". In: *Nuclear Science and Engineering* 37.3 (1969), pp. 410–422. doi: [10.13182/NSE69-A19116](https://doi.org/10.13182/NSE69-A19116). eprint: <https://doi.org/10.13182/NSE69-A19116>. URL: <https://doi.org/10.13182/NSE69-A19116>.

- [68] L. Desorgher, F. Lei, and G. Santin. “Implementation of the reverse/adjoint Monte Carlo method into Geant4”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 621.1 (2010), pp. 247–257. issn: 0168-9002. doi: <https://doi.org/10.1016/j.nima.2010.06.001>. URL: <https://www.sciencedirect.com/science/article/pii/S0168900210011836>.
- [69] Valentin Niess et al. “Backward Monte-Carlo applied to muon transport”. en. In: *Computer Physics Communications* 229 (Aug. 2018). arXiv: 1705.05636, pp. 54–67. issn: 00104655. doi: <10.1016/j.cpc.2018.04.001>.
- [70] Valentin Niess. “The PUMAS library”. In: *Computer Physics Communications* 279 (2022), p. 108438. issn: 0010-4655. doi: <https://doi.org/10.1016/j.cpc.2022.108438>. URL: <https://www.sciencedirect.com/science/article/pii/S0010465522001576>.
- [71] Mengyun Guan et al. *A parametrization of the cosmic-ray muon flux at sea-level*. 2015. doi: <10.48550/ARXIV.1509.06176>. URL: <https://arxiv.org/abs/1509.06176>.
- [72] D. Heck et al. CORSIKA: *A Monte Carlo code to simulate extensive air showers*. Tech. rep. 51.02.03; LK 01; Wissenschaftliche Berichte, FZKA-6019 (Februar 98). 1998. doi: <10.5445/IR/270043064>.
- [73] S. Haino et al. “Measurements of primary and atmospheric cosmic-ray spectra with the BESS-TeV spectrometer”. In: *Physics Letters B* 594.1 (2004), pp. 35–46. issn: 0370-2693. doi: <https://doi.org/10.1016/j.physletb.2004.05.019>. URL: <https://www.sciencedirect.com/science/article/pii/S0370269304007567>.
- [74] K. Carstensen et al. “The Sea Level Muon Momentum Spectrum and Charge Ratio at 75°”. In: *International Cosmic Ray Conference*. Vol. 6. International Cosmic Ray Conference. Aug. 1975, p. 2082.
- [75] Moritsu Manabu. *Track direction identification by track fitting*. 35th COMET Collaboration Meeting. Internal presentation. Dec. 2021. URL: <https://kds.kek.jp/event/40211/contributions/201557/>.
- [76] Mehdi Mirza and Simon Osindero. “Conditional Generative Adversarial Nets”. In: *ArXiv* abs/1411.1784 (2014).

- [77] Ashish Vaswani et al. “Attention is All You Need”. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. NIPS’17. Long Beach, California, USA: Curran Associates Inc., 2017, pp. 6000–6010. ISBN: 9781510860964.

List of Figures

1.1	Feynman diagram for the weak decay of the muon. Note how lepton flavour is conserved by the presence of a muon neutrino and an electron anti-neutrino.	3
1.2	Feynman diagrams of processes allowing charged lepton flavour violation in the SM extended with neutrino masses. Although these processes enable CLFV, their branching ratios are heavily suppressed to unobservable levels because of the lightness of neutrinos compared to the weak scale [16].	4
1.3	90%-confidence upper limit on the branching ratio of three charged lepton flavour-violating processes over time. The target material N is indicated for $\mu\text{-}e$ conversion experiments. Past experiment results were tabulated by Bernstein and Cooper [16]. Future data points are the expected sensitivities quoted in the MEG II [18], Mu3e [19], COMET Phase-I [20] and Mu2e [21] design reports.	6
1.4	Energy spectrum of decay-in-orbit electrons produced by muonic aluminium atoms, calculated by Czarnecki et al. [25]. The logarithmic scale reveals the shape of the high-energy tail up to the endpoint energy of 104.97 MeV.	8
1.5	Momentum distribution of reconstructed signal and decay-in-orbit electrons in COMET Phase-I, assuming a conversion branching ratio of 3×10^{-15} . A momentum resolution of 200 keV/c is assumed, and detector efficiency factors are extracted from the technical design report [20]. The vertical dashed line shows the lower bound of the momentum window at 103.6 MeV/c, which is used to discriminate background events from the conversion signal.	9

1.6	New vertices which arise in the effective Lagrangian of Equation 1.7 and allow CLFV. The photonic interaction can also mediate $\mu-e$ conversion by making the photon interact with an external quark.	10
2.1	Schematic top-down view of the COMET experiment. The beam pipe is represented in grey, and the light red rectangles along the beamline represent the solenoids that generate the magnetic field. Curved solenoids additionally help to select charge and momentum, as discussed in Section 2.3.	14
2.2	Timeline of a COMET event. Shortly after the proton collision, the beam flash floods the detector and muons begin to be bound in the muon stopping target. The Phase-I trigger window shown here starts at $t = 700\text{ ns}$ and lasts until the next proton bunch collision. Electron tracks with a momentum around $105\text{ MeV}/c$ observed within this timing window are considered as signal candidates.	15
2.3	Cutaway view of the pion-production section. The orange arrow indicates the path of the proton beam while the teal arrow shows the direction of backward-going pions captured by the magnetic field.	16
2.4	Momentum distribution of pions produced by simulating proton collisions with Geant4, depending on whether their initial direction is forward or backward with respect to the proton beam direction.	17
2.5	Cutaway view of the Phase-I transport solenoid. The curved solenoid (in red) combined with collimators (in teal) select particles depending on their charge and momentum.	17
2.6	Effective drift of particles of various momenta and either charge as they progress along the Phase-I transport solenoid, calculated using Equation 2.2. The pitch angle is assumed to be 45° and the bands show the effect of a $\pm 10^\circ$ difference. As shown on the right, a 0.05 T dipole field allows negative particles around $45\text{ MeV}/c$ to stay on axis while higher- and lower-momentum particles drift in opposite directions and positive particles all drift one way. By use of collimators at the end of the transport section, only particles of the right momentum and charge are allowed through.	18

2.7	Cutaway view of the StrECAL detector. The Straw Tracker stations are shown in blue and the ECAL in yellow.	20
2.8	Cutaway view of the Cylindrical Detector, composed of the Cylindrical Drift Chamber (in teal) and Cylindrical Trigger Hodoscope (in yellow). The muon stopping target disks, in purple, sit in the centre of the detector system.	21
2.9	Conversion electron trajectory as observed by the Cylindrical Detector. The effect of stereo angles is visible at the start of the trajectory where hits have alternating azimuthal positions around the actual electron track. Upon reaching the CTH, the electron produces consecutive hits in four adjacent counters, satisfying the fourfold coincidence trigger criterion. .	22
3.1	Data flow in the ICEDUST framework. The colour of arrows represents the data format. By design, simulated detector data and real data share a common format such that they can be processed identically by the calibration, reconstruction and analysis stages.	28
3.2	Layout of <code>oaEvent</code> files which contain MC or real data. Blue arrows in Figure 3.1 indicate steps of the data flow where data is stored in this format on disk.	28
3.3	Cutaway views of the simulation geometries implemented in <code>SimG4</code> and visualised with <code>DisplayCore</code> . The experiment hall is also modelled in the simulation but was hidden here for clarity.	31
3.4	Hits, shown as red dots, produced by the CDC sensitive detector class when a particle deposits charge inside the gas. One instance of <code>IG4HitGas</code> is created every time the particle traverses a CDC cell, i.e. roughly every 16 mm.	34
3.5	Top-down cutaway view of the running configuration for MC5, showing 1.6×10^5 overlaid events (1% of a bunch). The orange line shows the sampling boundary where particles that cross into the detector region via the beamline or wall are recorded. Only charged-particle trajectories are shown for clarity.	36

3.6	Still frames of an animation rendered by the visualisation tool. The event shown outlines how a conversion electron would be seen by the CyDet system among background hits.	37
4.1	Schematic diagram of the data flow in the GAN training procedure. Real samples from the training dataset and fake samples from the generator are iteratively evaluated by the discriminator to compute a score. The loss for both networks is estimated by comparing the score to the sample's associated truth label (real or fake). By <i>backpropagation</i> , the loss is then numerically differentiated with respect to the network weights. The weights are then adjusted in the direction of decreasing loss.	42
4.2	Structure of one training sample. Hits are arranged into fixed-length sequences in the GEANT4 order, i.e. whereby hits from the same track are contiguous.	48
4.3	Comparison of hit patterns in the CDC depending on whether or not a reconstructible track occurred. The area of each hit is proportional to the amount of energy deposited, and the colour denotes particle type.	49
4.4	Distributions of the three continuous features used to train the GAN.	50
4.5	Temporal (1D) convolution operation with kernel size 3. For clarity, both the input and output sequences have one channel, i.e. $C_{in} = C_{out} = 1$. The blue arrow shows the direction along which the kernel slides across the input sequence.	51
4.6	Residual blocks used in both networks to facilitate gradient flow. Blue arrows indicate an activation function.	52
4.7	Network architectures. The layer types are: FC for fully-connected, Conv for a standard convolution and ConvT for a transposed (fractionally-strided) convolution. In convolutional layers, kernel size and stride are denoted by k and s , respectively.	53
4.8	Loss curves as a function of training epochs.	56
4.9	Comparison of individual samples, comprised of 512 hits. Hits are represented by a circle whose area is proportional to the energy deposit and whose colour shows distance of closest approach. Lines connect consecutive hits in the sequence.	57

4.10 Comparison of feature distributions and correlations when generating a dataset of the same size as the training dataset.	58
4.11 Kullback-Leibler divergence between the distributions of real and generated hits as a function of training epochs.	58
4.12 Comparison of quantities not explicitly learned by the GAN model. . .	59
4.13 Fréchet Inception Distance computed by the pre-trained Inception-v3 network over training epochs of the CDC GAN. Since Inception requires a 2D image as input and our hit sequences have a third temporal dimension, the hit data is projected along its axes and a separate FID is calculated for each projection.	61
4.14 Fréchet Inception Distance computed over training epochs of the CDC GAN by two external neural networks. Unlike in Figure 4.13, hit sequences can be processed at once without projecting them along an axis. The 3D CNN is trained on Monte Carlo hit data, in the same way as the front and back layers added to Inception-v3 for transfer learning.	62
4.15 UMAP computed from feature vectors extracted by the TL Inception-v3 network, for reconstructible, noise-like and generated hit sequences. The number of neighbours in the UMAP transform was set to 50.	63
5.1 Comparison of the forward and backward (or adjoint) configurations of Monte Carlo simulation, in the case where a small detector volume is surrounded by a comparatively large particle source. Reversing the flow of particle transport means that sampled events are more likely to contribute toward the rate estimation.	66
5.2 COMET Phase-I simulation geometry used in the cosmic background study.	68
5.3 Spectrum of the atmospheric muon flux (multiplied by $E_\mu^{2.7}$) calculated at an altitude of 1600m using a CORSIKA simulation. This spectrum, tabulated in both energy and zenith angle, is used as the flux source in the COMET Phase-I backward MC. Here, it is compared with the parametrisation of the muon flux at sea level from Guan et al. [71].	70

5.4	Validations of the backward MC-estimated flux with experimental observations of the atmospheric muon flux by two independent experiments. Events from the backward MC simulation are selected to reflect the angular acceptance of the corresponding experiment.	71
5.5	Average rate of inbound atmospheric muons over each face of the CRV as a function of momentum. Figures in parentheses in the legend are the integrated rate averaged over each face.	72
5.6	μ - e conversion background rates from atmospheric muons estimated via backward MC simulation. The error bars show the statistical uncertainty on the sampled flux over backward sampling trials. In the presence of the CRV, the background rate is dominated by events where an atmospheric muon sneaks through the openings.	74
6.1	Initial position distribution of signal electrons before and after weighting them by the likelihood of a muon being bound in each bin.	77
6.2	Momentum spectrum of electrons in the DIO sample.	79
6.3	Potential DIO-induced background with momentum $p = 103.7 \text{ MeV}/c$. .	79
6.4	Sneaking atmospheric μ^- background	80
6.5	Momentum spectrum of conversion, DIO and atmospheric events around the conversion energy	84
6.6	Sneaking atmospheric μ^+ background event	85
6.7	Signal and background event counts versus reconstructed track momentum, over $T_{\text{DAQ}} = 146$ days and assuming $\mathcal{B}_{\text{conversion}} = 3.61 \times 10^{-15}$. . .	87
6.8	Integrated atmospheric background count versus CRV inefficiency . .	88
A.1	Software packages that compose the original backward MC simulation stack for the Phase-I study, and organisation of their integrated ICEDUST counterparts.	102

List of Tables

2.1	Expected number of background events from each potential source [20]. The total count is much smaller than one, such that the observation of a single signal electron could be evidence that charged lepton flavour is violated.	24
2.2	Efficiency factors used to determine the signal acceptance in the COMET Phase-I TDR [20]. The net acceptance is the product of all efficiency factors. “Cuts” refers to track quality cuts, used to reject events with irregular tracks that cannot be accurately fitted and reconstructed.	26
4.1	Physical features which make up each hit in the training dataset.	47
4.2	Training hyperparameters.	55
6.1	Summary of atmospheric background counts for $T_{\text{DAQ}} = 146$ days under different scenarios for particle identification (PID) and CRV efficiency (ϵ_{CRV}).	88