UNIVERSITÀ
DEGLI STUDI
FIRENZE

DOTTORATO DI RICERCA IN
Fisica e Astronomia

CICLO XXXIV

COORDINATORE Prof. Raffaello D'Alessandro

Fixed-target physics with the LHCb experiment at CERN

Settore Scientifico Disciplinare FIS/04

Dottorando
Mariani Saverio

Tutore
Dott. Graziani Giacomo

Tutore
Prof. Adriani Oscar

Coordinatore
Prof. D'Alessandro Raffaello

Anni 2018/2021

EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH (CERN)



December 17, 2021

Fixed-target physics with the LHCb experiment at CERN

Author: Saverio Mariani

Supervisors: Giacomo Graziani
Oscar Adriani

INFN and Università degli Studi di Firenze

Abstract: Owing to its detector geometry covering the forward direction and its excellent track reconstruction and particle identification performance, the LHCb experiment at CERN started in 2015 a pioneering study of beam-gas collisions by injecting noble gases in the LHC accelerator. In this thesis, my contributions to this physics programme, addressing the demand for extending the physics reach of the LHC complex stressed in the 2020 European Strategy for Particle Physics Update, are presented. Using a sample of proton-helium collisions collected in 2016, I performed the measurement of the non-prompt antiproton production at $\sqrt{s_{\rm NN}} = 110$ GeV, complementing a previous analysis only considering the prompt component. The results provide a key ingredient to a more accurate description of the antiproton production in collisions between primary cosmic rays and the interstellar medium, representing the dominant limitation for the interpretation of data on the antiproton flux in space. This is being measured by experiments, notably the AMS-02 spectrometer, searching for an excess over the expected flux that could provide an indirect evidence for a hypothetical Dark Matter particle annihilation or decay process. Being the particle identification performance in fixed-target data limited by the size of calibration samples, I conceived and developed a machine-learning and data-based approach to its parametrization and demonstrated that it achieves a better performance with respect to models based on the LHCb detailed simulation. An upgrade of the fixed-target device is ongoing, mainly consisting in the installation of a 20-cm-long cell confining the gas target in a region upstream the nominal LHCb interaction point. During my *Ph.D.* programme, I have been responsible for the software activities aiming to allow beam-beam and beam-gas data to be simultaneously acquired. I tested and adapted the full event reconstruction sequence on simulated beam-gas data and implemented a selection strategy within the challenging real-time framework being deployed for the LHCb upgrade. In view of the significant increase of the luminosity and of the wider choice of gas species offered by the target upgrade, this work is paving the way to the implementation of a unique full-scale fixed-target experiment at the LHC.

© 2021 CERN for the benefit of the LHCb collaboration. CC BY 4.0 licence.

A tutti i luoghi e le persone
che sono per me casa, ovunque essi siano.

Preface

Starting from 2010, when the first proton-proton collision was delivered, the data collected at the LHC accelerator at CERN have been exploited in a rich physics programme, ranging from the discovery of the Higgs Boson jointly announced in 2012 by the CMS and ATLAS collaborations, to the LHCb precision measurements of flavour physics observables and to the ALICE studies of the quark-gluon plasma. All these results are aimed to test with increasing precision the predictions of the Standard Model of Particle Physics, the most complete description of the fundamental particles and interactions available nowadays. Observations of phenomena beyond the Standard Model might motivate its extension and shed light on the many questions still unanswered. Among these, the evidence that a significant fraction of the mass in the Universe, called dark matter, is not explained by visible matter has been reported by many cosmological and astrophysical observations. Its possible nature in terms of hypothetical particles not contemplated in the Standard Model is being investigated at accelerator facilities, where these might be directly produced in the collisions, and by experiments in space. Indeed, a possible dark matter annihilation or decay process could manifest as an excess of the antimatter abundance in space with respect to the expected production in collisions between cosmic rays and the interstellar medium, mostly composed of hydrogen and helium. The interpretation of the data collected in space are thus relying on the precision of theoretical models for the description of these known sources. In 2015, when the AMS-02 experiment published its first results for the antiproton abundance in cosmic rays, the lack of experimental data on the antiproton production was the dominant systematic uncertainty and limited the conclusions that could be drawn by comparing the results with the predicted secondary antiproton flux.

In 2011, the LHCb experiment at CERN was equipped with a System for Measuring Overlap with Gas (SMOG), intended to reduce the systematic uncertainty on the production cross-section of heavy hadrons related to the measurement of the LHC luminosity. Thanks to the flexibility of its detector and operations and to the possibility to inject noble gases into the LHC accelerator beam-pipe with SMOG, the LHCb experiment started in 2015 a pioneering programme exploiting collisions between the LHC beams and gaseous fixed targets. The collision energy range, intermediate between lower-energy experiments operating for example at the SpS accelerator and those at colliders such as the LHC or RHIC, is poorly explored by previous experimental measurements. The fixed-target programme at the LHCb experiment is an opportunity for different fields, including heavy-ion physics, nucleon structure characterization and astrophysics. As an example, by injecting helium in the LHC beam-pipe, the antiproton production cross-section in pHe collisions at $\sqrt{s_{\mathrm{NN}}} = 110\,\mathrm{GeV}$ was determined for the first time and contributed to improve the precision of models for the antimatter production in space. Nevertheless, SMOG was not originally conceived for production measurements and the programme suffered from some operational limitations. In particular, the integrated luminosity for the beam-gas mode could not be directly measured and, giving the overlap between the beam-gas and beam-beam interaction regions, the data-taking was mostly carried out in short and dedicated runs of the LHC. As a consequence, large statistical uncertainties affected the measurements both directly and indirectly via the uncertainty on the calibration samples.

With the end of the LHC Run2 period of data-taking, in 2018, LHCb started a general upgrade of the detector and of the data acquisition strategy, moving *de facto* to a brand-new experiment. Within this scenario, also the fixed-target system is being upgraded to SMOG2, whose core idea consists of the installation of a confinement cell for the gas upstream the nominal pp collision point. With a well-defined 20 cm-long region for the beam-gas collisions, the areal density of the gas target could be increased by up to two orders of magnitude with the same gas flow as SMOG and also non-noble gases like hydrogen, deuterium or oxygen could be exploited. Moreover, the separation between the beam-beam and beam-gas collision regions will open the possibility to simultaneously acquire data for the two collision systems, operating LHCb at the same time as a fixed-target and a collider experiment. These improvements, together with the upgrade of the gas injection system, will further enrich the fixed-target physics programme providing unprecedented statistics to study production of heavy charm and beauty hadron states, the Drell-Yan process, rare central-exclusive channels and other probes for the internal structure of nucleons. The programme of cosmic ray interest will also be extended.

The SMOG project and its upgrade are included in a wider scenario aimed to the extension of the LHC physics reach, discussed within the Physics Beyond Collider forum [1] and recognised as a key activity in the European strategy for particle physics update [2, 3]:

> The multi-TeV LHC proton- and ion-beams allow for the most energetic fixed-target (FT) experiments ever performed, opening the way for unique studies of the nucleon and nuclear matter at high x, of the spin content of the nucleon and of the nuclear matter phases [...]. Such LHC-FT studies have strong connections to high-energy neutrino and cosmic-rays physics. **The physics reach of the LHC complex can greatly be extended at a limited cost [...]. The efforts of the existing LHC experiments to implement such a programme, including specific R&D actions on the collider, deserve support.**

In this thesis, my contributions to the Run2 LHCb fixed-target programme and to the preparation to the data-taking with SMOG2 are discussed in five parts.
In the first one, the theoretical framework of this work is introduced, covering in Chapter I.1 and I.2 some elements of the Standard Model of Particle Physics and the experimental searches for dark matter through the measurements of the antimatter abundance in space, respectively. Chapter I.3 and I.4 introduce the experimental context, describing the LHCb experiment and the Run2 SMOG programme, respectively. The second and third parts deal with my main contributions to the Run2 SMOG analysis programme and operations. Chapters II.1 and II.2 present the analysis I performed during my *Ph.D.* programme for the non-prompt (detached) antiproton production in pHe collisions, *i.e.* those originating in decays of strange baryons. This production mechanism is expected to account for 25-40% of the prompt antiproton one, but it is poorly constrained by previous experimental data. Two different approaches are followed for this measurement and described in the two chapters: in the former, all antiprotons produced in pHe collisions are reconstructed and selected and then distinguished between prompt and detached according to their consistency to the pHe collision vertex; in the latter, antiprotons from the $\overline{\Lambda} \to \overline{p}\pi^+$ decay with a promptly-produced $\overline{\Lambda}$ particle are exclusively measured. The results from the two approaches are then compared in Chapter II.3, drawing the analysis conclusions.
Another important contribution I gave to the Run2 SMOG programme, discussed in

Part III, is aimed to improve the particle identification performance for fixed-target analyses, suffering from the large statistical uncertainty on the data-driven calibration. With the exploitation of machine-learning techniques, the classifiers that are defined at LHCb to identify the particles are described as a Gaussian Mixture Model and the dependence of all its parameters on the relevant observables is learned through a set of Neural Networks. The particle identification classifier can then be predicted on smaller-size samples, according to their distribution of the same observables. A proof-of-principle physics-case and other applications on fixed-target and *pp* samples are presented. Several ongoing analyses, including that described in the Part II of this thesis, are profiting from the introduction of this approach.
The fourth part deals with the preparation to the data-taking with SMOG2, that will operate from 2022. Chapter IV.1 describes the upgrades that the LHCb detectors and data acquisition strategy are currently facing. Chapter IV.2 is focused on the SMOG2 project, scrutinises the main characteristics of the installed cell and the opportunities for physics research. During my *Ph.D.* programme, I implemented the selection algorithms to acquire data with SMOG2, extending the LHCb fixed-target programme to the LHC Run3. As documented in Chapter IV.3, I tested and adapted the reconstruction algorithms to the fixed-target collisions upstream the LHCb nominal interaction point within the real-time reconstruction framework that will operate with the upgraded LHCb detector. Exploiting the separation between the beam-beam and beam-gas interaction regions, I studied the feasibility to simultaneously operate LHCb with the two collision systems.
The conclusions of this work are drawn in part V.

Contents

Preface **iv**

Contents **vii**

I *Introduction* **1**

I.1 Elements of theory **2**
I.1.1 The Standard Model of Particle Physics . . . 2
I.1.2 Quantum chromodynamics . . . 4
I.1.3 The Dark Matter . . . 7

I.2 Cosmic antiprotons **10**
I.2.1 Introduction to Cosmic Rays . . . 10
I.2.2 Measurements of the antiproton flux . . . 15

I.3 The LHCb experiment **27**
I.3.1 The CERN facility . . . 27
I.3.1.1 The LHC accelerator . . . 28
I.3.2 The LHCb detector . . . 31
I.3.2.1 The LHCb tracking system . . . 32
I.3.2.1.1 The VELO detector . . . 33
I.3.2.1.2 The magnet and the tracking stations . . . 36
I.3.2.2 The LHCb particle identification system . . . 40
I.3.2.2.1 The RICH system . . . 40
I.3.2.2.2 The calorimetric system . . . 42
I.3.2.2.3 The muon system . . . 44
I.3.3 The LHCb data trigger and processing . . . 45

I.4 The SMOG system **49**
I.4.1 The luminosity measurement with SMOG . . . 49
I.4.1.1 Direct luminosity measurements . . . 49
I.4.1.2 The gas injection system . . . 50
I.4.2 The SMOG gaseous target . . . 51
I.4.2.1 The SMOG physics opportunities . . . 53
I.4.2.2 The prompt antiproton production measurement . . . 55

II *The non-prompt antiproton production measurement* 59
II.0.1 Analysis motivation . . . 60

II.1 Inclusive approach 62
II.1.1 Analysis strategy . . . 62
II.1.2 Simulated samples . . . 63
II.1.3 Antiproton selection . . . 65
II.1.4 Template fit . . . 68
II.1.5 Experimental efficiency . . . 78
II.1.5.1 Simulation efficiency . . . 83
II.1.5.2 Reconstruction and selection efficiencies . . . 89
II.1.6 Systematic uncertainties . . . 92
II.1.6.1 Prompt antiproton $\chi^2_{\rm IP}$ template . . . 92
II.1.6.1.1 High-$\chi^2_{\rm IP}$ prompt antiprotons . . . 92
II.1.6.1.2 Template data-driven cross-check with $\overline{\Lambda}(1520) \to \overline{p}K^+$ 94
II.1.6.2 Simulated $\overline{\Lambda}$-to-$\overline{\Sigma}^-$ ratio . . . 101
II.1.6.3 Simulated $\overline{\Xi}^+$-to-$\overline{\Lambda}$ ratio . . . 105
II.1.6.4 Particle identification . . . 110
II.1.6.5 Tracking . . . 117
II.1.7 Results for the inclusive approach . . . 121

II.2 Exclusive approach 127
II.2.1 Analysis strategy . . . 127
II.2.2 Simulated samples . . . 128
II.2.3 $\overline{\Lambda}_{\rm prompt} \to \overline{p}\pi^+$ decays selection . . . 129
II.2.4 Signal extraction . . . 136
II.2.5 Experimental efficiency . . . 143
II.2.5.1 Generation efficiency . . . 148
II.2.5.2 Simulation efficiency . . . 148
II.2.5.3 Reconstruction and selection efficiencies . . . 153
II.2.5.4 Total efficiency . . . 155
II.2.6 Systematic uncertainties . . . 157
II.2.6.1 Data-driven VELO track reconstruction efficiency . . . 159
II.2.6.2 Fit procedure . . . 162
II.2.6.3 Interactions with the detector material . . . 163
II.2.7 Results for the exclusive approach . . . 167

II.3 Analysis conclusions 173

III *A machine-learning approach to particle identification* 177
III.1.1 Limitations of SMOG particle identification . . . 178
III.1.2 Machine-learning approach . . . 179
III.1.3 Model application on SMOG data . . . 180
III.1.3.1 Selection of the calibration channels . . . 181
III.1.3.2 Feature selection . . . 186
III.1.3.3 Model preprocessing and training . . . 188

III.1.3.4 Model validation . . . 189
III.1.4 Applications of the approach . . . 190
III.1.4.1 PID model for pHe and pAr data . . . 191
III.1.4.2 PID for prompt and detached antiprotons in pHe . . . 195
III.1.4.3 Understanding kaon calibration samples in pp data . . . 199

IV *Preparation to Run3 LHCb fixed-target physics* **207**

IV.1 The LHCb upgrade **208**
IV.1.1 LHCb upgrade overview . . . 208
IV.1.2 The LHCb upgrade detectors . . . 209
IV.1.2.1 The tracking system . . . 209
IV.1.2.2 The PID system . . . 214
IV.1.3 The LHCb upgrade data acquisition strategy . . . 215

IV.2 The SMOG2 project **219**
IV.2.1 Motivations for the upgrade . . . 219
IV.2.2 The SMOG2 cell and the gas feed system . . . 220
IV.2.3 The SMOG2 physics opportunities . . . 224

IV.3 The data processing **227**
IV.3.1 The beam-gas data-taking strategy . . . 227
IV.3.2 The event reconstruction sequence . . . 230
IV.3.2.1 The VELO track reconstruction . . . 232
IV.3.2.2 The primary vertex reconstruction . . . 239
IV.3.2.3 The long track reconstruction . . . 248
IV.3.2.4 The HLT1 reconstruction sequence timing . . . 256
IV.3.3 The SMOG2 selection algorithms . . . 258
IV.3.3.0.1 Minimum-bias selection . . . 262
IV.3.3.0.2 Multi-body selections . . . 262

V *Conclusions* **264**

Acknowledgements **268**

Acronyms **270**

References **274**

Part I

Introduction

A unified description of most of the observable phenomena is nowadays included into the Standard Model of Particle Physics, embracing electromagnetic, weak and strong interactions. Unanswered questions still remain, notably the evidence of a mass in the Universe called dark matter only gravitationally interacting with baryonic particles. To investigate the nature of this mass, still unknown, experiments are measuring the quantity of antimatter in space with an increasing precision and extending the energy range. An excess over the expected abundance from collisions between cosmic rays and the interstellar medium could evidence a possible dark matter decay or annihilation process towards a particle-antiparticle final state. An accurate modelling of the secondary antimatter production is thus needed.
When the AMS-02 experiment published its results for the antiproton flux in cosmic rays, the scarce available data for the antiproton production cross-sections limited the precision of the predictions and no conclusion could be drawn. The LHCb experiment, one of the main four operating at the LHC accelerator at CERN, contributed to reduce this uncertainty with the first determination of the antiproton production in pHe collisions at a nucleon-nucleon centre-of-mass $\sqrt{s_{\mathrm{NN}}} = 110$ GeV energy. This is performed exploiting the injection of helium in the LHC beam-pipe through a system called SMOG, the LHCb detector geometry, instrumenting small angles with respect to the beam axis, and its excellent tracking and particle identification performance. The results provided a first example of the capabilities to exploit LHCb studies of proton-gas collisions, providing a unique fixed-target facility at the highest available energy.

Chapter I.1 and I.2 introduce some elements of the Standard Model of Particle Physics and describe the experimental searches for antiprotons in space, respectively. Chapter I.3 and I.4 are focused on the experimental setup. The former describes the LHCb experiment and the main aspects of its detectors and data processing sequence; the latter the SMOG gas injection system, its operation and physics reach during the LHC Run2 period of data-taking, from 2015 to 2018, and particularly focuses on the antiproton production measurement in pHe collisions.

I.1 Elements of theory

In this chapter, a brief introduction to the theoretical context of the research activities documented in this thesis is presented. Section I.1.1 reminds the basic ingredients of the Standard Model of Particle Physics; Section I.1.2 deals with Quantum ChromoDynamics (QCD), describing the strong nuclear forces binding the quarks into hadrons. Section I.1.3 lists some of the evidences for the Dark Matter and introduces some of the experimental strategies for its study.

I.1.1 The Standard Model of Particle Physics

The need to group and organise in a well-defined and coherent set of laws the huge variety of phenomena we observe in the Universe has always motivated scientists in their research. In modern physics, starting from the unification of all electric and magnetic effects in Maxwell's laws, the relativity and the quantum mechanics paved the way for the development of Quantum ElectroDynamics (QED), embracing all electromagnetic phenomena. The interaction ruling nuclear beta decays, called weak, was then included, giving birth to the electroweak theory. Another fundamental interaction, responsible for the binding of nucleons in nuclei and of quarks in nucleons, called strong, is described by Quantum ChromoDynamics (QCD). All of these are included in the Standard Model of Particle Physics (SM), the most complete quantum field theory available nowadays to describe the fundamental particles and their interactions. Only the definition of the gravitational force as a quantum field theory within the SM is still missing. With respect to the other interactions, it can usually be neglected for particle physics.

The SM defines three possible particle fields with:

- spin 1, carrying the fundamental interactions: the photon (γ) for the electromagnetic, the $W^{\pm}$ and Z bosons for the weak and the gluons for the strong one;
- spin $\frac{1}{2}$, the fundamental constituents of matter. These are divided between leptons, interacting through the gravitational, weak and, when charged, the electromagnetic forces and quarks, also subjected to the strong one;
- spin 0, the Higgs boson, produced by an excitation of the Brout-Hengler-Higgs field, called after the names of the theoreticians that formulated it, and recently

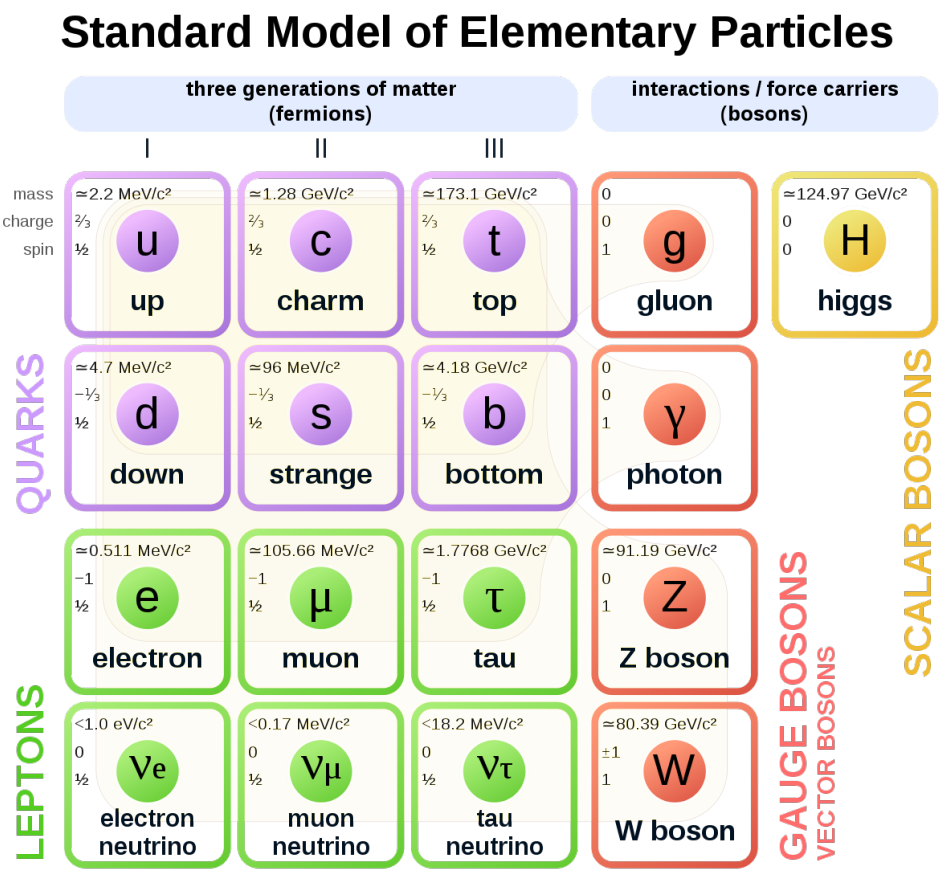


Figure I.1.1: Scheme for the fundamental particles described by the Standard Model of Particle Physics. For all particles, the mass, the electric charge and the spin are indicated.

discovered at CERN [4, 5]. The interaction of quarks, leptons and spin-1 bosons with the Higgs particle is proportional to their mass.

A scheme of all the particles described by the SM is presented in Fig. I.1.1, with the mass, the electric charge and the spin indicated. For each particle, a respective anti-particle exists with all additive quantum numbers, such as the electric charge, of opposite sign. By convention, antiparticles of quarks, neutral leptons and some composite particles are indicated with a bar over the particle symbol, like for the antiproton, denoted with $\overline{p}$. Leptons are organised in three families with a negatively-charged and a neutral particle and are distinguished by a quantum number called lepton flavour. Among the charged leptons, the lightest one is the electron (e^-) and it is stable. The other two, the muon (μ^-) and the tauon (τ^-), present the same electric charge as the electron but a higher mass and decay towards lighter states. For each lepton, a respective neutrino exists: ν_e, ν_μ, ν_τ. The SM describes all neutrinos to be massless, but evidences for their oscillations in flavour with time have proved that at least two of them must have a non-null mass. Also quarks are distinguished in three families by their hadronic flavour and all contain one $+\frac{2}{3}\,e$ and one $-\frac{1}{3}\,e$ charged particles. The quarks of the first family, up (u) and down (d), determine the quantum numbers of protons (with a uud combination) and neutrons (with a udd combination), constituting the ordinary matter. Quark-antiquark pairs of all

families are virtually formed in the nucleons by gluons exchange. The two other families include four flavours: strange (s), charm (c), bottom (b) and top (t), in increasing mass order. Composite particles can be grouped into two categories:

- mesons, composed of a quark-antiquark pair and all unstable;
- (anti)baryons, composed of three (anti)quarks. Among these, the proton and its antiparticle only have never been found to decay.

The two categories are distinguished by the baryon number, an additive quantum number defined as $B = 1/3(n_q - n_{\bar{q}})$, being $n_q(n_{\bar{q}})$ the number of constituent (anti)quarks. Mesons have a zero baryon number, (anti)baryons ∓ 1. Recently, states foreseen by the theory from the combination of two quarks and two antiquarks or a quark-antiquark pair and three (anti)quarks called tetraquark and pentaquarks, respectively, have been observed, mostly by the LHCb collaboration at CERN.

I.1.2 Quantum chromodynamics

Quantum ChromoDynamics (QCD) is the theory describing the strong interactions among hadrons and, as for QED, it is a gauge field theory, *i.e.* it descends from the assumption that there is a symmetry that leaves the theory unchanged. For QCD, this is postulated to be exact and is described by a SU(3) group, meaning that the respective quantum number, called colour, can assume one of three values conventionally set to red (r), green (g) and blue (b). Each quark carries a colour charge, antiquarks present anti-colours ($\bar{r}$, $\bar{g}$ and $\bar{b}$) and gluons carry a colour-anticolour charge pair. Only colour-neutral combinations (colour singlets) can be observed. Quarks in mesons present thus the same colour-anticolour combination, while baryons (anti-baryons) can only be formed by quarks (antiquarks) with all different (anti-)colours.
Since, differently than photons, gluons carry a colour charge, the QED and QCD theories are characterised by an opposite evolution of the interaction intensity with the transferred energy. According to the former, virtual e^+e^- pairs surrounding fermions have a screening effect on their charge and, with the decreasing of the length scale or, equivalently, the increasing of the transferred squared four-momentum Q^2, the interaction strength increases. On the contrary, gluon interactions with the virtual particles surrounding quarks, and notably the gluon-gluon ones, have a net result to induce an antiscreening effect and, as a consequence, a weaker scale at high Q^2. In the limits of an infinite Q^2, the asymptotic freedom, *i.e.* a null intensity of the strong interaction, occurs. In this regime, quarks are expected to behave as free particles and QCD can be treated perturbatively. On the other side, the separation of two quarks require an infinite amount of energy and, when a quark is produced for example in a high-energy proton-proton collisions, composite particles appearing as jets of hadrons are formed. The opposite energy dependence of the electromagnetic and strong interactions, shown in Fig. I.1.2, is described as a non-constant value of the coupling constants:

$$\alpha_{QED}(Q^2) \simeq \alpha \cdot \left[1 - \frac{\alpha}{3\pi} ln(\frac{Q^2}{m_e^2})\right]^{-1} \qquad Q^2 \gg m_e^2 \tag{I.1.1}$$

$$\alpha_{QCD}(Q^2) \simeq \alpha_{QCD}({Q^2}_0) \cdot \left[1 + \frac{33 - 2n_f}{12\pi} \alpha_{QCD}({Q^2}_0) ln(\frac{Q^2}{{Q^2}_0})\right]^{-1}$$

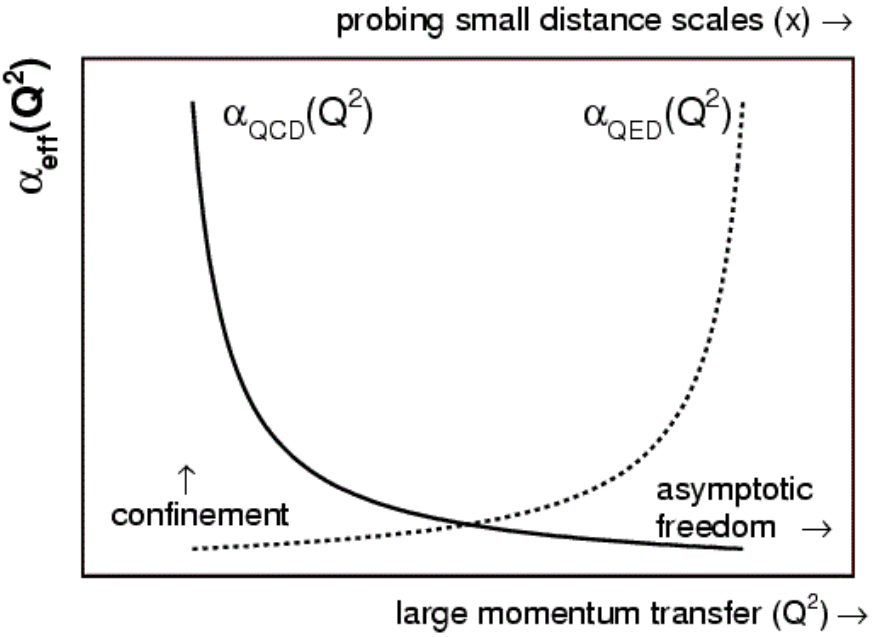


Figure I.1.2: Evolution with Q^2 of the QED and QCD coupling constants. While the former increases at larger values, the latter decreases as a consequence of the exact colour symmetry and of the antiscreening effect.

being:

- $\alpha \simeq 1/137$ the fine structure constant;
- m_e the electron rest mass;
- n_f the number of active quark flavours at the Q^2 energy scale;
- ${Q^2}_0$ a reference energy scale.

Conventionally, the dependence on the reference energy scale is hidden into the $\Lambda_{\rm QCD}$ constant, determining the minimum scale where a perturbative description of QCD is possible:

$$\Lambda^2_{QCD} = {Q^2}_0 \cdot exp\Big[\frac{12\pi}{2n_f - 33}\frac{1}{\alpha_{QCD}({Q^2}_0)}\Big]. \tag{I.1.2}$$

By substituting in Eq. I.1.2:

$$\alpha_{QCD}(Q^2) = \Big[\frac{33 - 2n_f}{12\pi} ln(\frac{Q^2}{\Lambda^2_{QCD}})\Big]^{-1}. \tag{I.1.3}$$

The behaviour of quarks in the nucleon system was extensively studied in lepton-nucleon deep inelastic scatterings and described by Feynman in the parton model. According to it, each nucleon is composed of partons, carrying a fraction of the total momentum called Bjorken-x. A scattering between two particles can be considered as the sum of the single interactions among the partons, assumed to be free in the short interaction time. To describe a lepton-nucleon scattering, the variable

$$x = \frac{Q^2}{2Mc^2\nu}, \tag{I.1.4}$$

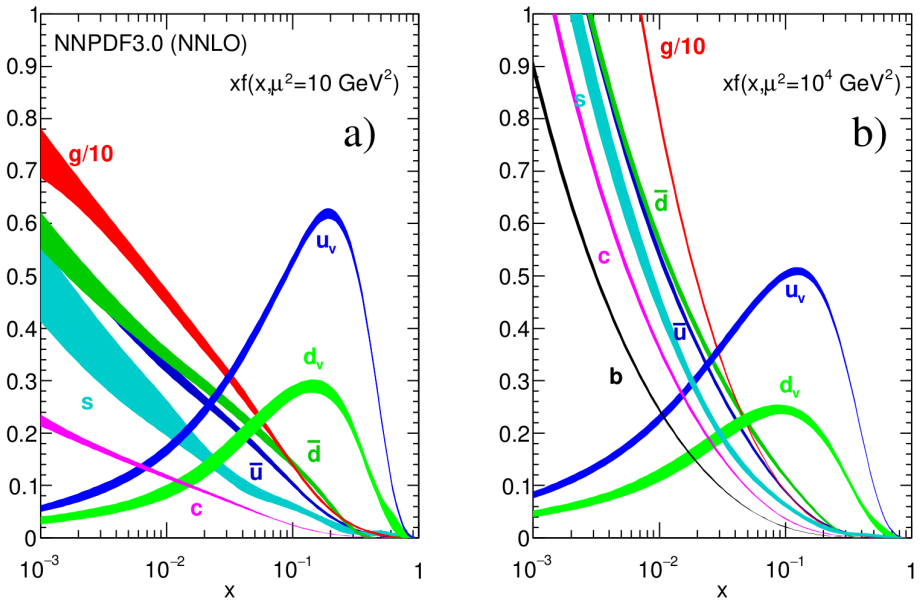


Figure I.1.3: Unpolarised parton distribution functions multiplied by x at the energy scales $Q^2 = 10\,\mathrm{GeV}^2$ (a) and $Q^2 = 10^4\,\mathrm{GeV}^2$ (b). Coloured band widths represent the theoretical uncertainty. Figure from Ref. [6].

being ν the lepton energy loss and M the proton mass, is defined. According to the Bjorken scaling law [7], the charge distribution within the nucleon only depends on x. The partonic content of the nucleon can be thus described with only two structure functions:

$$F_1(x) = \frac{1}{2}\sum_i e_i^2 f_i(x) \tag{I.1.5}$$

$$F_2(x) = \sum_i x e_i^2 f_i(x) \tag{I.1.6}$$

where e_i denotes the parton fractional electric charge, being i an index running over the parton species. The functions $f_i(x)$ are called Parton Distribution Function (PDF), are shown in Fig. I.1.3 and describe the probability density for a parton to present a given Bjorken-x value at a Q^2 energy scale. They can not be calculated using perturbative QCD, but fitting observables to experimental data. On the opposite, the PDFs evolution with the Q^2 is described by the DGLAP equations [8], allowing the results from previous experiments to be applied at higher energies to modern accelerators. Figure I.1.3 compares the PDFs at two energy scales. In both cases, valence quarks, determining the nucleon quantum numbers, peak at high-x values and are more dominant the lower the Q^2 energy scale. The low-x region is dominated by sea quarks, that are continuously produced as quark-antiquark pairs in gluon splitting, the latter demonstrated to carry about half of the nucleon momentum.

In hadronic collisions, an observable that is used to infer the x values of the colliding partons is the Feynman-x x_F, defined as the longitudinal momentum of the observed produced particle p_L^* with respect to its maximum value in the centre-of-mass frame:

$$x_F = \frac{p_L^*}{|\max(p_L^*)|}. \tag{I.1.7}$$

It varies in $[-1, 1]$ and can be demonstrated with the energy and momentum conservation laws:

$$x_F = x_1 - x_2, \tag{I.1.8}$$

being x_1 and x_2 the Bjorken-x values of the two colliding partons.

For energies below Λ_{QCD}, where a perturbative QCD description is not possible, effective theories or phenomenological models are used in place of the first principles. As an example, according to the Gribov-Regge field theory [9], the parton-parton interactions are described as the exchange of colour-neutral objects called pomerons. With the increasing of the distance between the two partons, causing the strong interaction intensity to increase according to the evolution of the coupling constant discussed above, the continuous emission of decreasing-x particles is interpreted as a colour field string. Its breaking into fragments produces the hadrons. These models are employed by the MonteCarlo event generators, producing the simulated events that are needed to analyse the data collected at the experiments. Among these, EPOS-LHC [10], able to describe minimum-bias data collected in proton-nucleus collisions at the LHC accelerator, will be considered for the analysis of a sample with pHe collisions discussed in this thesis. A comparison of its characteristics with other widely-used models describing the hadron-nuclear collisions that are relevant for Cosmic Ray (CR) physics, such as QGSJETII04 [11], can be found in Ref. [12].

I.1.3 The Dark Matter

Despite the predictions of the SM have been tested and confirmed with increasing precision at several experimental facilities on Earth, astrophysical and cosmologic observations demonstrate that it is only able to describe ∼5% of the energy and ∼ 15% of the mass composition of the Universe. The other ∼ 85% mass, corresponding to ∼26.4% [6] of the total energy according to the conventional interpretation of cosmological measurements, has an unknown nature and is called Dark Matter (DM). The first evidence for an unexplained mass in space was reported in 1933 by Fritz Zwicky. By observing the Coma cluster, he noticed that, according to the visible mass, galaxies could not be bound [13, 14]. The result was only confirmed after four decades, when Vera Rubin measured the evolution of the M31 galaxy rotational energy with the radial distance. While, as illustrated in left Fig. I.1.4, a dependence $v \propto r^{-1/2}$ is expected from the classical laws of mechanics, a constant rotational velocity was observed [15]. Only including a non-visible mass in the galaxy halo, the same previously assumed by Zwicky, data and expectations could agree. In more recent times, two types of observations particularly contributed to establish the existence of DM in the Universe:

- the gravitational lensing, a consequence of the description of gravity according to the general relativity. This is defined as the deviation of the light propagation caused by a gravitational field and allows an observer to look to a luminous body even if there is another massive one in between. This can be an astrophysical body, hence observable, or DM. As an example, right Fig. I.1.4 shows a composite image resulting from the superimposition of the observations of a collision of two clusters of galaxies in different regions of the electromagnetic spectrum [16]. Pink-shaded areas represent

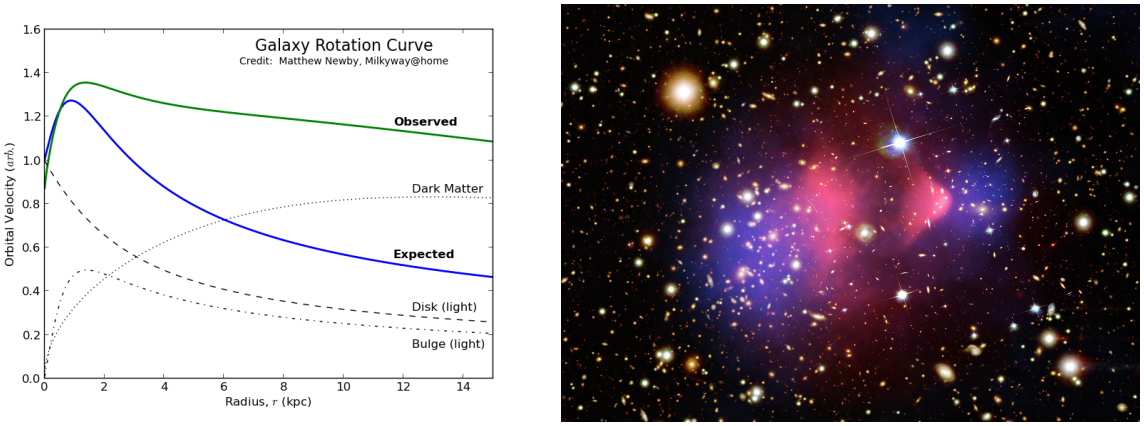


Figure I.1.4: Two of the main evidences for dark matter. The left plot presents the expectation and the measurement of a galaxy rotational velocity as a function of the radial distance from the galactic centre, as re-elaborated from the pioneering measurement by Vera Rubin [15] in the 1970s. A deviation from the evolution $v \propto r^{-1/2}$, predicted by the classic mechanics laws, can be explained as a dark matter contribution. The right plot shows a composite imagine for the observation of the so-called "bullet cluster" [16]. Pink areas correspond to the hot gas measured by the Chandra telescope; white and orange points are the galaxies as observed in the optical spectrum by the Magellan and Hubble telescopes; blue regions represent the mass distribution, as resulting from the gravitational lensing effects. The separation between the pink and blue areas demonstrates that most of the mass is dark.

the hot gas, as observed from its X-rays emission by the Chandra telescope, white and orange points are galaxies, detected in the optical by the Magellan and Hubble telescopes and the blue regions represent the mass distribution, as resulting from the gravitational lensing of distant sources. The separation between the hot gas and the mass regions demonstrates that most of the mass in the clusters is dark. Moreover, the different shapes between the pink and blue distributions demonstrate that DM does not interact with ordinary matter other than gravitationally. Indeed, while baryonic gas is subjected to the electromagnetic force and thus slows down with respect to the galaxies, the gravitational lensing shows that the mass concentrates far from the collision;

- the cosmological evidences, such as the the observation of the anisotropies of the Cosmic Microwave Background (CMB) [17] by the WMAP [18] and Planck [19, 20] experiments. The description of their data for the power spectrum of the CMB temperature fluctuations is indeed only possible by a cosmological model accounting for a 26.4% DM contribution to the energy composition of the Universe.

In order to explain the nature of a mass that does not emit, absorb or scatter lights, since it is not observable with traditional telescopes, that was formed in the early stages of the Universe and constitutes $\sim 84\%$ and $\sim 26\%$ of the total mass and energy, physics beyond the SM is required. Indeed, there is no evidence that it is subject to any interaction other than the gravitational one and that its nature is compatible with a known particle. The proposed candidate DM particles fall into three categories, according to the velocity:

- cold DM, in case it is slow with respect to the speed of light. The GeV- TeV

energy range was considered since a natural extension of the SM. This is indeed populated by most of the known particles and would explain the DM abundance as thermal relics from the early stage of the Universe. At the time I am writing this thesis, several candidates grouped under the name of Weakly-Interacting Massive Particles (WIMP) have been proposed by possible extensions of the SM, such as the SuperSymmetry (SUSY), but never confirmed. At higher mass scales, large and dense objects such as primordial black holes, neutron stars or non-luminous objects like planets have been proposed, but are currently disfavoured by the astrophysical constraints [21];

- hot DM, in case it has ultrarelativistic properties. Other than the SM neutrinos, that are disfavoured since their mass is too light [22], the axions, particles that are introduced in the context of the QCD theory, are being studied;

- warm DM, with intermediate properties between cold and hot DM.

Regarding the WIMPs, these are extensively searched for using the following three main experimental strategies. Direct searches exploit a possible scattering that DM can have with an ordinary particle, which would manifest as a low-energy recoil of the latter and a signal such as the emission of scintillation light. These experiments are usually installed deep underground to minimise the background from incoming cosmic rays and instrument large experimental volumes with cryogenic or noble liquid detectors. As an example, the DAMA/LIBRA [23] experiment is installed at the Gran Sasso INFN national laboratories and equipped with $\sim$250 kg of NaI(Tl) scintillators. By cumulating data over many years, the experiment has claimed an annual modulation caused by the Earth motion around the Sun and thus a different relative velocity of the detector with respect to the DM halo [24, 25]. No confirmation from an independent result of this claim has been given yet. Another followed experimental strategy measures the products of a possible annihilation or decay DM process in space and compares these with the expectations from known sources. An example of these searches, addressing antiprotons, will be discussed in details in the next chapter, being connected with the research activities object of this work. Finally, several searches at colliders, notably the Large Hadron Collider (LHC) at CERN, are ongoing. In this case, a DM particle could be produced in the pp collisions and indirectly detected as the recoil of a SM particle together with a large amount of missing energy. Also, decays or interactions of a long-living DM particle could give signal in existing or proposed LHC experiments [26], such as CODEX-b.

I.2 Cosmic antiprotons

In this chapter, the key elements of indirect Dark Matter searches are summarised. Sec. I.2.1 recalls the main characteristics of cosmic rays, Sec. I.2.2 particularly focuses on the theoretical and experimental advances for the measurement of the antiproton flux in cosmic rays, motivating part of the research activities documented in this work.

I.2.1 Introduction to Cosmic Rays

Defined as the flux of particles and nuclei coming from space, Cosmic Rays (CRs) have played a fundamental role in the development of particle physics before the advent of the accelerator facilities and, still nowadays, offer a unique laboratory for fundamental research in High-Energy Physics (HEP). Indeed, while at accelerator facilities on Earth a maximum energy scale of 10^{13} eV has been accessed, CRs with an energy E up to 10^{20} eV have been measured. The CRs energy spectrum is described by a steeply falling near-power law:

$$\frac{dN}{dE} \propto E^{-\gamma}, \tag{I.2.1}$$

with the spectral index γ ranging between 2.5 and 3.1 depending on the energy range. As presented in left Fig. I.2.1, rapid variations of the spectral index, which are expected to correspond to a change in the physics processes involved in the CR acceleration or propagation, are located:

- at the so-called first and second knee, at $\mathcal{O}$ (10^{15} eV) and $\mathcal{O}$ (10^{17} eV) energies, respectively, where the spectral index increases. The origin of the first knee, where γ increases from 2.7 to 3.1, is explained as the breakdown of the galactic CRs acceleration mechanism, as suggested by the cut-off energy dependence on the atomic number;
- at the so-called ankle, at $\mathcal{O}$ (10^{18} eV) energy. The spectral index here flats back to $\gamma \simeq 2.7$, this being interpreted as a higher-energy flux of extra-galactic particles overtaking the galactic ones.

Finally, a steep cut-off is observed over 10^{20} eV as expected from the CRs interaction with the photons of the CMB:

$$p\gamma_{CMB} \rightarrow \Delta^{+}, \tag{I.2.2}$$

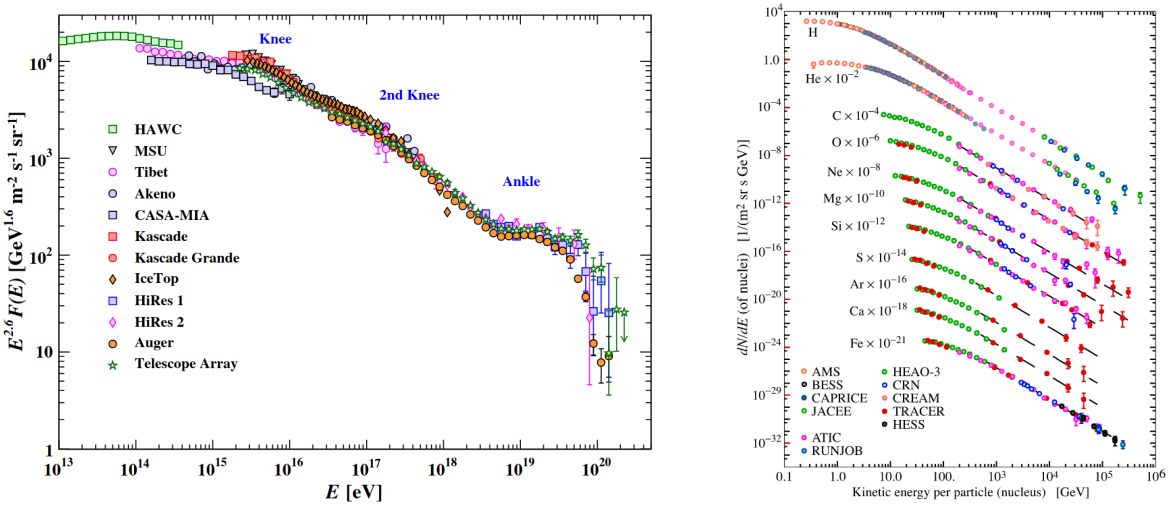


Figure I.2.1: Cosmic rays energy spectrum for all particles, drawn multiplied by $E^{2.6}$ to highlight the energy values with a changing slope (left), and for separate nuclear species (right). Both plots result from the combination of the measurements of several experiments, indicated in the legends. Figures from Refs. [6] and [27], respectively.

followed by a $\Delta^+ \to p\pi^0$ or a $\Delta^+ \to n\pi^+$ decay. For lower energy values, the spectrum is measured splitting by element. This is illustrated in right Fig. I.2.1 for nuclei ranging from hydrogen to iron and a similar evolution is found when rescaling by the number of nucleons.

Depending on their production, CRs are divided into two groups:

- primary, when accelerated by astrophysical sources. Among the primary particles, $\sim 74\%$ are free protons, $\sim 18.2\%$ α particles and the rest is divided among nuclei (especially C, O and Fe), electrons, photons and neutrinos;
- secondary, when originating from fragmentation processes with the InterStellar Medium (ISM), mostly composed of hydrogen and helium.

Concerning the nuclei, the CRs elemental composition (in black) is compared with the measured one in the solar system (in red) in Fig. I.2.2. The observed consistency for most of the nuclei suggests a stellar CR origin, but, on top of the see-saw evolution, caused by the odd-Z nuclei weaker binding energies, two evident differences in the Li-Be-B and Sc-V-Mn groups emerge. All these nuclei are mostly secondary. Measurements of the relative elemental or isotope abundances are performed to characterise the CRs propagation. For example, with a secondary-to-primary nucleon ratio, notably with boron and carbon because of their low atomic numbers, the quantity of ISM traversed by the CR during its propagation is measured. With a ratio between a stable and a decaying isotope of the same secondary element, such as ^{9}Be/^{10}Be, the CRs propagation time in the galaxy can be measured [29].

Three strategies are being followed to the CRs characterization:

- space experiments, located outside of the Earth atmosphere, directly measure the CR components. Considering the detector size that is allowed by the operations

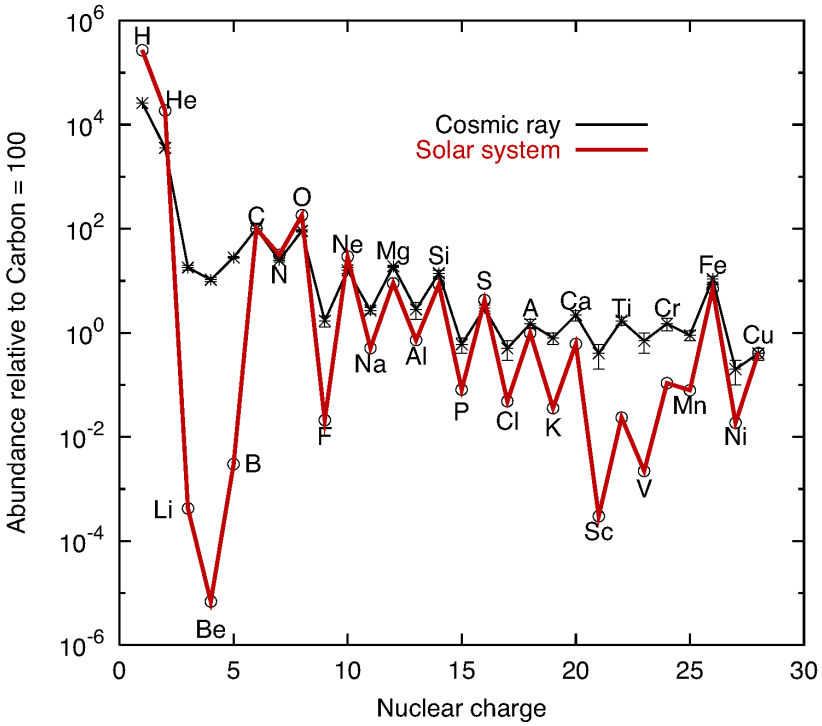


Figure I.2.2: Comparison between the elemental composition measured in CRs (in black) and in the solar system (in red) relative to carbon, arbitrarily set to 100. A similar see-saw evolution is observed, except for the Li-Be-B and Sc-V-Mn groups, produced in fragmentation processes with the interstellar gas. Figure from Ref. [28].

in space, they have a limited exposure to data and are thus conceived for energies below the first knee, where the flux is such that a sufficient statistics can be collected. Among these, some are based on a spectrometer, such as the PAMELA [30] and AMS-02 [31] experiments that will be discussed in more details below, others are mainly equipped with a calorimetric system, such as CALET [32];

- ground-based experiments address ultra-high energy CRs that are detected by reconstructing the shower of particles they produce when impinging on the Earth atmosphere, with the same working principle than a calorimeter. By arranging in a large area a matrix of detectors, the core of the shower, correlated to the point of the atmosphere hit by the primary CR, is reconstructed. As an example, the Pierre Auger observatory [33] extends over a 3000 km^2-wide region in Argentina and has measured CRs with an energy up to 10^{20} eV;

- balloon-flight-based experiments are intermediate between the previous two categories, suffer from CR interactions with a residual fraction of the atmosphere, but are much cheaper and easier to design and to operate with respect to experiments in space. An example of these is GAPS [34].

To a correct interpretation of the collected data, all the steps of the CR propagation up to the detection must be considered. For galactic CRs, the sources are well established [35] to be supernovae remnants. Active galactic nuclei, compact and intense luminous regions located in the centre of some galaxies or fast emissions of γ particles known as Gamma

Ray Bursts could be the origin of extragalactic CRs. Regarding the galactic CRs, the mechanism allowing the CR particles to be accelerated was firstly proposed in the 1950s by Enrico Fermi who, despite the scarcity of experimental data, was able to reproduce the energy spectrum power law even if with a softer spectral index [36]. The core idea of his theory was that a charged particle of a CR is iteratively accelerated as a consequence of the scatterings with the randomly-distributed magnetic fields in interstellar clouds. The resulting average energy gain $\langle \Delta E/E \rangle$ is related to the cloud velocity β as [37]:

$$\left\langle \frac{\Delta E}{E} \right\rangle \sim \frac{1+\beta^2/3}{1-\beta^2} - 1 = \frac{4}{3}\beta^2. \tag{I.2.3}$$

The effect, being proportional to β^2 and called for this reason second-order Fermi acceleration mechanism, is rather inefficient to explain the CRs acceleration and their observed power-law energy spectrum. Fermi improved thus the theory describing the cloud as a fluid and including the acceleration at the shock waves. The interaction of the CR with a planar shock, defined as a rapid discontinuity of some physical quantity such as the gas density, induces on average an acceleration [37]:

$$\left\langle \frac{\Delta E}{E} \right\rangle = \frac{4}{3}\beta. \tag{I.2.4}$$

The first-order Fermi acceleration mechanism is linearly proportional to the shock wave velocity β (that can reach 10^3 times the interstellar cloud one) and can be demonstrated to give a power-law energy spectrum for the CR particle.
The evolution of the CR density $\psi_i(\vec{r}, p, t)$ with the position $\vec{r}$, the momentum p and the time t, being i an index denoting the particle species, is described by the propagation equation. This takes into account the production, re-acceleration and all loss effects occurring during the propagation as [38]:

$$\begin{aligned} \frac{\partial \psi_i(\vec{r},p,t)}{\partial t} = {} & q_i(\vec{r},p,t) + \vec{\nabla}\cdot(D_{xx}\vec{\nabla}\psi_i) + \frac{\partial}{\partial p}\left[p^2 D_{pp}\frac{\partial}{\partial p}\left(\frac{\psi_i}{p^2}\right)\right] - \frac{\partial}{\partial p}(\dot{p}\psi_i) \\ & -\vec{\nabla}\cdot(\vec{v}_c\psi_i) + \frac{\partial}{\partial p}\left[\frac{p}{3}(\vec{\nabla}\cdot\vec{v}_c)\psi_i\right] - \frac{\psi_i}{\tau_{i,\,frag}} - \frac{\psi_i}{\tau_{i,\,decay}} \end{aligned} \tag{I.2.5}$$

where:

- $q_i(\vec{r}, p, t)$ represents the local production rates, for both the primary CR accelerator and the secondary fragmentation or decay processes of other particles;
- D_{xx} and D_{pp} are the diffusion tensor terms in the position and momentum space, respectively;
- $\dot{p}$ is the momentum time derivative, describing the momentum losses caused by the radiative processes, especially for electrons, or by interactions with the ISM and with the galactic magnetic field;
- $\vec{v}_c$ is the convection velocity, describing the CR interaction with galactic winds;
- τ_{frag} and τ_{decay} are the time-scales for the CR nuclear fragmentation or radiative decay losses, if any.

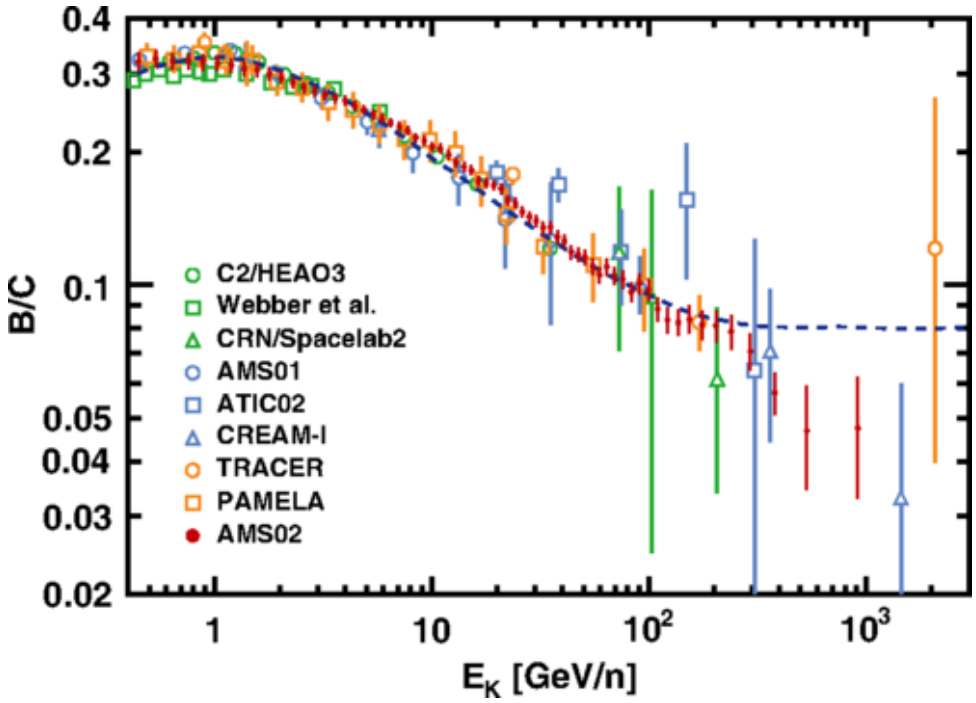


Figure I.2.3: Evolution of the ratio between the boron and carbon abundances in cosmic rays as a function of the kinetic energy, as resulting from the observation of different experiments, indicated in the legend. Figure from Ref. [39].

Regarding the propagation, the galactic magnetic field B presents a large-scale symmetry, oriented as the Galaxy spiral arms, and irregularities on lower scales. These, acting as scattering centres, induce a stochastic and isotropic propagation of the CR particle and hence its diffusive description. The particle rigidity, defined as $R = pc/Ze$ being Ze its charge, is introduced to describe the propagation. Indeed, a particle with velocity $\vec{v}$ interacting with a magnetic field is described by:

$$m\gamma\frac{d\vec{v}}{dt} = Ze\,(\vec{v}\times\vec{B}). \tag{I.2.6}$$

Being the deflection only determined by the transverse velocity component $v_\perp = vsin\theta$, with θ the angle between the velocity and magnetic field vectors, the resulting motion is circular with radius:

$$r = \frac{m\gamma v}{Ze}\frac{sin\theta}{B} = \frac{pc}{Ze}\frac{sin\theta}{Bc} = \frac{R}{Bc}sin\theta. \tag{I.2.7}$$

Other than the galactic magnetic fields, CRs with a rigidity below 10 GV are also affected by the solar and geomagnetic effect before the detection.
The CR fragmentation in the ISM can be described as:

$$\frac{\partial\psi_i}{\partial t} = -\frac{\psi_i}{\tau_{frag}} = -n_{ISM}v\sigma_i\psi_i \tag{I.2.8}$$

being n_{ISM} the ISM particle density and σ_i the interaction cross-section. For the former, the hydrogen particle density is $n_{ISM,\,\mathrm{H}} \simeq 1\,\mathrm{cm}^{-3}$, the helium one is $n_{ISM,\,\mathrm{He}} \sim 0.1\cdot n_{ISM,\,\mathrm{H}}$ and heavier elements are only present in traces [40]. As discussed in the next section for

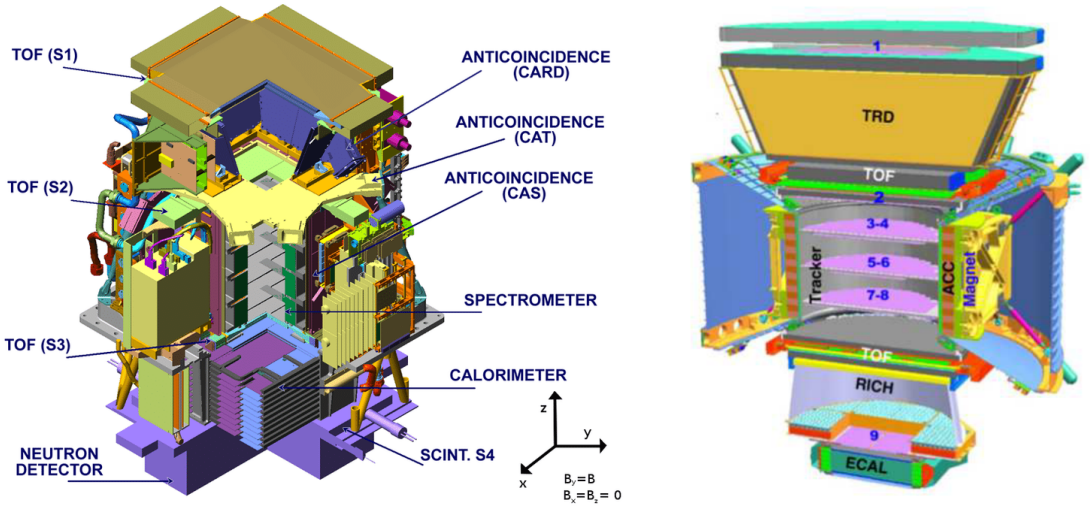


Figure I.2.4: Sketches of the PAMELA (left) and AMS-02 (right) detectors. Figures from Refs. [41] and [42], respectively.

processes producing antiprotons, σ_i is obtained by combining the cross-sections with the separate ISM chemical elements as a target weighted by the respective abundances. In particular, by applying the propagation equation to a secondary and a primary CR, such as boron (N_B) and carbon (N_C), the production rate of the former can be expressed as:

$$\frac{\partial N_B}{\partial t} \simeq n_{ISM} v \sigma_{C \to B} N_C \simeq \frac{N_B}{\tau} \qquad \Longrightarrow \qquad \frac{N_B}{N_C} \simeq n_{ISM} v \sigma_{C \to B} \tau \tag{I.2.9}$$

where the stationary hypothesis $\partial N_B / \partial t \simeq N_B / \tau$ is assumed and indicating $\sigma_{C \to B}$ the carbon-to-boron fragmentation cross-section. Figure I.2.3 presents the superimposition of the measurements of the N_B/N_C ratio (also denoted with B/C, as in the figure) by different experiments, notably the recent AMS-02 one covering rigidity values up to 2.6 TV [39], as a function of the per-nucleon kinetic energy. The B/C ratio is found to decrease with the energy with no structures and is used to constrain the CRs propagation parameters.

I.2.2 Measurements of the antiproton flux

For DM indirect searches, measurements of the antimatter flux in CRs are ongoing. Antimatter is indeed expected to be rare and only produced in the CRs-ISM processes and its flux is thus very sensitive to an exotic contribution such as a DM decay or annihilation process. The most recent measurements for antimatter fluxes in space have been carried out by two experiments:

- the Payload for Antimatter Matter Exploration and Light-nuclei Astrophysics (PAMELA) experiment [30] operated between 2006 and 2016, installed on a satellite. It focused on the detection of the CRs charged component, particularly positrons and antiprotons over the sensitivity ranges $50\,\text{MeV} - 270\,\text{GeV}$ and $80\,\text{MeV} - 190\,\text{GeV}$, respectively. Its physics goals also included the search for anti-nuclei, with a sensitivity of 10^{-7} for the antihelium-to-helium ratio and the study of the solar

modulation effects for low-energy CR particles. Its detector, illustrated in left Fig. I.2.4, was $\sim$ 1.3 m high with a 470 kg mass and instrumented a magnetic spectrometer. With a permanent magnet producing a 0.43 T magnetic field and six silicon-based tracking stations, the trajectory of charged particles was reconstructed, measuring the charge and the rigidity. Particle identification was performed with an electromagnetic calorimeter and two Time-Of-Flight (TOF) scintillator planes, also providing the trigger for data acquisition. The PAMELA results for the antimatter fluxes will be briefly presented below, while a summary of all its main scientific observations can be found in Ref. [43];

- the Alpha Magnetic Spectrometer (AMS-02) [31] experiment was installed onboard the International Space Station in 2011 and is still acquiring data. Its physics programme is similar to PAMELA, but the larger dimensions, the detector improvements and the longer planned duration of the operations are expected to increase the measurements precision and to extend the covered energy (up to $\mathcal{O}$ (TV) rigidities) and Z (up to 26) ranges. Similarly to PAMELA, the AMS-02 detectors, shown in right Fig. I.2.4, are arranged around a permanent magnet, producing a 0.15 T magnetic field, provide the measurement of the charge and the rigidity with nine silicon tracking planes over a level arm of $\sim$ 3 m and of the energy with an electromagnetic calorimeter. The data acquisition is triggered by the TOF measurement stations. Two additional detectors with respect to PAMELA are installed to improve the particle identification performance: a Transition Radiation Detector (TRD), exploiting the emission of the transition radiation to distinguish electrons and positrons from (anti)protons and the energy loss to identify the nuclei, and a Ring Imaging CHerenkov detector (RICH) detector, offering the redundancy to distinguish among different particle species.

Both experiments have performed indirect DM searches considering positrons, antiprotons and other light nuclei such as anti-deuterium and anti-helium. For the first channel, while electrons can be produced by the primary CRs sources, positrons are assumed to only be secondary and a decreasing positron fraction with the energy would be expected. The PAMELA [44] (top left) and AMS-02 [47] (top right) results are presented in Fig. I.2.5 and both observe a clear disagreement with the pure secondary positron origin (solid black line) for $E > 10$ GeV. To understand the positron rising, several explanations have been proposed and are drawn in bottom Fig. I.2.5. Some introduce a positron primary emission as a DM annihilation or decay process [48] (in green) or by pulsar stars [49] (in red), others suggest a description of the electrons and positrons propagation in the galaxy taking into account the spiral arm patterns of the sources, as suggested by astrophysical observations [50] (in blue). The best description of the AMS-02 results is now provided by the pulsar emission hypothesis, but more data are required to decrease the uncertainty and to further extend the covered energy range.

A second measurement that has been considered for indirect DM searches is the antiproton-to-proton ratio. In this case, the secondary production expectations, confirmed by balloon-borne experiments [51] were at $\mathcal{O}$ (10^{-4}) for $E \simeq 10$ GeV and, being antiprotons more difficult to produce in known astrophysical processes, an observed excess over the expected flux would be expected at higher energy, if the positron excess was due to DM annihilation. On the contrary, if antiprotons are found to be consistent with the secondary

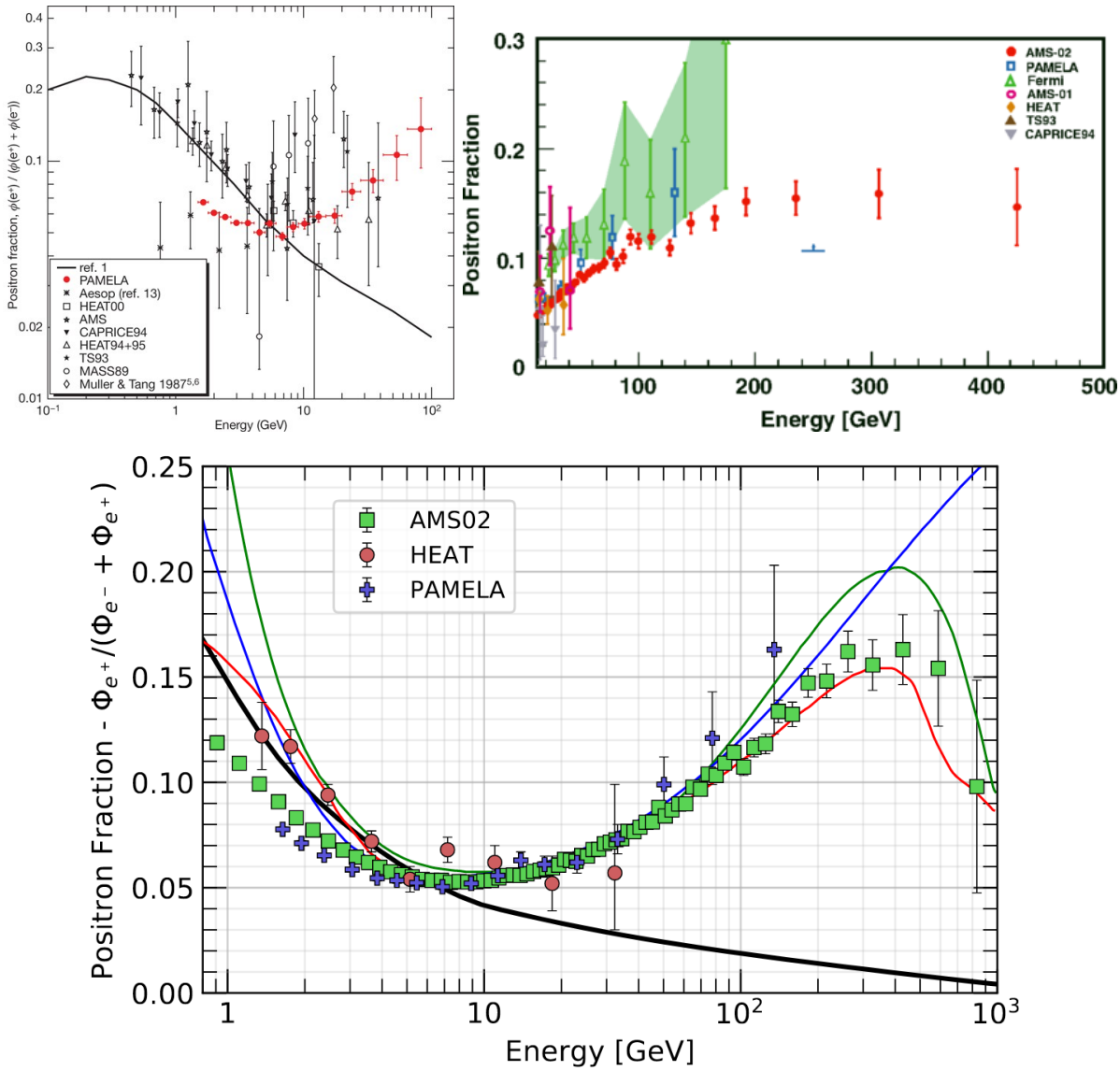


Figure I.2.5: Measurement of the positron fraction in cosmic rays by the PAMELA (top left) and AMS-02 (top right) experiments. The bottom plot compares the two results with a previous measurement, with a prediction from a pure secondary positron origin (black solid line), with a dark matter decay (in green), with a different propagation description (in blue) and with the emission from pulsars (in red). Figures from, in order, [44], Refs. [45] and [46].

production hypothesis, this would support a non-exotic positron source. A first result for the antiproton-to-proton ratio by PAMELA [52] up to a 100 GeV kinetic energy (top) and its extension by AMS-02 [53] (bottom) up to a 450 GV rigidity, shown in Fig. I.2.6, demonstrated that the antiproton fraction flattens to $\sim 2 \cdot 10^{-4}$ for $E > 10$ GeV. For the latter, a sample of 65 billion CR events have been recorded and antiprotons are selected combining the TRD, TOF, tracking and energy loss information to be compatible with $|Z| = 1$. The contamination of the selected events from charge-confusion candidates and other particle species produced in interactions with the detector material is suppressed considering the TOF, RICH and calorimeter particle identification information. The proton and antiproton fluxes are measured in rigidity bins and then divided so that

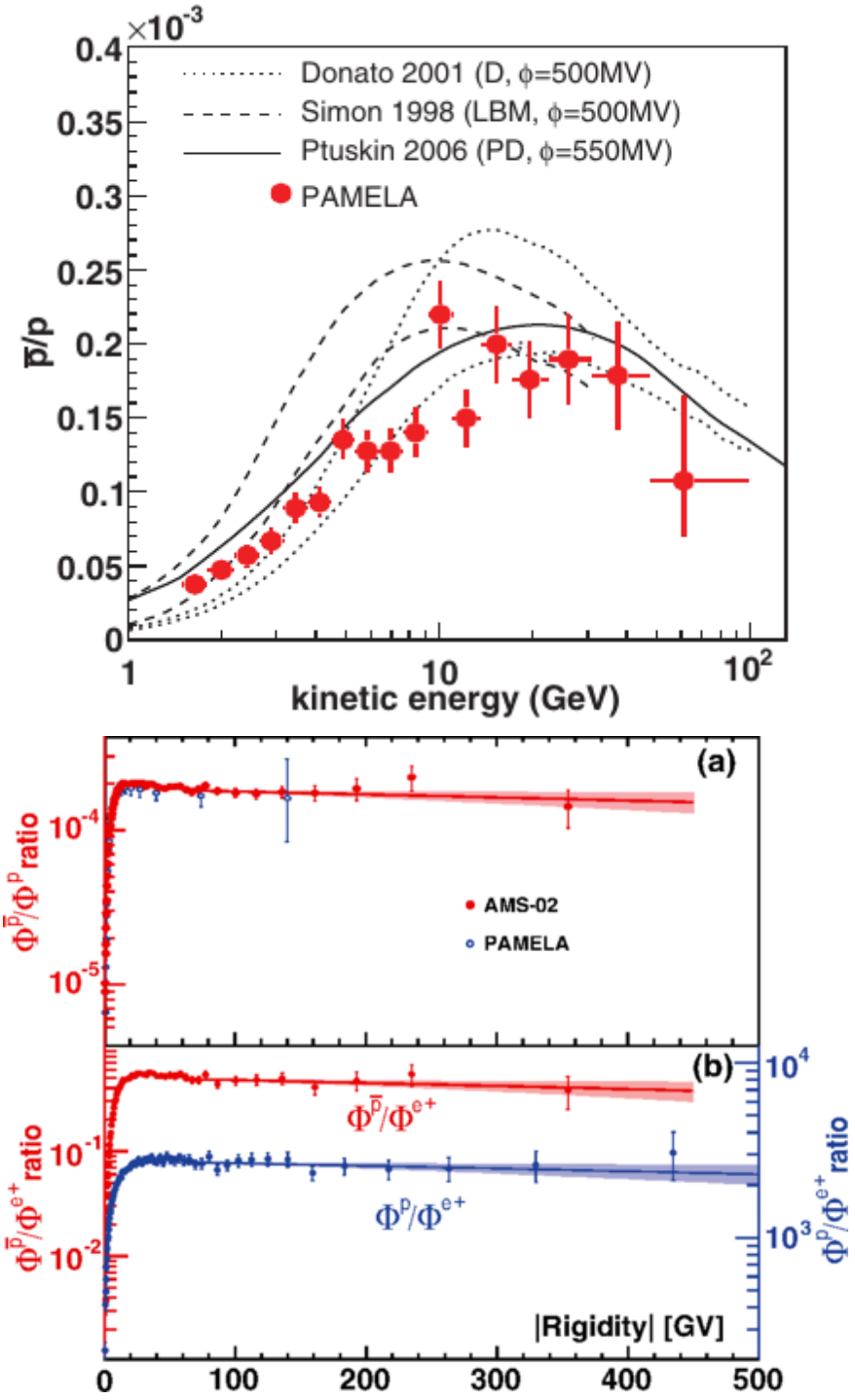


Figure I.2.6: Measurements of the antiproton-to-proton flux as a function of the energy by the PAMELA (top) and of the rigidity by the AMS-02 experiments (bottom). For $E \geq 10\,\text{GeV}$, an approximately constant ratio at $\sim 2 \cdot 10^{-4}$ is found. Figures from Refs. [52] and [53].

most of the systematic uncertainties cancel out. Recently, the AMS-02 results have been updated up to a 525 GV antiproton rigidity and improved in statistics and accuracy [54].

The comparison between the experimental results for the antiproton flux in

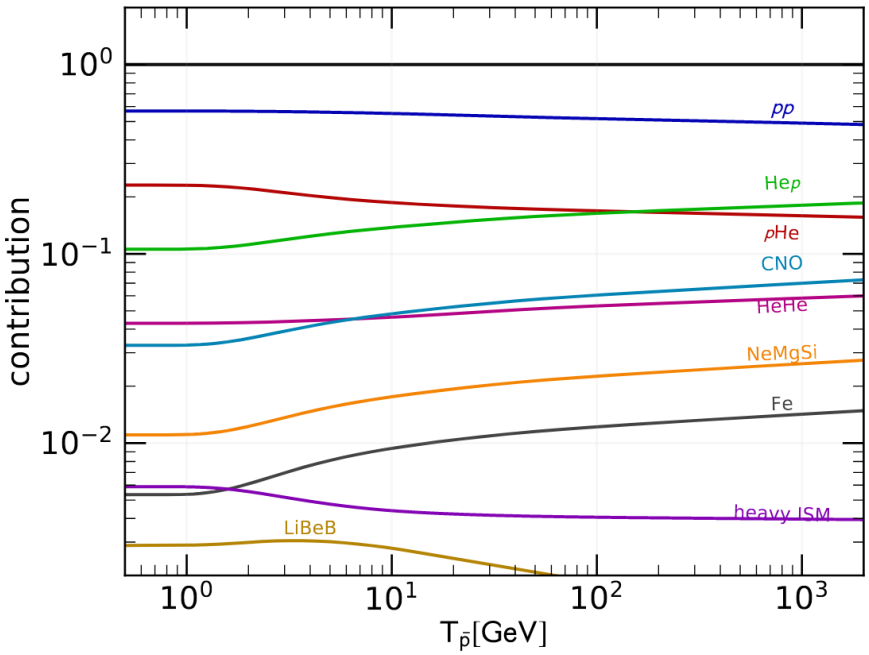


Figure I.2.7: Relative contributions to the antiproton production in CRs-ISM collisions as a function of the energy. The leading term, ∼60% and constant with the energy, comes from *pp* collisions; channels involving helium as a target or projectile account for ∼ 40%, while heavier nuclei only account for few per-cents. Figure from Ref. [55].

CRs and its expected production in CRs-ISM collisions can evidence a possible DM contribution. The main ingredients of the theoretical model and the related leading uncertainties are summarised in the following. Starting from the CRs propagation description in Eq. I.2.1, the source term q_i for purely secondary antiprotons is given by [55]:

$$q_i(T_{\bar{p}}) = \sum_j \int_{T_{th}}^{\infty} dT_i \, 4\pi n^j_{ISM} \psi_i(T_i) \frac{d\sigma_{ij}}{dT_{\bar{p}}}(T_i, T_{\bar{p}}), \tag{I.2.10}$$

being:

- i and j indices running on the incoming CR species and on the ISM components;
- ψ_i the incoming CR flux;
- T_{th} the energy threshold for the antiproton production;
- 4π a factor coming from the assumed isotropy of the incoming CR particle.

Rather than energy-differential $d\sigma/dT_{\bar{p}}$, the cross-section is conveniently expressed in a Lorentz-invariant form $\sigma_{inv}(\sqrt{s_{\rm NN}}, p_{\rm T}, x_R)$ as a function of the colliding nucleon-nucleon centre-of-mass energy $\sqrt{s_{\rm NN}}$, of the antiproton transverse momentum and of a scaling variable such as x_F or x_R, defined as x_F but considering the particle energy in place of the momentum $x_R = E^*_{\bar{p}}/E^{*,max}_{\bar{p}}$. The relation between the two expressions is evaluated

in Ref. [56] as:

$$\frac{d\sigma}{dT_{\overline{p}}}(T, T_{\overline{p}}) = \int d\Omega \frac{d^3\sigma}{dE_{\overline{p}} d\Omega} = \int_0^{2\pi} d\phi \int_{-1}^{1} d(cos\theta) p_{\overline{p}}^2 \frac{dp_{\overline{p}}}{dE_{\overline{p}}} \frac{d^3\sigma}{p_{\overline{p}}^2 dp_{\overline{p}} d\Omega} = \tag{I.2.11}$$

$$= 2\pi \int_{-1}^{1} d(cos\theta) p_{\overline{p}}^2 \frac{E_{\overline{p}}}{p_{\overline{p}}} \frac{1}{E_{\overline{p}}} \sigma_{inv}(\sqrt{s_{\rm NN}}, p_{\rm T}, x_R) = 2\pi p_{\overline{p}} \int_{-1}^{1} d(cos\theta) \sigma_{inv}(\sqrt{s_{\rm NN}}, p_{\rm T}, x_R).$$

Considering the ISM elemental composition, the antiproton production cross-sections and the primary CRs fluxes, the relative antiproton ij contributions are presented in Fig. I.2.7 as a function of the antiproton kinematic energy. The leading term comes from pp collisions, about 60%, while processes involving helium as the projectile or the target account for $\sim$ 40%. All other interactions involving heavier nuclei sum up to few per-cent at most. With respect to the CRs-ISM production processes, experimental data for $\sigma_{pp\to\overline{p}X}, \sigma_{pHe\to\overline{p}X}, \sigma_{\alpha H\to\overline{p}X}, \sigma_{\alpha He\to\overline{p}X}$ are thus needed. Before the LHCb measurement of the prompt antiproton production in pHe collisions at $\sqrt{s_{\rm NN}} = 110\,\text{GeV}$, described in Sec. I.4.2.2, no data on the channels involving helium were available. Instead, a phenomenological extrapolation was performed from the $pp\to\overline{p}X$ reaction and considering simulation [57] or the available data in pC or other pA collision systems [58]. The underlying hypothesis is that the rescaling is proportional to A^D, being A the atomic mass number and D a coefficient varying between 2/3 if the nucleus is classically described as a sphere and 1 if all nucleons are assumed to interact independently.

The leading contributions to the uncertainty affecting the theoretical prediction for the antiproton-to-proton flux ratio in CRs as known in 2015 are illustrated in Fig. I.2.8 [59]: the primary CR spectrum (top left), the knowledge of the antiproton production cross-sections (top right), the CRs propagation (bottom left) and the solar modulation effect (bottom right). The uncertainty on the primary spectra increases at large energy, where available data are scarce, while solar modulation, as discussed in the previous section, is mostly relevant at low energies. The leading contributions to the uncertainty come from the CRs propagation and from the knowledge of the cross-sections, only poorly constrained by the available data. The model is compared to the PAMELA and AMS-02 antiproton-to-proton ratio measurements as a function of the kinetic energy in bottom Fig. I.2.8. For $E > 10\,\text{GeV}$, data indicate a milder dependence on the energy than expected, with a hint for an excess with respect to the secondary antiproton production and thus a possible exotic antiproton primary component. Anyhow, the systematic uncertainty affecting the model, shown in the figure as the sum of the coloured bands, is such that no conclusive statements can be drawn. Advances in the description of the CRs propagation or experimental measurements for the relevant cross-sections were needed.

Two experimental results were published in 2017 and 2018 by the NA61 and LHCb collaborations, respectively. The former is for the antiproton production and spectra in pp collisions at $\sqrt{s_{\rm NN}} = 6.3$, 7.7, 8.8, 12.3 and 17.3 GeV [60]; the latter is the first determination of $\sigma(p\text{He}\to\overline{p}_{\rm prompt}X, \sqrt{s_{\rm NN}} = 110\,\text{GeV})$ [61]. Both only addressed the prompt antiproton production, excluding the contribution from anti-hyperon decays. Considering that significant scaling violations are expected for $\sqrt{s} > 50\,\text{GeV}$, for which no accurate prediction is available, the new results at high energies had a significant impact in decreasing the modelling uncertainties. This is exemplified by Fig. I.2.9, comparing

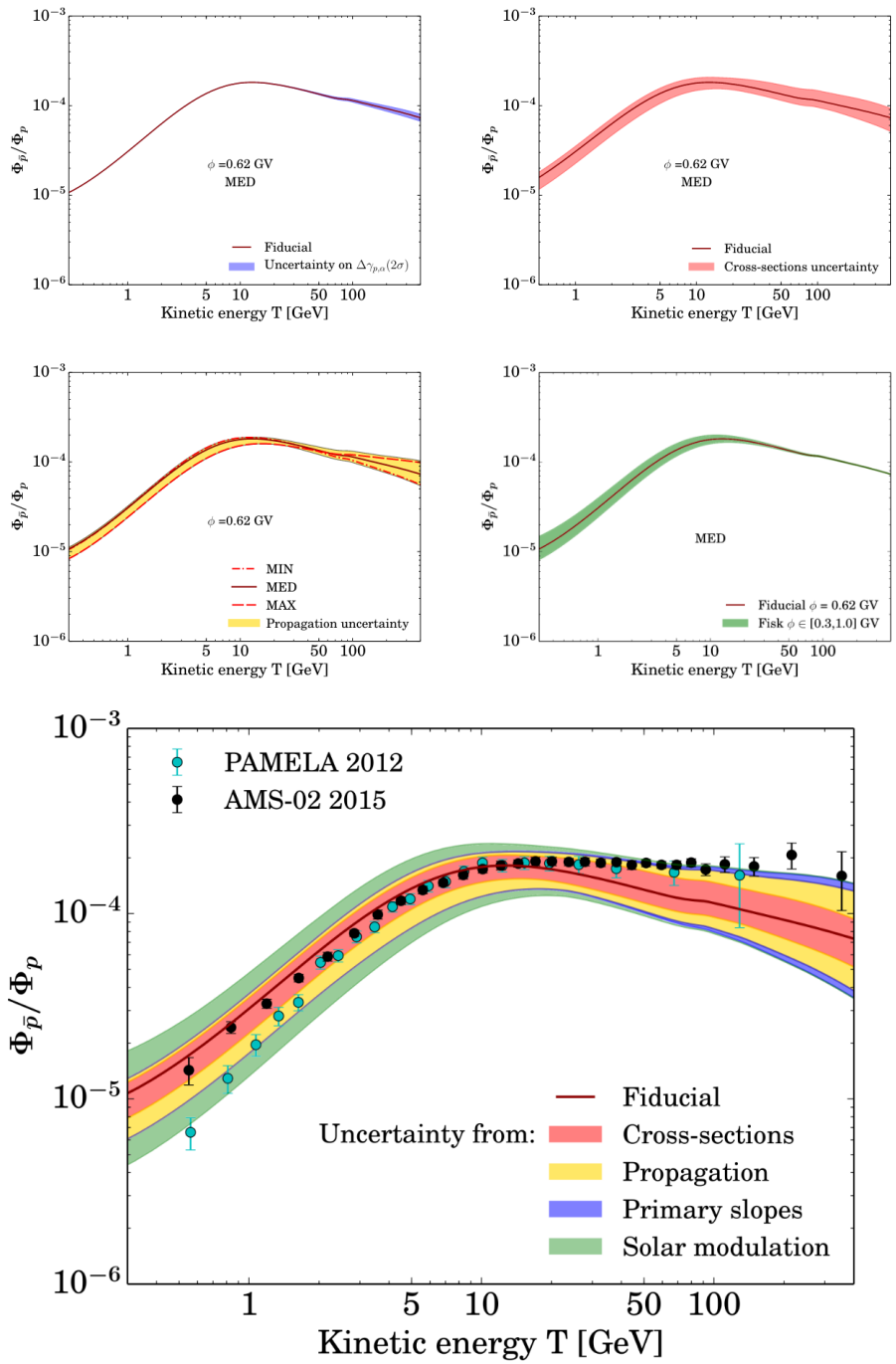


Figure I.2.8: Uncertainties affecting the theoretical prediction for the antiproton-to-proton flux ratio in CRs in 2015: the incoming primary CR spectrum (top left), the antiproton production cross-sections (top right), the CRs propagation (bottom left) and the solar modulation (bottom right). The bottom plot compares the theoretical model and the related sum of all the uncertainties with the experimental data collected by the PAMELA and AMS-02 experiments, which indicate a milder dependence on the energy than the prediction for $E > 10\,\text{GeV}$. Due to the large uncertainty affecting the theoretical prediction, this is still consistent to data and no conclusive statement could be drawn. Figures from Ref. [59].

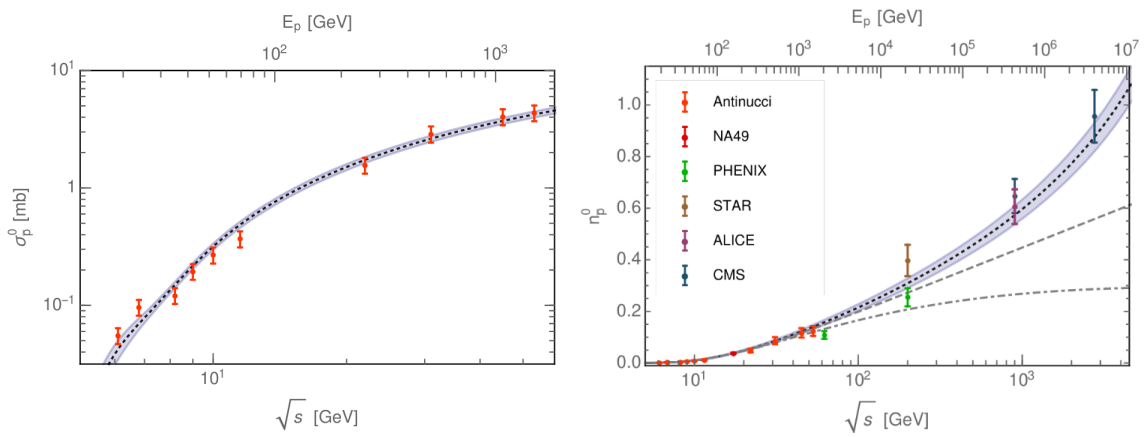


Figure I.2.9: Evolution with the energy of the prompt antiproton production cross-section (left) and multiplicity (right). A scaling law is depicted on the left plot and as the dotted-dashed grey lines on the right plot and is only able to describe experimental data up to $\sim 50\,\text{GeV}$. For higher energies, a violation of the scaling law occurs. Figure from Ref. [64].

$\sigma(pp \to \overline{p}_{\text{prompt}} X)$ for energies up to $\sim 50\,\text{GeV}$ (left) and to $\sim 2\,\text{TeV}$ (right). In the former, the available data by the NA49 collaboration [62] were used to demonstrate that for $\sqrt{s_{\text{NN}}} \in [10, 50]\,\text{GeV}$ the antiproton production invariant cross-section obeys to the scaling law [63, 64]:

$$\sigma_{inv}(\sqrt{s_{\text{NN}}}, p_{\text{T}}, x_R) \quad \to \quad \sigma_{inv}(p_{\text{T}}, x_R) \propto (1 - x_R)^{7.76} exp\Big[- \frac{\sqrt{p_{\text{T}}^2 + m^2}}{0.168} \Big]. \qquad \text{(I.2.12)}$$

The extrapolation of the same function at higher energies, illustrated as the dotted-dashed grey line in right Fig. I.2.9, fails in describing the LHC data. Only by including a violation of the scaling, *i.e.* an explicit dependence of the cross-section from $\sqrt{s}$, the model correctly describes the data [64].

Two examples of the improvement in the scaling parametrization are presented in Fig. I.2.10, comparing the LHCb data with predictions from different models as a function of the x_F variable (top) and of the momentum (bottom). Such models are clearly constrained by the new data, resulting into a reduced uncertainty on the energy evolution of the $\sigma(p\text{He} \to \overline{p}X)$ cross-section. An updated evaluation of the antiproton abundance in CRs is then performed [66] and presented in Fig. I.2.11 as a function of the rigidity. When compared to the AMS-02 data, a better agreement is found with respect to I.2.8, thus disfavouring the hypothesis of a primary contribution to the antiproton flux. Anyhow, as resulting from the plot of the residuals of the model application to the AMS-02 data in the bottom, the knowledge of the antiproton production cross-section is still one of the leading contributions and further data are needed. The same conclusion is drawn in Ref. [54], where the most recent AMS-02 data for the antiproton-to-proton flux ratio are compared with theoretical models including or not a DM antiproton contribution. Despite the opposite assumption, a similar agreement with the data is found.

At the $\mathcal{O}$ (10-110 GeV) energy scales, the LHCb fixed-target programme can provide experimental inputs covering all the antiproton production components to increase the accuracy of the models. The invariant antiproton production cross-section can indeed be

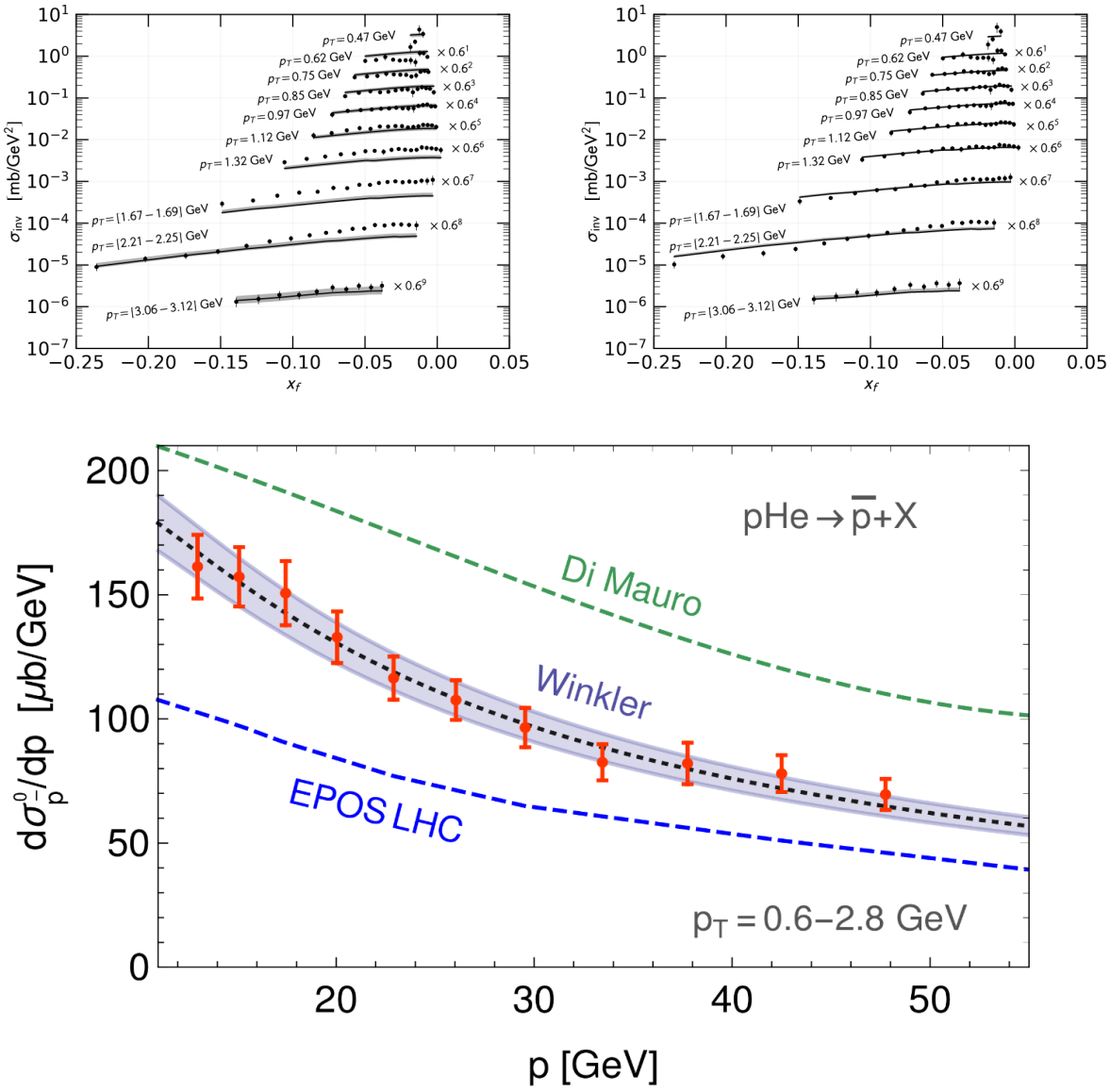


Figure I.2.10: Comparison of the LHCb data for the antiproton production in pHe collisions as a function of the Feynman-x variable (top) and the momentum (bottom) with parametrizations assuming different models for the scaling violation. Data clearly indicate one model, thus decreasing the uncertainty on the cross-section predictions. Figures from Ref. [55] and presented in Ref. [65] based on Ref. [64], respectively.

expressed as:

$$\sigma_{inv} = \sigma^0_{inv}(2 + \Delta_{IS} + 2\Delta_{\bar{\Lambda}}), \tag{I.2.13}$$

where:

- σ^0_{inv} is the prompt $\overline{p}$ production, which is dominant and the only one addressed by the 2018 LHCb results, as discussed in Sec. I.4.2.2. For the future, the measurements can be extended towards lower energies in the centre-of-mass frame, this ensuring a larger coverage of the antiproton production phase-space and allowing the gap to the SpS energy scales to be covered. As discussed in Sec. I.4.2.1, other than

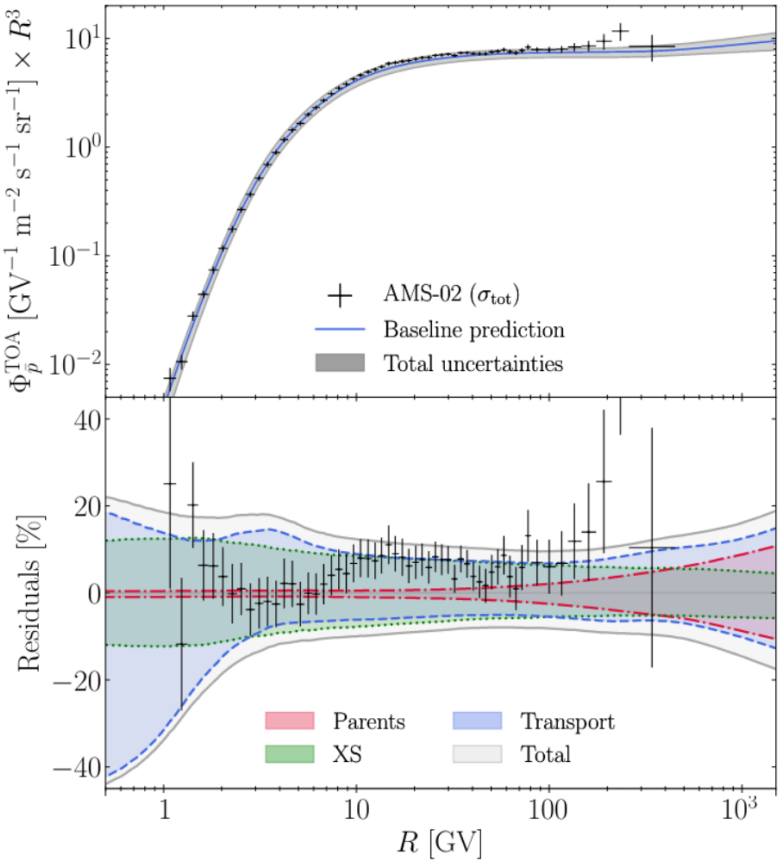


Figure I.2.11: Theoretical prediction for the antiproton abundance in CRs as a function of the rigidity in a 2019 publication [66]. The top plot compares the prediction with the AMS-02 data and finds a better agreement with respect to Fig. I.2.8. The bottom plot presents the residuals of the model and the single contributions to the uncertainty. The knowledge of the production cross-section, denoted as XS in the figure, is still one of the leading contributions. Figure from Ref. [66].

the sample at $\sqrt{s_{\rm NN}} = 110\,\text{GeV}$ analysed for the 2018 result, another one with pHe collisions at $\sqrt{s_{\rm NN}} = 87\,\text{GeV}$ was collected in 2016. Lower energy values, possibly down to $\sqrt{s_{\rm NN}} = 29\,\text{GeV}$, are expected to follow, giving a better coverage of positive x_F values, as discussed in Sec. IV.2.3;

- $\Delta_{\bar{\Lambda}}$ is the ratio between the decay- and promptly-produced antiprotons, defined as

$$\Delta_{\bar{\Lambda}} = \frac{\bar{\Lambda}}{\bar{p}_{\rm prompt}} \times \mathcal{B}(\bar{\Lambda} \to \bar{p}\pi^+) + \frac{\bar{\Sigma}^-}{\bar{p}_{\rm prompt}} \times \mathcal{B}(\bar{\Sigma}^- \to \bar{p}\pi^0), \tag{I.2.14}$$

where $\bar{\Lambda}(\bar{\Sigma}^-)/\bar{p}_{\rm prompt}$ is the ratio between the $\bar{\Lambda}$ ($\bar{\Sigma}^-$) and prompt antiproton cross-sections and $\mathcal{B}$ the branching ratios for the indicated decays. With symmetry arguments, $\bar{\Sigma}^-/\bar{\Lambda} \sim 0.33 \pm 0.08$, which leads to $\Delta_{\bar{\Lambda}} = (0.81 \pm 0.04)(\bar{\Lambda}/\bar{p}_{\rm prompt})$. The scarce available data for $\bar{\Lambda}/\bar{p}_{\rm prompt}$ are presented in top Fig. I.2.12 as a function of the centre-of-mass energy and indicate it is constant at ~ 0.3 for $\sqrt{s} < 50\,\text{GeV}$ energies, at ~ 0.6 for $\sqrt{s} > 500\,\text{GeV}$. The energy scale where the strangeness enhancement

occurs is accessible to LHCb in a fixed-target collision and, considering that the other available data are affected by large uncertainties, a dedicated measurement could significantly contribute to improve the model, shown in the figure as the dashed black line. A measurement addressing $\Delta_{\bar{\Lambda}}$ in pHe collisions at $\sqrt{s_{\rm NN}} = 110\,\text{GeV}$ is part of my *Ph.D.* programme and is presented in Part II. Being all the experimental results displayed in top Fig. I.2.12 obtained from pp collisions, this is the first measurement of this quantity in the pHe system;

- $1 + \Delta_{IS} = 1 + (\sigma^0_{inv}(\bar{n})/\sigma^0_{inv}(\overline{p}) - 1)$ is a term accounting for the antiproton production from antineutron decays, that is not directly measured at accelerators. The Δ_{IS} term alone is a measurement of the possible isospin violation, leading to a different production cross-section of prompt antiprotons and antineutrons and its measurements as a function of the centre-of-mass energy are presented in bottom Fig. I.2.12. For $\sqrt{s} < 100\,\text{GeV}$, the available data hint for a sizeable Δ_{IS} deviation from zero. The comparison of the antiproton production cross-section with injected hydrogen or deuterium, only possible with the ongoing upgrade of the fixed-target programme discussed in Part. IV, is expected to provide a significant experimental input to decrease the current large uncertainty on the isospin violation effect.

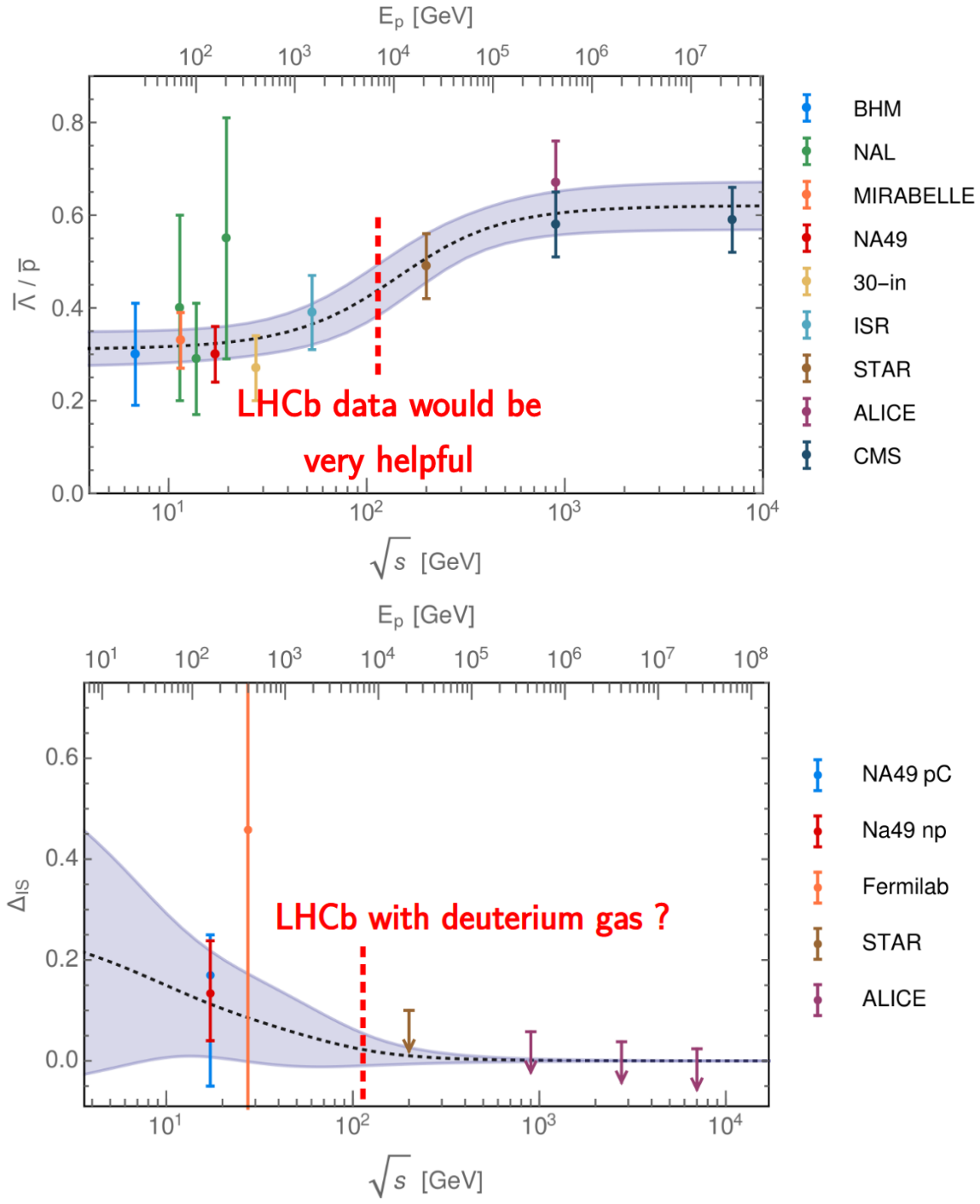


Figure I.2.12: Available data from several experiments, indicated in the legend, for the ratio between the $\overline{\Lambda}$ and the prompt $\overline{p}$ production (top) and for the difference of the production cross-section for prompt antiprotons and antineutrons (bottom) as a function of the centre-of-mass energy. The former is found to be constant at ~ 0.3 for $\sqrt{s} < 50\,\text{GeV}$ energies, at ~ 0.6 for $\sqrt{s} > 500\,\text{GeV}$; for the latter, the available data suggest a sizeable deviation from zero for $\sqrt{s} < 100\,\text{GeV}$. In both cases, according to the author himself, a measurement by LHCb would be very helpful, exactly covering the scale where the strangeness enhancement or the deviation from zero occurs, respectively. Figures presented in Ref. [65] starting from Ref. [64].

I.3 The LHCb experiment

In this chapter, the experimental framework of the research activities discussed in this thesis is introduced. Section I.3.1 describes the CERN complex, particularly focussing on the LHC accelerator where the data used for this work were collected. Section I.3.2 and I.3.3 are devoted to the description of the LHCb experiment, dealing with its detector and its data processing strategy, respectively.

I.3.1 The CERN facility

Le Conseil européen pour la recherche nucléaire (CERN) was officially born in 1954, when 12 European founding members, among which Italy, decided to start a scientific cooperation after the tragedy of the World War II. From that moment, it has grown continuously up to the largest HEP laboratory ever built. In 2017, more than 17500 scientists of 110 nationalities and coming from more than 70 countries have joined the efforts in the conception and operations of the experiments or in the analysis of the collected data. Some of the historical milestones, exemplifying CERN missions of pushing the fundamental research frontiers, developing new technologies and training scientists, are:

- the 600- MeV Synchrocyclotron came into operation in 1957, was the first accelerator at CERN and firstly allowed the $\pi^+ \to e^+\nu$ and $\pi^+ \to \pi^0 e^+\nu$ decays to be observed [67];
- in 1959, the Proto-Synchrotron (pS), with a 628 m circumference and a maximum 26 GeV beam energy, started its operations. The accelerated protons hit an internal target producing five secondary beams. In 1973, one of these beams led to the discovery of the neutral currents [68] with the Gargamelle bubble chamber. Still, as described below, the pS is operating as one of the acceleration stages for the LHC;
- in 1971, the Intersecting Storage Ring (ISR), the first hadron collider built at CERN with a 470 m circumference and a design 62 GeV centre-of-mass energy, established the evolution from fixed-target to collider machines. Also, the ISR paved the way for later accelerators with the introduction of the stochastic cooling technique and used to reduce the beam transverse dimensions and energy spread;

- in 1976, the Super Proto-Synchrotron (SpS), with a ~ 7 km circumference and a 450 GeV collision energy, was switched on. Exploiting the techniques developed at the ISR accelerator, it was used from 1981 to 1991 as a proton-antiproton collider and led to the discovery of the W and Z bosons [69]. The SpS accelerator is still in operation, providing beams to fixed-target experiments and to the LHC complex;
- in 1989, the first electron and positron beams circulated in the Large Electron-Positron (LEP) collider, in the same ~ 27 km-long circular tunnel that is now hosting LHC. With its maximum 100 GeV centre-of-mass collision energy, upgraded to 209 GeV in 1995, a detailed study of the bosons discovered at the SpS was performed, precisely determining their masses and their probabilities to decay to the allowed final states. At the end of their operations, the experiments also claimed a possible Higgs Boson evidence with a mass around 115 GeV/c^2, but the available statistics was not enough to draw a conclusion;
- in 1989, searching for an automated way to share information among scientists around the world, Tim Berners-Lee gave birth to the World Wide Web [70], which was then released with an open license by CERN;
- in 2008, the inauguration of the LHC took place. In 2012, the CMS and ATLAS collaborations jointly announced the discovery of the Higgs Boson [4, 5] and were mentioned in 2013 in the Nobel Prize assignment to Peter Higgs and Francois Englert.

I.3.1.1 The LHC accelerator

Approved by the CERN Council in 1994, the Large Hadron Collider (LHC) [71] is the accelerator currently operating at CERN and the largest ever built. It is hosted in the same tunnel of the previous LEP, with a 27 km-long circumference and an average 100 m depth underground. Differently than LEP, delivering electron-positron collisions, protons or ions, notably the lead ones, circulate into LHC and a higher collision energy is reached. Moreover, in contrast to proton-antiproton machines where antiprotons are difficult to produce, accumulate and uniformly squeeze in direction and energy, the employment of protons in both beams ensures a higher accelerator capability to deliver collisions, the luminosity. This is a key requirement to study rare processes like the decays of the Higgs boson or to search for physics beyond the Standard Model, while the benefit of using antiprotons would be minor at the LHC energies, where interactions between low-x partons, mainly gluons and sea quarks, are dominant. In the first data-taking period, the so-called Run1 between 2009 and 2012, LHC operated with a $\sqrt{s} = 7$ TeV and then a $\sqrt{s} = 8$ TeV collision energy and reached 13 TeV, 1 TeV below the design energy, with the start of Run2 in 2015. For the start of the Run3 operations, foreseen in 2022, the LHC proton beams will circulate with a starting energy of 6.8 TeV, to be increased to the design one at a later stage. In a hadronic machine, the centre-of-mass energy between the two colliding partons $\sqrt{s_{\text{coll}}}$, ranging in $[0, \sqrt{s}]$, is determined by:

$$\sqrt{s_{\text{coll}}} = c \cdot \sqrt{(p_1 + p_2)^2} = c \cdot \sqrt{(x_1 P_1 + x_2 P_2)^2} \simeq \sqrt{x_1 x_2} \cdot \sqrt{s}, \tag{I.3.1}$$

being $x_{1,2}$ the Bjorken variables introduced in Sec. I.1.2, $p_{1,2}$ ($P_{1,2}$) the parton (beam particle) quadrimomenta and denoting with the square the relativistic norm. For each

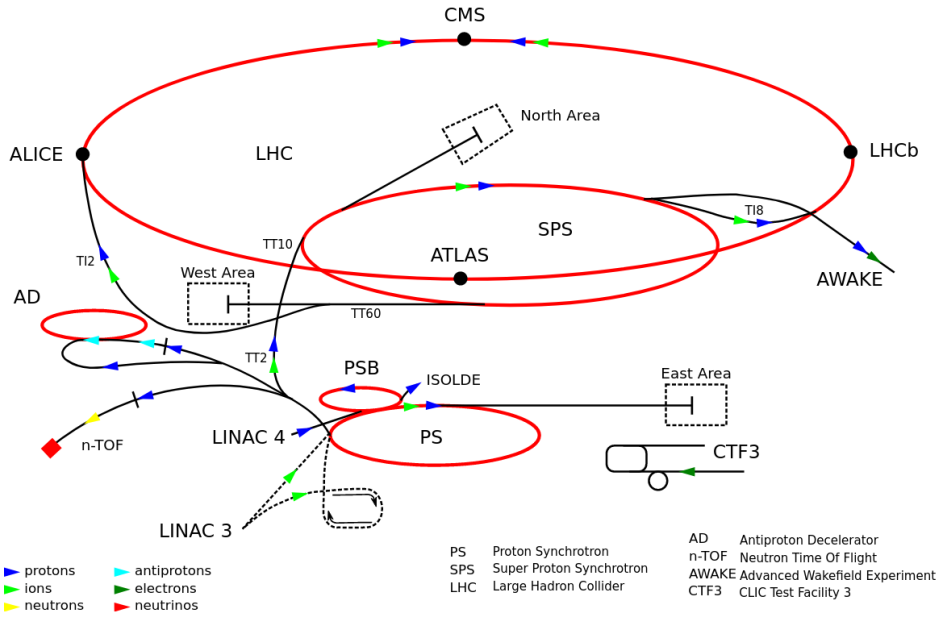


Figure I.3.1: Scheme of the CERN accelerator complex with the largest experiments indicated. Figure by Forthommel, released under CC-BY-SA 3.0.

parton-parton interaction, a different and unknown Lorentz boost along the beam axis, conventionally called z, connects the laboratory and the centre-of-mass frames. Conserved quantities can be built considering the other two directions: x, pointing towards the LHC centre and y, directed upwards and completing the right-handed orthogonal system. The transverse momentum $p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2}$ is typically used to distinguish soft and hard interactions. Alternatively, a polar system is used, with the polar and azimuthal angles defined as $\theta = atan(\sqrt{x^2 + y^2}/z)$ and $\phi = atan(y/x)$, respectively. The polar angle can also be expressed with the pseudorapidity, defined as $\eta = -ln(tan(\theta/2))$, which, in the ultrarelativistic limit, coincides with the rapidity y.

The proton beams acceleration is provided by different stages, summarised in Fig. I.3.1. Protons are stripped with an electric field from hydrogen molecules, while lead atoms are obtained by heating 2 cm of pure lead with an electric current. Particles are firstly accelerated by linear machines, called LINAC, that feed the pS and SpS accelerators where they reach a 26 and 450 GeV energy, respectively. They are then injected into the LHC and acquire the operational energy in about 20 minutes. Within the LHC accelerator, proton or lead ions are bended by a magnetic field with intensities up to 8 T produced by superconducting NbTi dipole magnets cooled with liquid helium. Other multi-pole magnets are then used to squeeze the beam dimensions at the four collision points, where the main experiments are located:

- Compact Muon Solenoid (CMS) [72] and A Toroidal LHC ApparatuS (ATLAS) [73] are general purpose experiments, being conceived to the study of the Higgs boson, to the analysis of the top quark and to search for physics beyond the SM. Since how a SM extension will manifest is not known, the geometry of the two detectors covers the largest possible fraction of the solid angle and is instrumented with a versatile detector structure. The main differences between the two experiments are

the magnetic field that is employed to reconstruct the particle trajectories, which is purely solenoidal for CMS and also toroidal for ATLAS and the installed detector technologies;

- A Large Ion Collider Experiment (ALICE) [74] is dedicated to the study of the matter under extreme temperature and pressure conditions, where Quark-Gluon Plasma (QGP) with quarks free from the confinement in hadrons is formed. The ALICE detector is optimised to study heavy-ion collisions, where the QGP formation is expected;

- Large Hadron Collider beauty (LHCb) [75] was initially conceived as a dedicated experiment to high-precision measurements of the b- and c-quark decays. As described in the following, its versatile detectors, their excellent performance and the flexibility of the data-taking strategy soon converted it into a general purpose experiment covering the forward region.

The protons circulating in the LHC are organised in bunches separated by a multiple of 25 ns and, for protons, with a typical fill of $\mathcal{O}$ (10^{11}) particles per-bunch, a 8 cm longitudinal length and $\mathcal{O}$ $(\mu\,\mathrm{m})$ transverse size at the interaction points. Considering the length of the LHC circumference, a maximum of 3600 bunches could be hosted but, in order to allow the beam to be injected and dumped, about 2800 were injected in Run2. The averaged per-bunch number of proton-proton collisions is denoted with ν. Since this also comprises elastic interactions, *i.e.* that do not involve the proton internal constituents, a more used number is μ, the average number of pp interactions that are visible in the detector. For the LHCb experiment, $\mu \simeq 0.7\nu$. At each bunch-bunch interaction, the actual number of visible pp collisions n follows the Poisson statistics $P(n) = \mu^n/n! \cdot e^{-\mu}$. The larger the μ value, the larger the pile-up, the harsher the trigger conditions.

A parameter describing the accelerator capability to deliver the collisions to the experiments is the instantaneous luminosity $\mathcal{L}$, defined as:

$$\frac{dN}{dt} = \mathcal{L}\sigma(\sqrt{s}), \tag{I.3.2}$$

being dN/dt and $\sigma(\sqrt{s})$ the rate and the cross-section of a given process, respectively, at the $\sqrt{s}$ energy scale. For all colliders with two circulating highly-relativistic beams as the LHC, this can be expressed as [76]:

$$\mathcal{L} \simeq 2c \cdot cos^2\alpha \cdot n_b N_1 N_2 \nu_{rev} \int \rho_1(x,y,z,t)\rho_2(x,y,z,t)\, dxdydzdt \equiv n_b N_1 N_2 \nu_{rev} \cdot \Omega \tag{I.3.3}$$

being α the beams crossing angle, ν_{rev} the revolution frequency, n_b and $N_{1,2}$ the number of circulating bunches and of per-bunch particles, $\rho_{1,2}$ the respective normalised particle density and Ω the beam overlap factor. Assuming the two bunch x and y spatial distribution are both Gaussian in shape and neglecting the crossing angle, this reduces to [78]:

$$\mathcal{L} \simeq n_b \nu_{rev} \frac{N_1 N_2}{4\pi\sigma_x\sigma_y}, \tag{I.3.4}$$

being $\sigma_{x,y}$ the widths of the two-dimensional Gaussian shape in the transverse plane with respect to the circulating direction. Some examples of the methods employed to measure

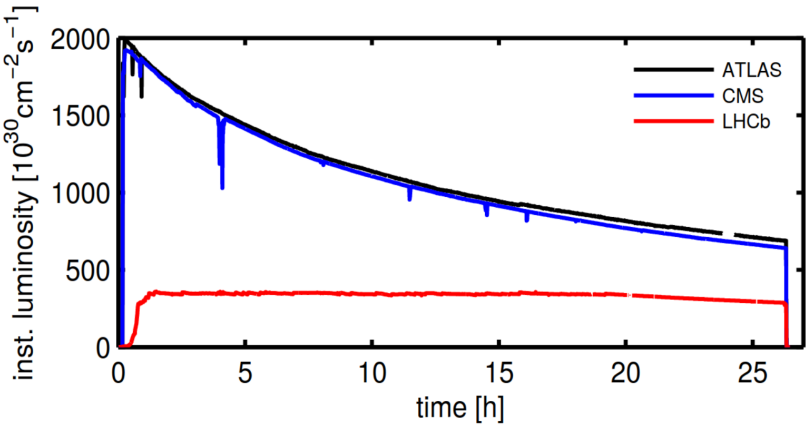


Figure I.3.2: Evolution of the instantaneous luminosity with the beam circulating time for the CMS (in blue), ATLAS (in black) and LHCb (in red) experiments. While an exponential decay is observed for general purpose experiments, a levelling to have a stable luminosity yet at a lower value is applied at LHCb. Figure from Ref. [77].

the LHC luminosity are discussed in Sec. I.4.1. The design pp target instantaneous luminosity was $10^{34}\,\mathrm{cm}^{-2}s^{-1}$, but in 2017 a value twice this limit was reached. During the nominal data-taking periods, as illustrated in Fig. I.3.2, the luminosity decreases as a function of time because of beam losses, changing $N_{1,2}$, and of several other effects increasing the $\sigma_{x,y}$ values. The total quantity of collected data is typically indicated with the integrated luminosity $\int \mathcal{L}\,dt$. The LHCb design operation luminosity in Run2 was $2 \cdot 10^{32}\,\mathrm{cm}^{-2}s^{-1}$, limited by the maximum occupancy and radiation hardness of the detectors and by the tolerable pile-up for the study of b-quark physics. In order to maximise the acquired sample size, a levelling of the LHC luminosity keeping it stable with time, represented as the red curve of Fig. I.3.2, is performed. The two beams near the LHCb interaction point are firstly displaced and then progressively re-aligned to counterbalance the luminosity decreasing with time. During Run2, the LHCb detector was operated at a maximum $4 \cdot 10^{32}\,\mathrm{cm}^{-2}s^{-1}$ luminosity, with a corresponding $\mu = 1.1$.

I.3.2 The LHCb detector

Among the LHC experiments, LHCb was initially conceived to be dedicated to heavy flavour physics. From this original scope, the LHCb physics programme has grown, as also exemplified by the results documented in this work, and also encompasses now QED measurements, hadron spectroscopy, heavy ion, central exclusive production and more. The instrumented region covered by the LHCb detector is $\theta \in [10, 300\,(250)]$ mrad in the horizontal (vertical) plane with respect to the beam axis, or, equivalently, the pseudorapidity values $\eta \in [2, 5]$. Such a choice is motivated by Fig. I.3.3, illustrating the production cross-section of $b\bar{b}$ pairs as a function of the emission angles of the two quarks with respect to the beam axis. A detector instrumenting the forward direction, collecting $\sim 24\%$ of the $b\bar{b}$ pairs as presented in red in Fig. I.3.3, significantly improves the balance between costs and benefits. Moreover, a different angular coverage with respect to other experiments ensures to complement their results and also implies an easier design, installation and

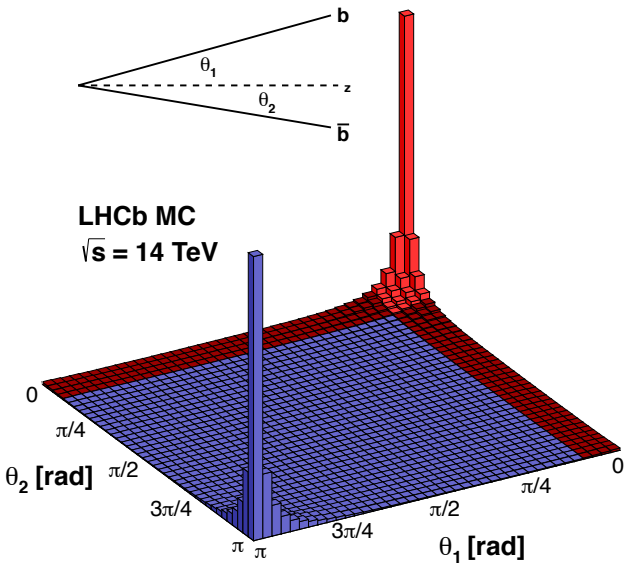


Figure I.3.3: Polar angular distribution with respect to the beam axis for the $b\bar{b}$ pairs production cross-section at a $\sqrt{s} = 14\,\text{TeV}$ energy, according to the PYTHIA generator [79]. Red squares only consider the $b\bar{b}$ pairs in the LHCb detector acceptance. Figure from Ref. [80].

maintenance for the detector supporting structures. The forward geometry is also ideal to the fixed-target configuration where, impinging the LHC beam on injected gas atoms, the produced particles are subject to a large Lorentz boost. A scheme of the spectrometer layout, with the main detectors indicated, is shown in Fig. I.3.4. Conceptually, these can be grouped into:

- the tracking system, measuring the charged particle momenta, identifying the collision vertices and allowing the particle decays to be reconstructed;
- the particle identification system, combining the momentum information with the velocity or the energy to separate the different charged particle species and indentifying some neutral particles such as neutral pions or photons.

In the following, both systems and some figures of merit for their performance are illustrated, while full information can be found in the indicated references.

I.3.2.1 The LHCb tracking system

In order to achieve a precise determination of the decaying particles mass, to measure the particles momenta and to reconstruct the positions of the pp collisions and of the decay vertices, the charged particle trajectories bending when subjected to a magnetic field is employed. At the LHCb detector, a magnetic field with a 4 Tm bending power is generated by a dipole magnet [81], while the particle trajectories are reconstructed combining the information of:

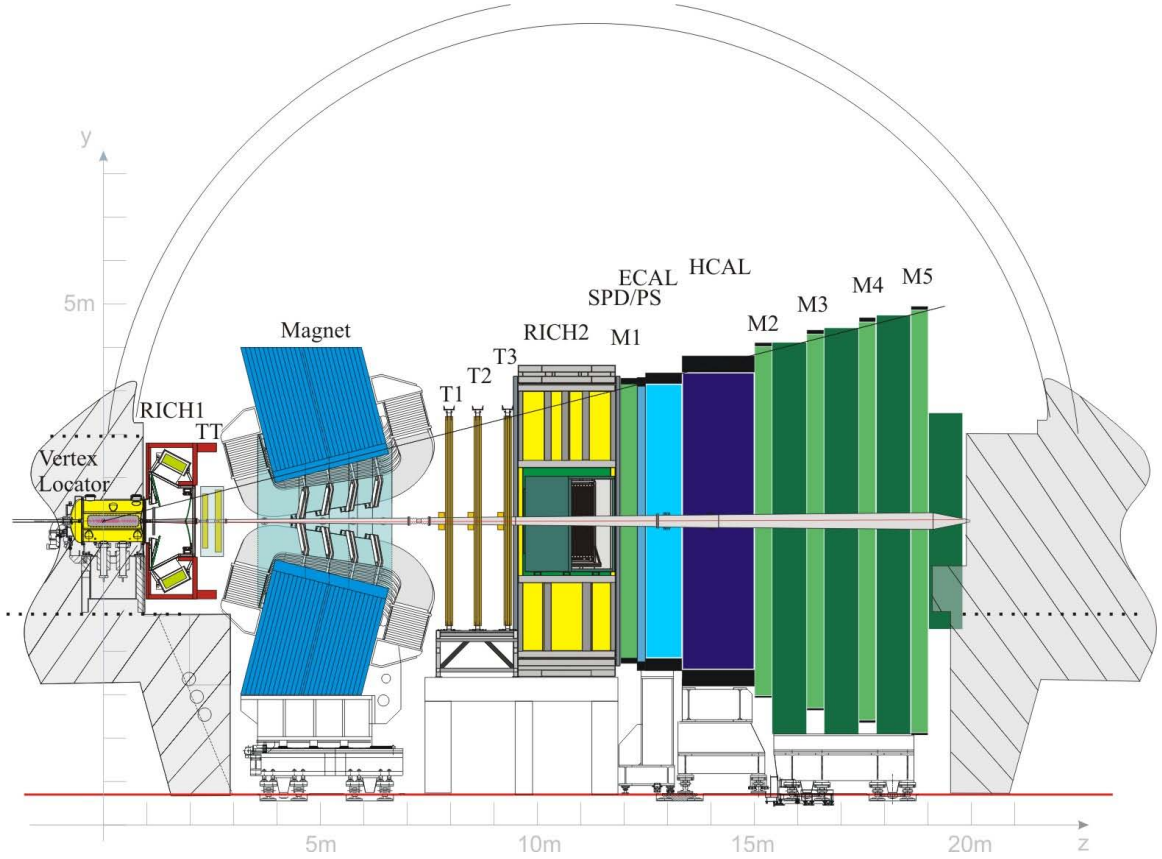


Figure I.3.4: LHCb spectrometer layout with the main detectors indicated. Figure from Ref. [75].

- the VELO detector [82], surrounding with silicon sensors the pp interaction region at a minimum distance of 8 mm;
- the Trigger Tracker (TT) [75], providing some trajectory reference coordinates upstream the magnet;
- the inner [83] and outer trackers [84, 85] (T1, T2 and T3), installed downstream the magnet.

Particles that are reconstructed with the information from the entire LHCb tracking system have a resolution on the momentum ranging from 0.5 to 1% for $5 < p < 200\,\text{GeV}/c$.

I.3.2.1.1 The VELO detector

The VErtex LOcator (VELO) detector is the first one traversed by the particles produced in the LHC collisions and is devoted to the reconstruction of track segments and to the determination of the position and composition of primary and secondary vertices. An excellent resolution in the discrimination between the two is needed, considering that a well-defined and displaced decay vertex is a typical signature for the production of a b- or c-composite particle. As illustrated in top Fig. I.3.5, the detector consists of 23 silicon tracking stations arranged in an orthogonal position with respect to the beam axis and covers in total a ~ 1 m-wide region in z. Each station is made up of two movable modules, which is required considering that the minimum sensor distance from the beam, 8.2 mm

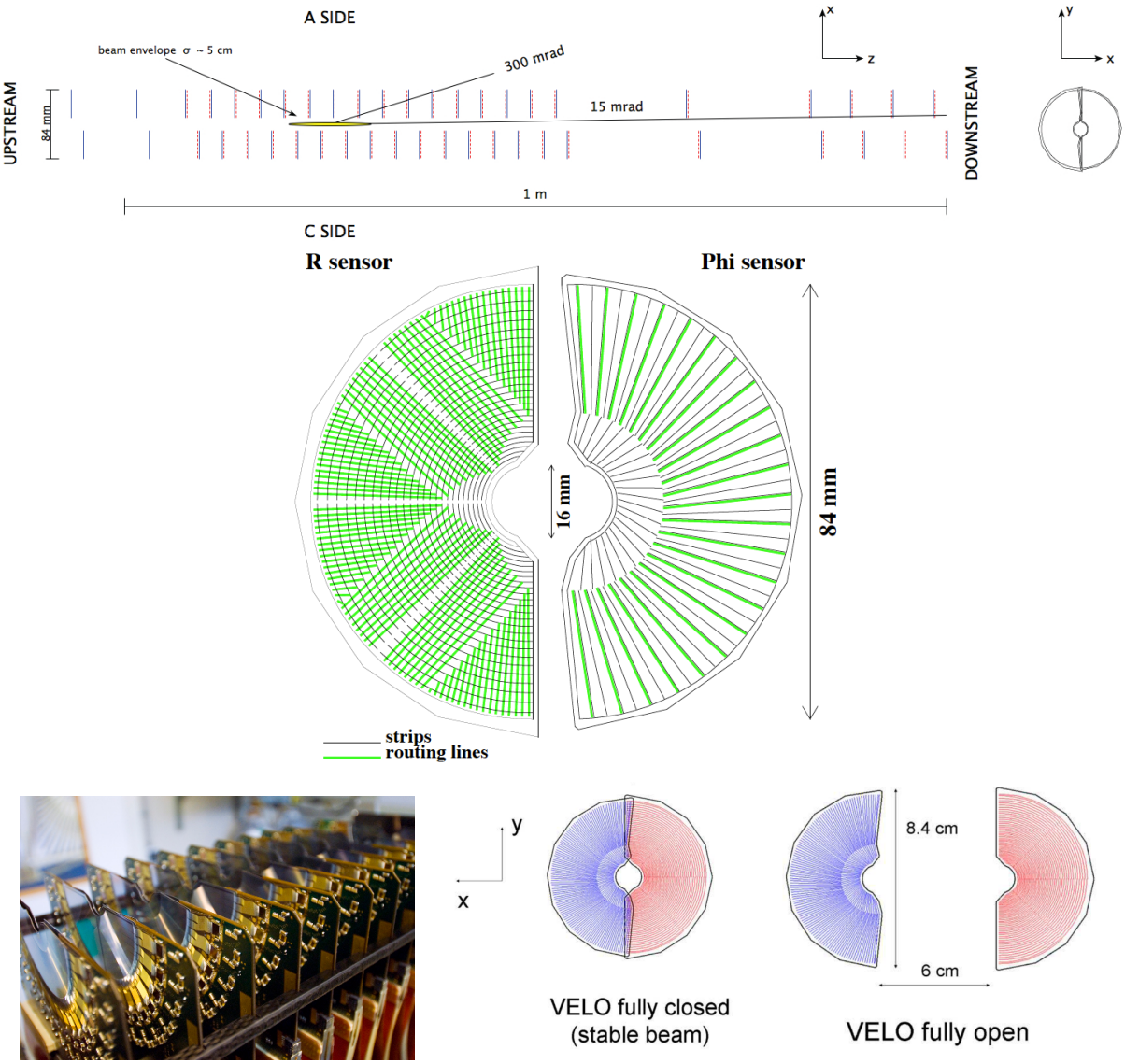


Figure I.3.5: Main characteristics of the VELO detector. The top figure shows the VELO positions along the beam axis; the central one illustrates for one sensor the strips measuring the radial and azimuthal coordinates. The bottom left figure is a picture of one half of the VELO detector, the bottom right compares the open and closed VELO positions. All figures are from Ref. [86] except the middle one, published in Ref. [75].

during data-taking, is lower than the radial aperture of the LHC beam during the injection or technical development periods. For these, a mechanical system is used to retract the two module halves by 3 cm and protect them from the radiation, as illustrated in bottom right Fig. I.3.5. When the beam safety conditions are restored, the xy beam-beam interaction point, which can vary by ± 5 mm from fill to fill, is firstly determined. Then, a closing procedure restores the VELO nominal position.

The first four VELO modules, called veto or pile-up, are instrumented with silicon strips only providing the measurement of the radial coordinate and are used to give an estimation of the number of particles produced in each bunch-bunch collision, input to the luminosity measurement. The events that, presenting a too large occupancy, would

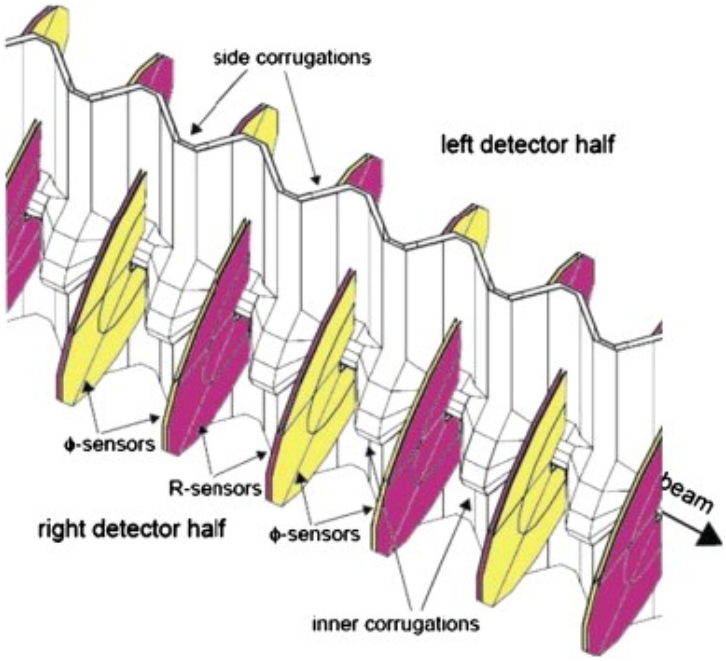


Figure I.3.6: Scheme for the RF-foil geometry with the closed VELO position, modelled according to the GEANT4 toolkit. Figure from Ref. [75].

require too much time to be reconstructed are discarded. Aside, these sensors improve the reconstruction efficiency for tracks with negative pseudorapidity, which improve the primary vertex reconstruction performance. All the other sensors, as shown in middle Fig. I.3.5, are equipped with silicon strips oriented in the radial and the azimuthal directions, with different skews to reduce the ambiguities in the track reconstruction. All the VELO modules are included in a vessel under vacuum. Its surfaces facing the beams are 300μ m-thick corrugated $AlMg_3$ sheets called RF-foils and provide:

- the separation between the primary LHC and the operational VELO vacuum;
- the shielding of the VELO sensors from the radiofrequency effects induced by the LHC circulating beams;
- the mirror wake fields guidance.

The effect on the particles impact parameter determination, which could degrade because of the scattering of the produced particles with the material, is minimised by the foils shapes and thinness and by the choice of a material with low atomic number such as the aluminium. As discussed in Sec. II.1.6.1, multiple scattering effects can still be relevant. In the closed VELO position, the RF foil geometry is designed in such a way to allow the two VELO halves to overlap, as illustrated in Fig. I.3.6.

The key requirements for the VELO detector performance are:

- a larger angular coverage than the nominal LHCb acceptance, which is achieved by the chosen sensor positions;

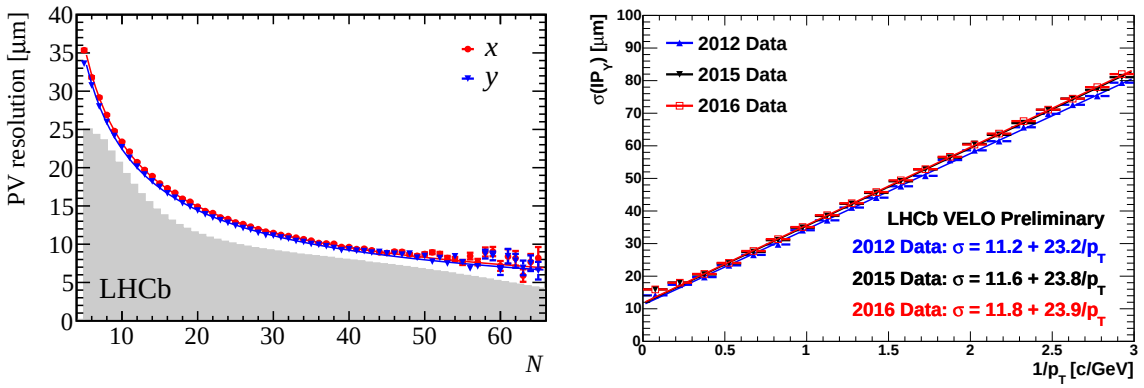


Figure I.3.7: Figures of merit for the performance of the VELO detector. The left plot presents the primary vertex resolution as a function of the number of particles originating in the vertex; the right one the impact parameter resolution measured on 2012 (in blue), 2015 (in black) and 2016 (in red) data. Figures from Refs. [87] and [88], respectively.

- an efficient reconstruction of straight-line tracks, being the magnetic field effect negligible in the VELO z region and a high rejection of fake tracks, *i.e.* those that are drawn connecting the energy deposits left by more than one particle. Both significantly affect the overall track reconstruction performance, considering that the VELO segments are then extrapolated looking for corresponding signals in the following tracking detectors;

- an excellent resolution in the determination of the position and composition of primary and secondary vertices. Most physics channels have indeed a signatures presenting large impact parameters or displaced vertices and the track-vertex assignment is a key requirement.

Two figures of merit illustrating the VELO performance are presented in Fig. I.3.7, while a detailed discussion can be found in Ref. [86]. The left plot shows the primary vertex resolution for 2012 data reconstructed with only one vertex as a function of the number of particles associated to the vertex. With the 2012 reconstruction, a primary vertex with a multiplicity of 25 tracks could be reconstructed with a resolution of $\sim 13\mu\,\mathrm{m}$ for the x and y coordinates and of $\sim 71\mu\,\mathrm{m}$ for the z. The evolution of the primary vertex reconstruction algorithm, improving the experimental performance throughout the Run2, is discussed in Sec. IV.3.2.2. The right plot presents the resolution on the y coordinate of the impact parameter, *i.e.* the distance between the linearly-extrapolated track and the reconstructed vertex position. This is affected by the multiple scattering and steeply increase with the inverse of the particle transverse momentum. Thanks to the proximity to the beam, the values, ranging from 10 to 80 $\mu\,\mathrm{m}$ for transverse momenta in $1/p_\mathrm{T} \in [0,3]\,(\,\mathrm{GeV}/c)^{-1}$, are the best among all LHC experiments.

I.3.2.1.2 The magnet and the tracking stations

The LHCb magnet, illustrated in the left plot of Fig. I.3.8, is formed by two coils that are inclined to match the LHCb detector acceptance. The magnetic field profile evolution with z is shown in the right Fig. I.3.8 compared to the positions of the tracking detectors. The bending power is about 4 Tm, with a maximum peak intensity of 1.1 T directed as

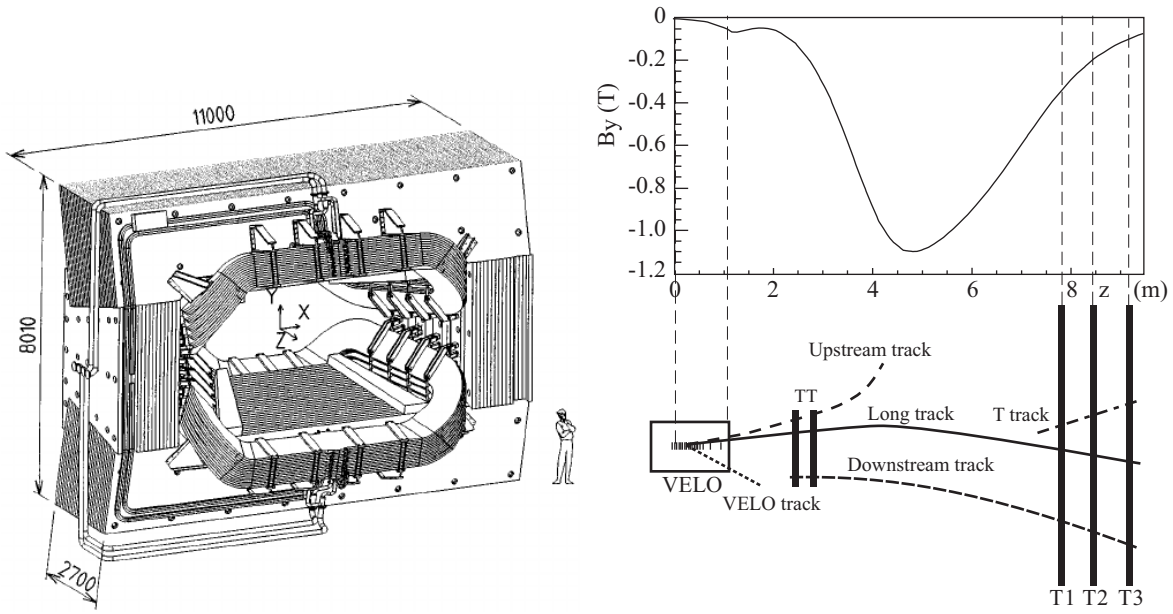


Figure I.3.8: Structure of the LHCb magnet (left) and evolution of the magnetic field strength as a function of the z coordinate (right). Figures from Refs. [75] and [87], respectively.

the y axis. The current in the magnet, and consequently the magnetic field polarity, is periodically inverted. The LHCb detector is indeed designed to be symmetric with respect to the x coordinate and this mitigates the charge-dependent detection asymmetries that could affect the measurements of charge conjugation and parity symmetry violation.

The tracking stations located upstream the magnet are called Trigger Tracker (TT) or *Tracker Turicensis*. These are placed immediately upstream the dipole magnet and set a reference for charged particle trajectories before the deflection. This improves the precision of the measurement of particle momenta, reduces the fraction of reconstructed tracks from spurious hits and allows long-living particles such as Λ or K^0_S that decay outside the VELO detector to be reconstructed [90]. By extrapolating the reconstructed track segments in the VELO to corresponding energy deposits in the TT detectors [91], a rough estimation of the particle momentum is available and its trajectory within the magnetic field can be predicted. The search for corresponding hits in the tracking stations downstream the magnet is limited to small tolerance windows, reducing the algorithm output data rate and improving the reconstruction performance. As illustrated in top Fig. I.3.9, the TT detector is made up of two stations distant $\sim 30\,\text{cm}$ with two layers each. All layers are equipped with 512 silicon strips with a $183\mu\,\text{m}$ pitch and cover the full LHCb acceptance. The first and last layer measure the x track coordinate, the other two are inclined by ± 5 degrees, forming two axes called u and v. By combining the information from all the sensors, ambiguities in the track reconstruction are removed.

Downstream the magnet, three other tracking stations with four detection planes each reconstruct the particle trajectory after its bending in the magnetic field. As shown in bottom Fig. I.3.9, two different techniques are employed. In the inner region, with a harsher environment due to the larger detector occupancy, silicon strips, presenting a high

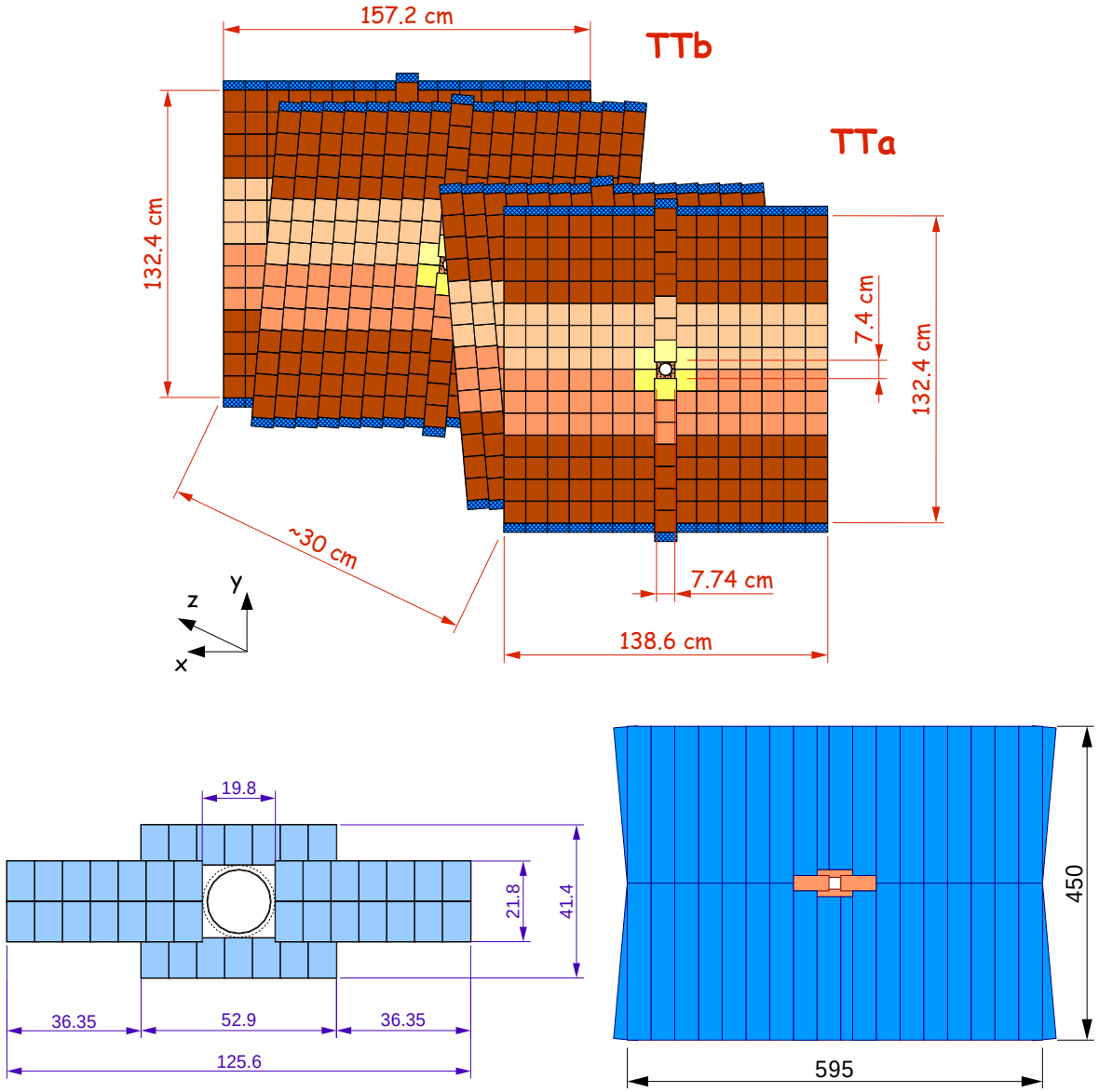


Figure I.3.9: Scheme of the TT (top) and front views of one station of the IT (bottom left) and OT (bottom right) detectors. Dimensions in the bottom plots are indicated in centimetres. Figures from Refs. [89] and [83].

radiation resistance, granularity and spatial resolution, are installed. For each station, the strips are oriented with the same *xuvx* geometry and installed with the same pitch as the TT detector. The geometrical coverage of the Inner Tracker (IT) detector is $\sim 2\%$ of the LHCb detector acceptance region, but, as a consequence of the increase of the track density at high pseudorapidities, corresponds to $\sim 20\%$ of the particle flux. At the radial value where the detector occupancy decreases to 10%, the IT detector is replaced by the Outer Tracker (OT) [84, 85], which employs drift tubes. These are 2.4 m long, have a 4.9 mm inner diameter and are filled with a $Ar/CO_2/O_2$ (70%/28.5%/1.5%) gas mixture, which guarantees a drift time below 50 ns and good ageing properties. The achieved spatial hit resolution are 50μ m and 200μ m for the IT and OT detectors, respectively.

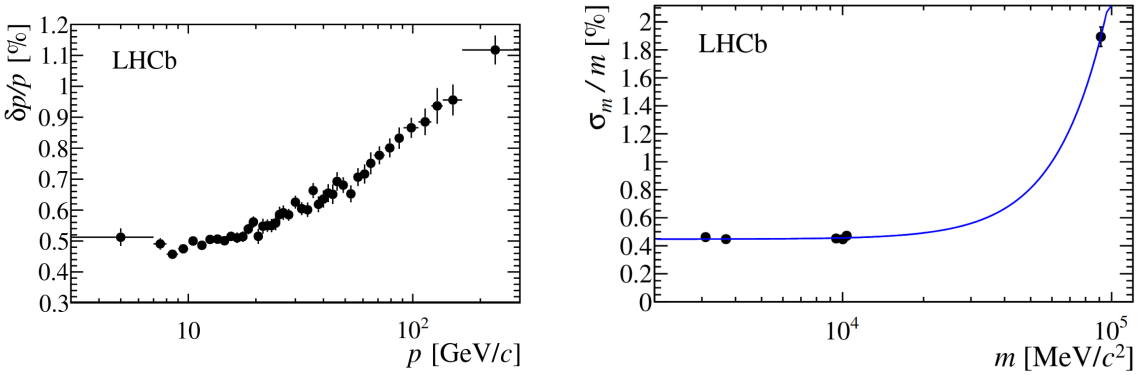


Figure I.3.10: LHCb relative resolution on the momentum (left) and on the mass for di-muon resonances (right). Figures from Ref. [87].

With the information from all detectors, tracks are reconstructed in two stages [92]. Firstly, pattern recognition tasks are executed to bundle together the energy deposits left by the particles in the detectors. Two strategies, called forward tracking and track matching, are followed. In the former, a starting track seed is reconstructed in the modules of the VELO detector. Since the magnetic field intensity is negligible at the VELO z coordinates, a straight-line model is assumed to determine the track positions and slopes. The reconstructed track segment is extrapolated to the TT first modules and corresponding energy deposits are looked for. With the available rough momentum estimation, the corresponding position in the T-stations is predicted assuming that the magnetic field effect corresponds to a kink in the particle trajectory. In the latter strategy, track seeds are reconstructed in the VELO detector and in the T-stations at the same time and are extrapolated forward and backward to the magnet plane. A compatibility criterion to match the two track segments is then set based on their distance in x and y at the magnet bending plane. Secondly, tracks are fitted in order to determine the track parameters and the particle momentum. A Kalman filter model [93] is employed in order to account for the multiple scattering effects. The resulting χ^2/ndf of the track fit is used to evaluate the track reconstruction quality. Since the two track finding strategies and other pattern recognition tasks dedicated to reconstruct the particles that are kicked out from the LHCb acceptance region are redundant, a clone killer [94] algorithm is then executed to remove tracks reconstructed twice. Also, a dedicated neural network classifier is employed to measure the probability that a track is reconstructed from energy deposits left from different particles [95]. In the following, the output of this classifier, called *GhostProb*, will be used in selection algorithms. The left Fig. I.3.10 presents the achieved momentum resolution as a function of the momentum, ranging between 0.5% and 2% for $p \in [5, 200]\,\text{GeV}/c$. For tracks originating in the same vertex, the invariant mass is calculated as:

$$Mc^2 = \sqrt{(\sum_i E_i)^2 - (\sum_i p_i c)^2} \tag{I.3.5}$$

indexing i the particles and indicating E_i and p_i their energy and momentum, respectively. The resolution on the determination of this quantity is presented in right Fig. I.3.10 for di-muon resonances and ranges from $\sim 0.5\%$ to $\sim 2\%$ for $M \in [3, 100]\,\text{GeV}/c^2$.

I.3.2.2 The LHCb particle identification system

With the available measurement of a charged particle momentum, its mass can be derived once its velocity or energy is known. The discrimination of all the particle species in the momentum range of interest ($\sim$ 2-150 GeV) demands for a redundancy of experimental strategies. Detectors devoted to Particle IDentification (PID) at the LHCb experiment are:

- the Ring Imaging CHerenkov detector (RICH) system [96], exploiting the Cherenkov effect to discriminate among charged particles, notably protons, pions and kaons;
- the calorimetric system [97], where the showers of particles produced by charged or neutral particles interacting with the detector material are absorbed to measure the incoming particle energy. The system is composed of an Electromagnetic Calorimeter (ECAL), optimised for the identification of neutral pions, photons, electrons and positrons, and a Hadronic Calorimeter (HCAL). To give a fast measurement of the detector occupancy employed for the hardware level of the trigger and to induce the start of the shower before the calorimeters, two scintillator planes called Scintillating Pad Detector (SPD) and PreShower (PS) are installed upstream ECAL.
- Muons are capable to escape HCAL and are tracked by five other dedicated stations [98] downstream the calorimeters. Matches between extrapolated tracks and energy deposits in these detectors provide thus the identification for muons.

I.3.2.2.1 The RICH system

A charged particle traversing a dielectric medium of refractive index n with a velocity $\beta > 1/n$ emits Cherenkov photons with a characteristic angle θ_c depending on its mass m and momentum p as [99]:

$$cos\theta_c = \frac{1}{n\beta} = \frac{1}{n}\sqrt{1 + \left(\frac{mc^2}{pc}\right)^2}. \tag{I.3.6}$$

In order to measure θ_c, allowing charged particles to be identified when combined with the momentum information from the tracking system, two RICH detectors located upstream and downstream the magnet and schematised in Fig. I.3.11 are used. The former is optimised for particles with $p < 60\,\text{GeV}/c$, which could exit the detector acceptance because of the magnetic deflection. It is filled with fluorobutane (C_4F_{10}, $n = 1.0014$ at a wavelength $\lambda = 400\,\text{nm}$) and covers the full LHCb acceptance. The latter is filled with tetrafluoromethane (CF_4, $n = 1.0005$) and efficiently identifies particles with $p \in [15, 110]\,\text{GeV}/c$ in the acceptance region $[15, 120]\,\text{mrad}$. During the LHC Run1, the RICH1 detector was also equipped with an aerogel radiator ($n = 1.030$) to optimise the performance for low-momentum particles. Before the start of Run2, it was removed since unable to cope with the increase in the total number of photons in the RICH detector because of the higher track multiplicity [100]. The photons emitted by the charged particles in the RICH radiators are conveyed with a system of spherical and planar mirrors outside the LHCb acceptance and, independently on their emission point, form a ring on a lattice of Hybrid PhotoDetectors (HPDs). Depending on the momentum and assuming a mass hypothesis, a test ring for each particle species is compared against the positions

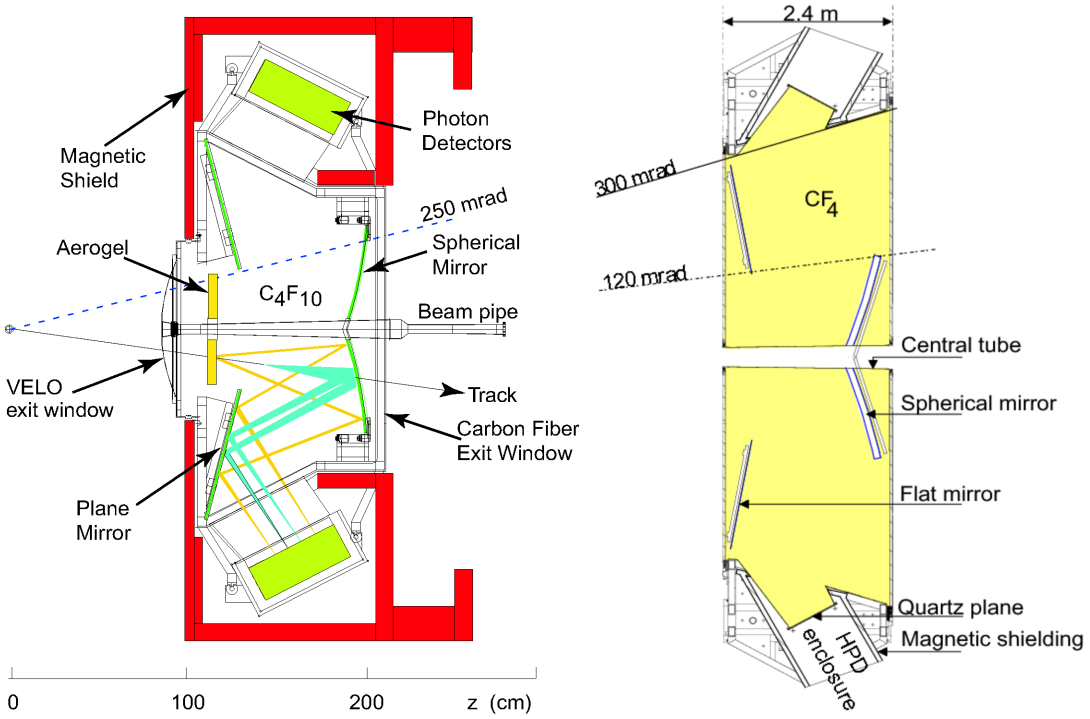


Figure I.3.11: Scheme of the RICH1 (left) and RICH2 (right) detectors. Figures from Ref. [75].

of the activated photodetectors. A RICH likelihood function is evaluated and, when combined with the information from the calorimeter and muon systems, is associated to tracks to define a particle identification classifier to be used in physics analyses. Usually, this is expressed as a log-likelihood difference between two h_1 and h_2 particle hypotheses:

$$\mathrm{DLL}_{h1,h2} = log\left(\frac{\text{h1 likelihood}}{\text{h2 likelihood}}\right). \tag{I.3.7}$$

For example, $\mathrm{DLL}_{p,\pi}$ is the log-likelihood difference between the proton and pion hypotheses, $\mathrm{DLL}_{p,K}$ between the proton and the kaon. These particle identification classifiers will be used in the prompt and non-prompt antiproton production measurements, discussed in Sec. I.4.2.2 and in Part. II, respectively, and in the machine-learning approach I developed during my *Ph.D.* programme to optimise the PID performance on fixed-target samples, described in Part III.

The theoretical evolution of the Cherenkov angle for all particle species as a function of the momentum and for all the radiators employed in the RICH system is presented in left Fig. I.3.12. The right figure presents the θ_c angle reconstruction performance in the RICH2 detector for tracks in 2010 and 2011 data reconstructed as isolated, *i.e.* whose associated Cherenkov rings do not overlap with any other one. Despite the resolution effects, an excellent proton-kaon separation up to $p \simeq 50\,\mathrm{GeV}/c$ is obtained, while the RICH2 refractive index limits the active identification of kaons and protons to $p \leq 9.3\,\mathrm{GeV}/c$ and $p \leq 17.7\,\mathrm{GeV}/c$, respectively. The main figures of merit parametrising the RICH performance are:

- the PID efficiency, defined as the fraction of particles correctly identified;
- the misidentification rate, *i.e.* the number of particles assigned to the wrong mass hypothesis;

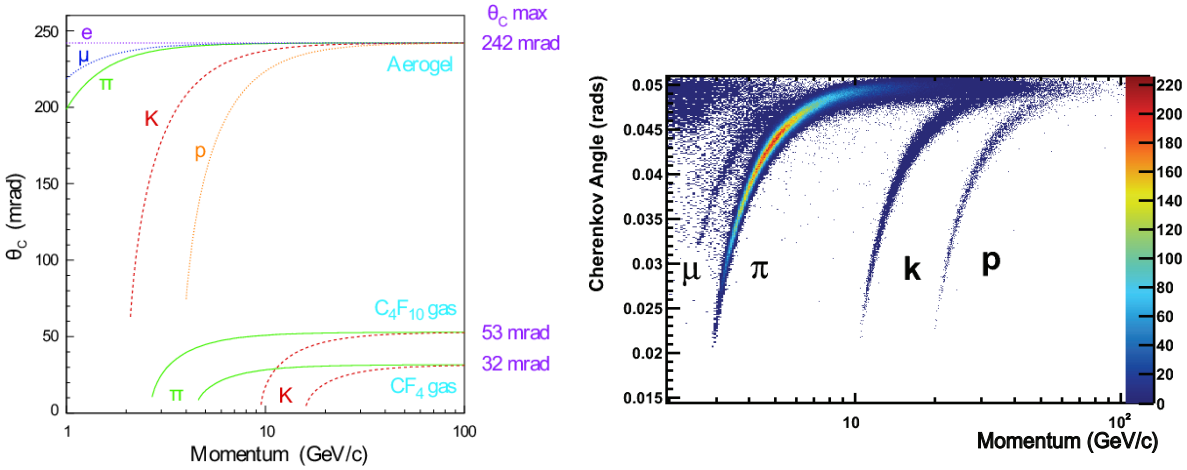


Figure I.3.12: Theoretical evolution of the Cherenkov angle θ_c with the momentum for all particle species in the radiators employed by the LHCb RICH detectors (left) and its measurement in the RICH2 detector for 2010 and 2011 data only considering reconstructed tracks producing isolated Cherenkow rings. Figures from Refs. [75] and [101], respectively.

Examples for both are presented in Fig. I.3.13 [101] with two applied thresholds on the relevant $\mathrm{DLL}_{h1,h2}$ variable. Combining the information from the two RICH detectors, protons are distinguished from kaons and pions in the full momentum range, while a steep decrease in the kaon PID efficiency is found. Integrating over the full momentum range, the kaon PID efficiency with the looser threshold is measured as $\sim 95\%$ with a pion misidentification rate of $\sim 10\%$, as $\sim 85\%$ and $\sim 3\%$ with the tighter one [101], respectively.

I.3.2.2.2 The calorimetric system

Within the calorimetric system, particles are identified through the characteristics of the showers they produce when interacting with the detector material. It is composed of two scintillator planes, the SPD and the PS separated by a lead converter, and two calorimeters, ECAL and HCAL. The system is intended to:

- distinguish among the hadron, $e^{\pm}$, π^0, γ hypotheses relying on the different induced showers, as schematised in Fig. I.3.14.;
- measure the particles transverse energy, used in the hardware data trigger;
- efficiently reconstruct π^0 and γ particles.

The PS and SPD detectors [102] are two polystyrene-based scintillating planes. They are separated by a lead converter, for a total radiation length of $6.5X_0$ ($2X_0$ for each detector and $2.5X_0$ for the lead converter). The former is used to distinguish photons from charged-particles-induced showers and facilitates the match with the information from the tracking detector; the latter discriminates electromagnetic and hadronic showers. The ECAL detector is a sampling calorimeter of *shashlik* type and is instrumented with 66 modules alternating 2 mm of lead and 4 mm of scintillator, for a total of 25 radiation lengths. To measure the longitudinal transverse shower development and to optimise

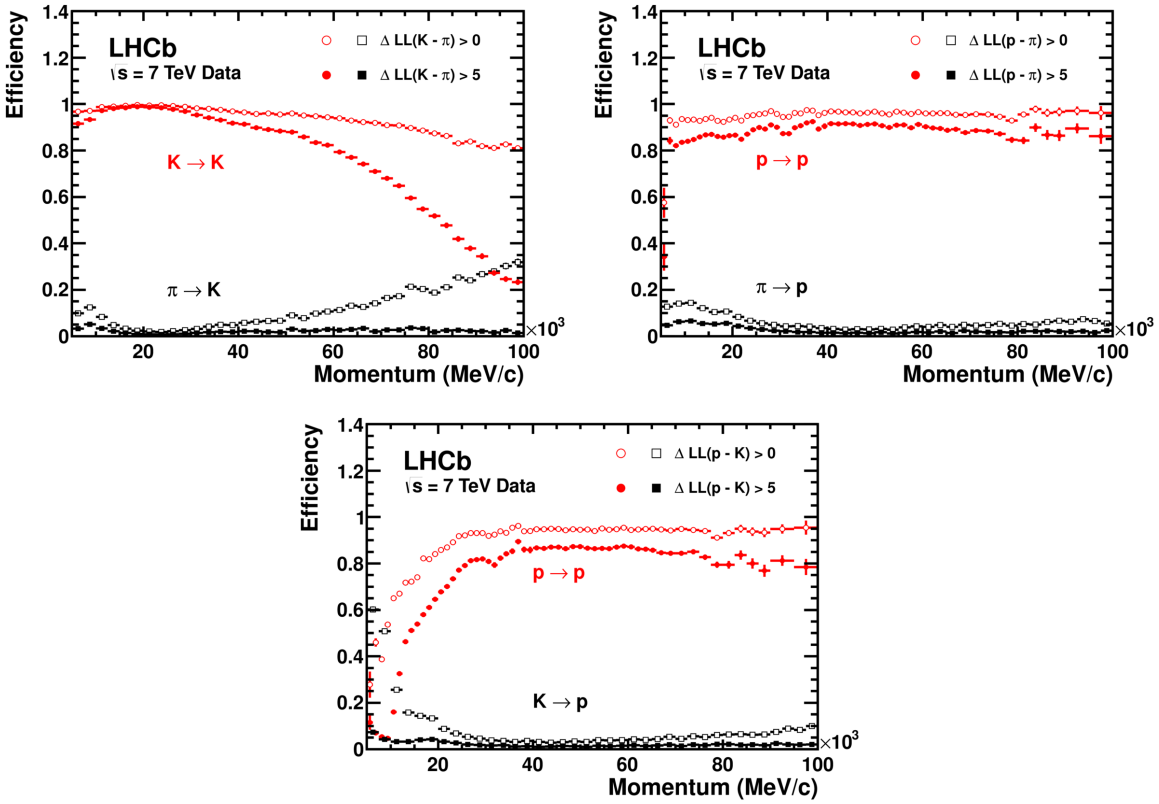


Figure I.3.13: PID efficiency and misidentification rates for the discrimination of kaons from pions (top left), protons from pions (top right) and proton from kaons (bottom) as a function of the momentum. For all plots, the efficiency and misidentification rates are compared considering two different applied thresholds on the respective $\mathrm{DLL}_{h1,h2}$. Figures from Ref. [101].

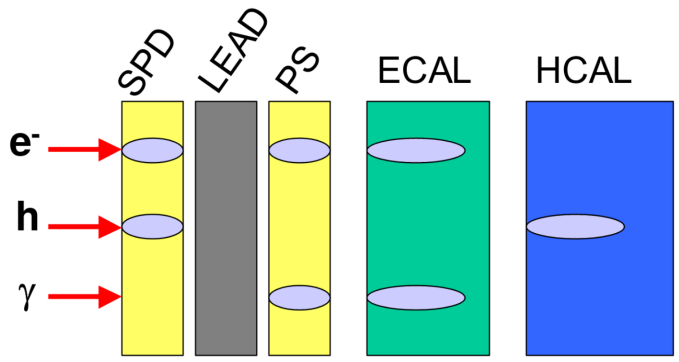


Figure I.3.14: Scheme for the shower development in the LHCb calorimetric system for different particle hypotheses. Figure from Ref. [102].

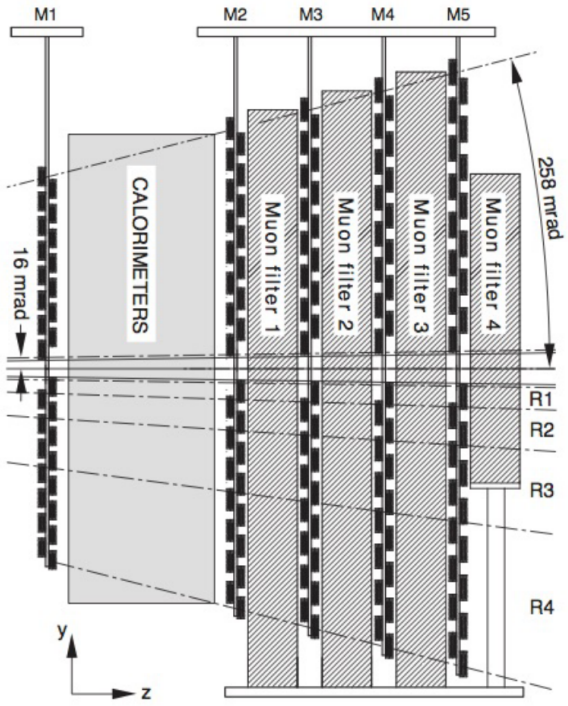


Figure I.3.15: Scheme for the muon chambers location in the LHCb detector. Figure from Ref. [75].

the separation between neighbour showers, each calorimeter layer is segmented in the transverse direction. The cell dimensions increase with the distance from the beam because of the lower occupancy in the detector and are 4.0×4.0 cm^2, 6.1×6.1 cm^2 and 12.1×12.1 cm^2 in the inner, middle and outer region, respectively. The achieved resolution on the energy measurement is

$$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E\,[\text{GeV}]}} \oplus 1\%, \tag{I.3.8}$$

indicating $\oplus$ the sum in quadrature $a \oplus b = \sqrt{a^2 + b^2}$.
Also HCAL is a sampling calorimeter but, differently than ECAL, employs alternating iron and scintillating tiles oriented in parallel to the beam axis to improve the angular resolution. A transverse segmentation is also available, with two cell sizes of 13.1×13.1 cm^2 and 26.2×26.2 cm^2 in the inner and outer regions, respectively. The total HCAL thickness is 5.6 hadronic interaction lengths, limited by the available space in the LHCb cavern. This is not enough for the full confinement of the induced hadronic shower, and the consequent achieved energy resolution is:

$$\frac{\sigma_E}{E} = \frac{69\%}{\sqrt{E\,[\text{GeV}]}} \oplus 9\%. \tag{I.3.9}$$

I.3.2.2.3 The muon system

The LHCb detector is completed by five stations for the muon identification and reconstruction [98], named from M1 to M5. The first one is located upstream the calorimeters

and is used to improve the early muon transverse momentum measurement, achieving $\sim 20\%$ resolution with the only muon stations, that is used in the hardware trigger. The other stations are installed downstream HCAL and are interleaved with 80- cm thick iron layers, corresponding to 20 interaction lengths and containing the hadronic showers. The angular acceptance of the chambers is [20,306] mrad ([16,258] mrad) in the horizontal (vertical) plane, while the minimum momentum required to muons to cross all the chambers is $p = 6\,\text{GeV}/c$. Each muon chamber is divided into four regions (R1-R4), whose dimensions increase like 1:2:4:8 to roughly accept the same fraction of particles. All regions in the chambers from 2 to 5 and for the M1 outer region install 1368 MultiWire Proportional Chambers (MWPC) detectors filled with an $Ar/CO_2/CF_4$ (40/55/5) gas mixture. In the innermost M1 region only, where the detector occupancy is too large to be sustained by MWPC, Gas Electron Multipliers (GEM) chambers, presenting a higher radiation resistance and a higher granularity and spatial resolution are employed to cope with the particle flux. The filling gas mixture for GEMs is $Ar/CO_2/CF_4$ (45:15:40). Muon identification [103] is firstly performed with a binary classifier based on the particle penetration through the calorimeters and the iron filters. A likelihood function for the muon and non-muon hypotheses is then computed, according to the match of the pattern of the energy deposits in the muon stations with the considered track. The resulting discriminating variable, called muDLL, is provided as the logarithm of the likelihood ratio between the muon and non-muon hypotheses. The muDLL function is then combined with the information from the RICH and calorimeter systems, providing $\text{DLL}_{\mu,h}$ classifiers similar to those mentioned above. The resulting muon identification efficiency and hadron misidentification probabilities depend on the momentum. With the combined likelihood approach, average values $\sim$93% with a hadron misidentification rate below 0.6% are achieved [103].

I.3.3 The LHCb data trigger and processing

Raw detector signals need to be processed [104–106] to reconstruct high-level objects like tracks, vertices and particles and to select events with interesting physics signatures. A graphical summary of the LHCb data processing sequence running within the GAUDI framework [107], with the names of the software applications dedicated to each step, is presented in Fig. I.3.16.

Early detector information is exploited online to discard the events that do not contain signatures for interesting physics signal. The trigger, as illustrated in Fig. I.3.17, is composed of three stages for both LHC Run1 and Run2. The first stage, called Level 0 (L0), is implemented on dedicated custom FPGA cards and runs synchronously with the LHC bunch-crossing rate with a fixed 4 µs latency. Based on the information from the calorimeter and muon systems, the L0 trigger reduces the data from 1 Tb/s originating from the 30-million per-second visible collisions to $\sim 1\,\text{MHz}$. Also events with no or minimum requirements have been acquired for analyses or to study experimental efficiencies with dedicated L0 trigger decisions. For example, for the pHe fixed-target sample analysed in this document, the L0SPD trigger line requiring any activity in the SPD detector to accept an event was used. NoBias-triggered events, *i.e.* selected randomly with a fixed 2 kHz output rate, were also acquired as a control sample. For the events

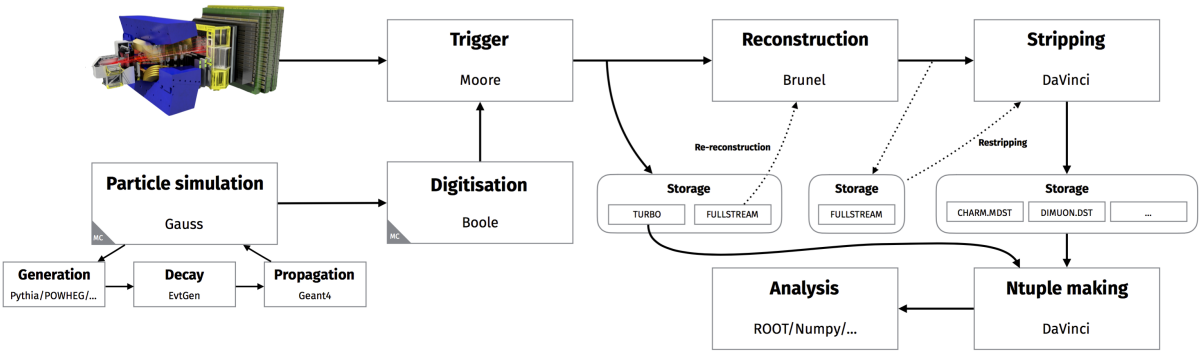


Figure I.3.16: Scheme for the LHCb data processing sequence. Figure from Ref. [108].

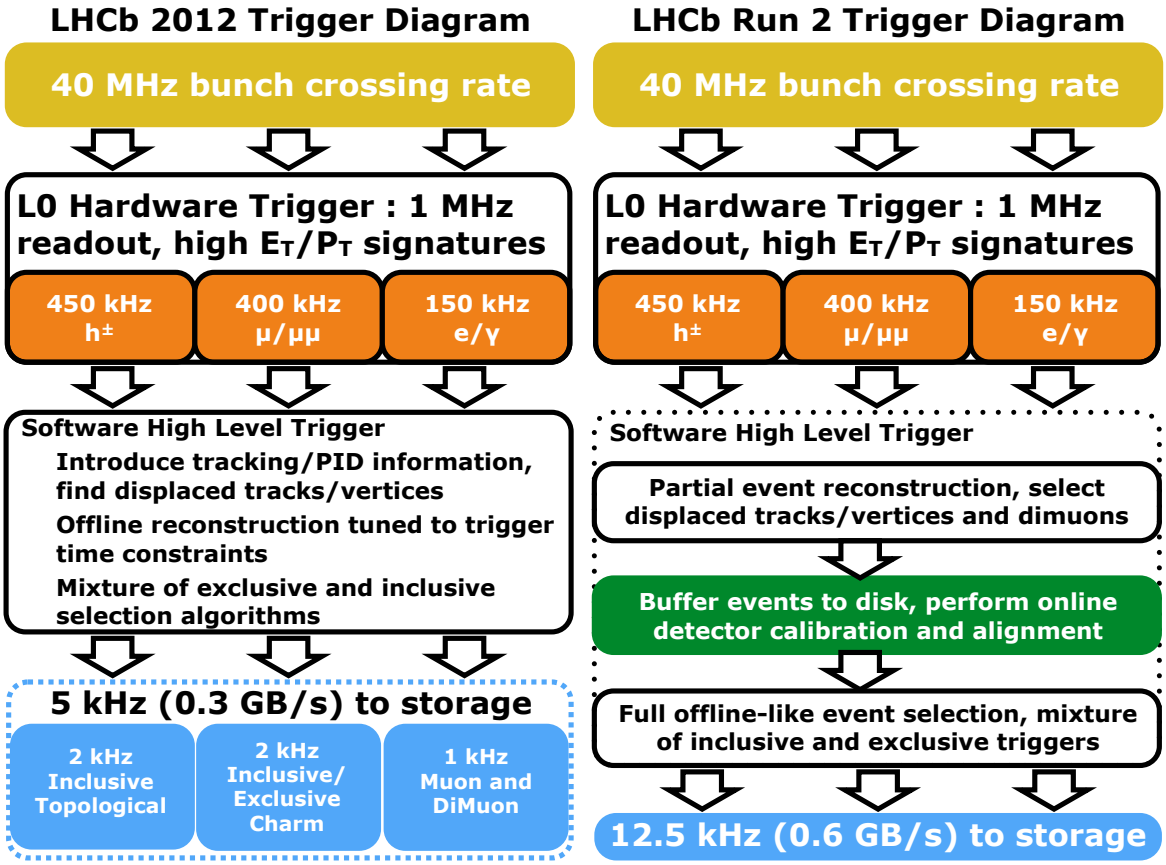


Figure I.3.17: Schemes for the LHCb data trigger software during the LHC Run1 (left) and Run2 (right) periods of data-taking. Figure from Ref. [109].

accepted by at least one of the L0 trigger selections, the information from all detectors is merged by dedicated acquisition boards and transferred to the software trigger. As presented in Fig. I.3.17, in both LHC Run1 and Run2, this comprises two stages and is controlled by the Moore application with the output data rate for the two periods indicated. In the former, HLT1, a partial event reconstruction is performed, including the finding of tracks and the muon PID algorithm. Based on these objects, inclusive trigger lines, *i.e.* selection algorithms dedicated to generic one or two-body signatures rather than a specific physics channels, are defined. As an example, to distinguish b- and c- from light-quark hadrons, which is crucial in the core physics programme of the LHCb experiment, the main signatures are the displacement of the decay vertex and the large transverse momentum of the decay products. The HLT2 trigger stage completes the event reconstruction with a more accurate fit of the reconstructed tracks and the decoding and processing of the calorimeter and RICH systems. With the complete description of the reconstructed events, both inclusive and exclusive trigger lines are acquired. During Run1 [110], the HLT2 reconstruction had a lower quality than offline, since calibration and alignment constants were not early available and the differences between the trigger and offline reconstruction made the evaluation of trigger efficiencies challenging. Moreover, given the available limited computing resources, only particles with $p_{\mathrm{T}} > 300\,\mathrm{MeV}/c$ could be reconstructed, this affecting the charm physics programme. For this reason, in the years between 2013 and 2015, the trigger strategy was redesigned to account for a large buffer interleaved between the two software trigger levels [111]. The HLT1 level still performed a partial reconstruction, but its output fed a disk with enough size to accommodate the events waiting for the determination of the detector alignment and calibration constants. These are provided in input to the HLT2 reconstruction, reproducing the offline quality.
Raw data of the events that are not discarded by the trigger are spread worldwide through the LHC computing grid [112] and reconstructed with an application called Brunel making use of the calibration and alignment constants that are progressively updated as a result of the continuous improvement in the understanding of the detector. The produced output is called full stream and contains all reconstructed objects describing the collision. This could be in principle used for physics analyses but, considering the needs of the full LHCb collaboration, is prohibitive from the computing point of view. Instead, a further data reduction step, called *stripping*, is tipically applied. This corresponds to a set of algorithms implemented offline to only select and save on disk the data that are relevant for physics analyses. The output of the *stripping* is finally available to analysts and is processed with the DaVinci application to produce an output data format that can be easily processed and analysed. With the start of the Run2 of data-taking, a more efficient paradigm for data-taking only reconstructing signal candidates that would be selected by the stripping, called Turbo [113], was conceived and developed, providing offline-quality reconstructed objects in real time. Being the pHe data sample analysed in this thesis acquired in full stream, this is not described in details here.

On the bottom left part of Fig. I.3.16, the steps required to produce simulated events are illustrated. These are:

- the event generation, *i.e.* the exploitation of theoretical models to simulate the particles produced in the collisions. Input parameters such as the LHC beam

circulation characteristics must be provided. For physics analyses of pp collisions at the LHCb detector, the PYTHIA generator [79] is mostly used, for the fixed-target programme discussed in this document the default generator is EPOS-LHC [10];

- the decay of unstable particles to their allowed final states, according to the available data for their decay probabilities. The software application employed at the LHCb experiment is EvtGen [114];

- the simulation of the generated particles propagation through the detector material, described by the GEANT4 toolkit [115, 116].

All three steps are controlled by the GAUSS application [117] and generated particles before the interaction with the detector are typically stored to be compared with those at the end of the processing sequence and to unfold the detector-induced experimental effects. The digitization phase, *i.e.* the conversion of the particles energy deposits in the detector signals, follows and is controlled by the BOOLE application. At this stage, the simulated data precisely reproduce the output of the LHCb detectors and the following steps are the same as real data. In the selection steps, simulated data are usually not removed but flagged as kept or discarded. In this way, the selection efficiencies during the data reduction process can be studied to unfold the experimental effects from the measured quantities.

I.4 The SMOG system

In this chapter, the implementation of the LHCb fixed-target programme in Run2 is discussed. Section I.4.1 introduces the system employed for the gas injection, called SMOG, illustrating its original conception to measure the LHC luminosity with the beam gas imaging technique; Section I.4.2 is focused on its use as a target to study beam-gas collisions, lists the related physics opportunities and covers the measurement of prompt antiproton production in pHe collisions at $\sqrt{s_{NN}} = 110\,GeV$.

I.4.1 The luminosity measurement with SMOG

As resulting from Eq. I.3.2, all cross-section measurements require the luminosity of the considered sample to be precisely known. The methods that have been proposed to its measurement are grouped as:

- indirect, when reconstructing and selecting a process whose cross-section is well known and obtaining the luminosity inverting Eq. I.3.2;
- direct, when determining the absolute luminosity calibration from the geometric properties of the colliding beams.

The former methods suffer from the theoretical uncertainties on the cross-section of the considered process and from the experimental efficiency of its reconstruction, the latter from systematic effects related to the description of the beam properties. A combination of techniques from both categories and information from the machine itself are thus typically needed. An example of indirect luminosity measurement employing proton-electron elastic scatterings in the LHCb fixed-target programme is described in Sec. I.4.2.2, while the two direct methods applied at the LHCb experiment [118] for the beam-beam samples are introduced in the following section.

I.4.1.1 Direct luminosity measurements

One method for the luminosity measurement was firstly proposed in 1968 at the Intersecting Rings Collider by Van der Meer. This was originally performed in the horizontal plane only and with non-bunched beams, but was then extended to two dimensions and to pulsed accelerators [119]. Still today, the Van Der Meer (VDM) scan is the most frequent

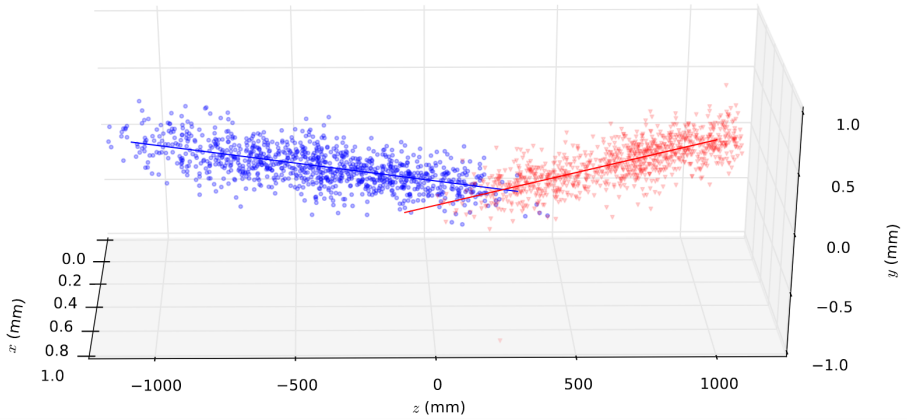


Figure I.4.1: Example of the three-dimensional reconstruction of the beam profiles through the application of the Beam Gas Imaging technique with 2013 data. Figure from Ref. [78].

technique used at the LHC. By displacing in the x and y directions one of the two beams, the rate of an arbitrary observable process with a simple experimental signature is measured, probing the particle distribution within the bunch. At the LHCb experiment, the beam position is controlled in real time with dedicated monitors and the considered counters are the recorded number of hits or tracks in some of the detectors. A precision between 2 and 3% is typically achieved, limited by systematic effects due to the luminosity variation with time, to the LHC optical system operations and to the validity of the assumption that the x and y directions can be factorised.

While all experiments at the LHC take VDM data during dedicated short periods, LHCb has the unique opportunity to decrease the uncertainty on the luminosity measurement complementing it with the Beam Gas Imaging (BGI) technique [78, 120]. As illustrated for example in Fig. I.4.1, the vertices of the collisions between protons and a noble gas injected in the accelerator vacuum are reconstructed. An image of the particles transverse distribution in the beam is obtained, providing a direct measurement of the beam overlap factor Ω of Eq. I.3.3. In contrast to VDM scans, no movements of the beams are required and the measurement can be parasitically performed during VDM-dedicated runs. On the other side, the reconstruction of beam-gas collisions drastically relies on the experiment vertex resolution, which, as discussed in in Sec. I.3.2.1.1, is excellent thanks to the forward geometry and the VELO track reconstruction performance. The uncertainty on the luminosity is decreased with the BGI technique up to $1.2 - 1.5\%$, the lowest among all LHC experiments.

I.4.1.2 The gas injection system

In order to improve the luminosity measurement with the BGI technique, the contamination of the LHC vacuum was progressively increased. When the method was firstly introduced, the pure residual gas in the LHC was used, with a pressure $\sim 10^{-9}$ mbar and a corresponding ~ 0.5 Hz HLT trigger rate per 10^{11} protons. In 2010 and 2011, the vacuum pumps serving the VELO detector volume were switched off during the BGI runs, increasing the rate by a factor ~ 4. Then, to further increase the statistics, a System for Measuring Overlap with Gas (SMOG) was conceived and installed in November 2011. It

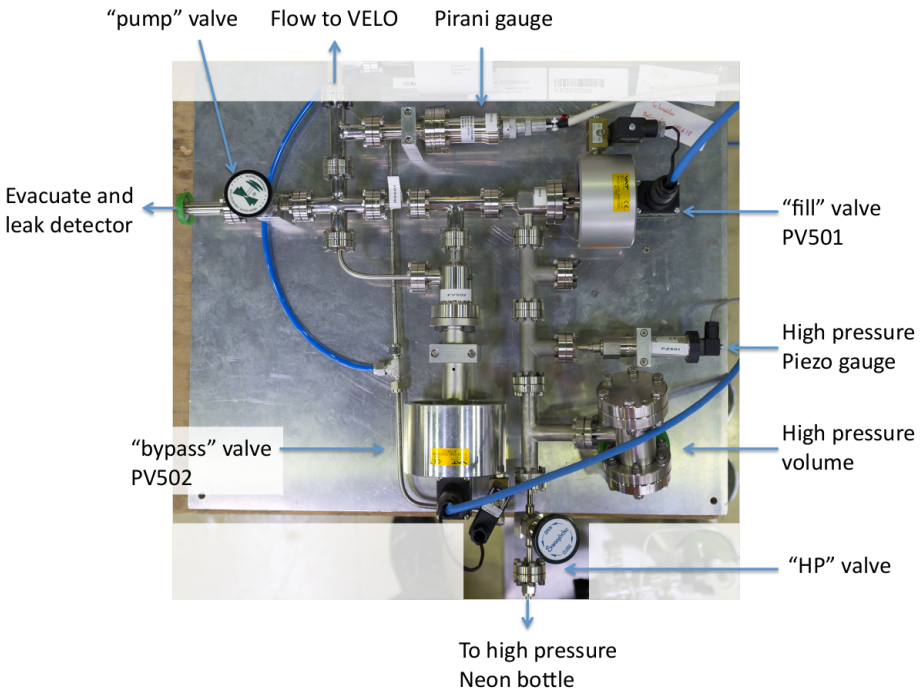


Figure I.4.2: Picture of the SMOG system main components. Figure from Ref. [78].

produced a local bump of neon gas with a $\sim 10^{-7}$ mbar pressure, two orders of magnitude higher than the LHC maximum contamination, around the LHCb nominal interaction point. The consequent recorded beam-gas events rate in 2012 was $\mathcal{O}(100\,\text{Hz})$ per 10^{11} protons. With the increased rate of collisions and improvements in the system [78], a 1.43% relative uncertainty on the luminosity measurement was reached in 2013.

A picture of the SMOG gas injection system is presented in Fig. I.4.2 with the main components indicated. The gas is refurbished by a high-pressure recipient and flows towards the VELO detector through a 4 m-long tube with $3 \cdot 10^{-5}\,\text{mbar} \cdot l/s$. During the operations, the injected pressure is monitored by four cold-cathode (Penning type) and one hot filament ionization gauge (Bayard-Alpert type) located around the VELO detector, but their absolute calibration has a $\sim 50\%$ uncertainty. The gas is extracted at $z = \pm 20\,\text{m}$ from the nominal interaction point, where the LHC vacuum pumps nearest to the LHCb interaction point are located.

I.4.2 The SMOG gaseous target

Starting from 2015, the fully-instrumented detector geometry covering the forward direction, its excellent performance in track reconstruction and particle identification and the possibility to inject noble gases into the LHC accelerator vacuum were employed in a pioneering fixed-target physics programme. As summarised in Fig. I.4.3, samples with different centre-of-mass energies and collisions systems employing both proton and lead beams and the injection of helium, argon and neon were collected during the LHC Run2, opening a rich and diverse physics programme described in Sec. I.4.2.1. During 2015,

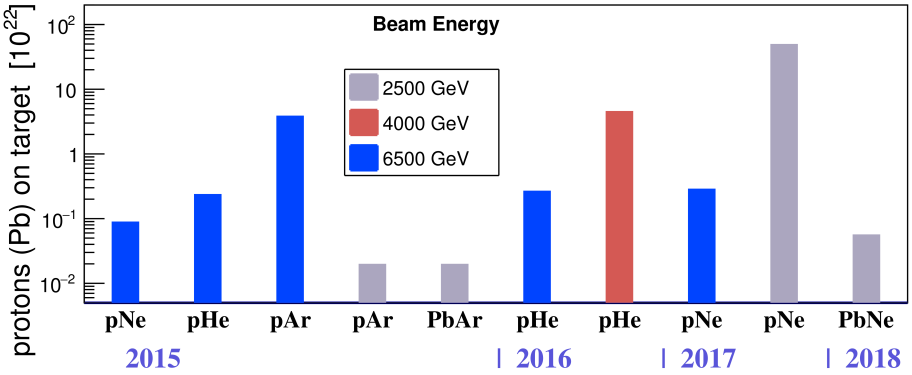


Figure I.4.3: Summary of the samples collected in Run2 with SMOG. The collision system, the statistics as protons or lead ions on target and the LHC beam energy of all samples are indicated. Figure from Ref. [121].

all three noble gases were injected in the LHC beam-pipe to test the system in different conditions. The 2016 data-taking was fully carried out with pHe collisions, to perform the first determination of the prompt antiproton production at $\sqrt{s_{\rm NN}} = 110\,\text{GeV}$, discussed in Sec. I.4.2.2. Also, the first measurement of charm production in fixed-target collisions at the LHC [122] was published based on the pHe and pAr samples at $\sqrt{s_{\rm NN}} = 87$ and $\sqrt{s_{\rm NN}} = 110\,\text{GeV}$, respectively. In both 2015 and 2016, fixed-target data were acquired in short dedicated periods and the size of the collected data samples was limited. While for antiproton and light charm state production measurements this was sub-leading with respect to the systematic uncertainties, notably due to the measurement of the luminosity of the analysed sample, an improvement in the data-taking strategy was needed to address heavier particles. At the end of 2017, neon gas was injected during a special pp run with $\sqrt{s} = 5\,\text{TeV}$ and the SMOG device was operated continuously and concurrently with the pp data-taking. The pNe collisions were only acquired for the bunches crossing the detector from the VELO side and not colliding in the LHCb detector. Even if with $\sim 10\%$ of the LHC bunches, LHCb proved the feasibility of a simultaneous operation in collider and fixed-target mode. The resulting pNe recorded number of protons of target, $\mathcal{O}$ (10^{23} nb^{-1}), is about two orders of magnitude larger than the other fixed-target samples. On the other hand, being the beam-beam and beam-gas luminous regions overlapped, the two programmes were not independent and a large contamination of the pNe sample from pp collisions caused by bunch-by-bunch proton migrations was found. Also, an additional background for pp physics due to the beam-gas programme had to be considered. Finally, in 2018, a PbNe sample was acquired with the same energy as the pNe one, in order to study nuclear effects in the two systems. The collected statistics for all samples is expressed as number of protons on target (pot), considering that, because of the lack of precise gauges for the injected gas pressure, the luminosity cannot be directly measured. For fixed-target samples, this is expressed as:

$$\mathcal{L} = \nu_{rev} \cdot n_b \cdot N_1 \cdot \sigma_t = \nu_{rev} \cdot n_b \cdot N_1 \cdot \frac{p_{gas} \Delta z}{K_B T} \tag{I.4.1}$$

being the target particles areal density σ_t expressed as a function of the gas pressure p_{gas}, temperature T and spread Δz, according to the perfect gas law. For a nominal SMOG

pressure of $p_{gas} = 2 \cdot 10^{-7}$ mbar, 10^{22} pot correspond per-metre of gas to about $5\,\text{nb}^{-1}$:

$$\int \mathcal{L}\, dt \sim 5\,\text{nb}^{-1} \times \frac{pot}{10^{22}} \times \frac{p_{gas}}{2 \cdot 10^{-7}\,\text{mbar}} \times \frac{\Delta z}{1\,\text{m}}. \tag{I.4.2}$$

I.4.2.1 The SMOG physics opportunities

The exploitation of fixed-target collisions of LHC proton or lead beams onto gas atoms opens unique opportunities. With a per-nucleon beam energy $E_N \in [0.45, 7]$ TeV, the former corresponding to the protons LHC injection and the latter to the design circulating energy, the beam-gas collision energy in the nucleon-nucleon centre-of-mass frame is:

$$\sqrt{s_{NN}} = \sqrt{2E_N M_N c^2} \in [29, 115]\,\text{GeV}. \tag{I.4.3}$$

This is an intermediate scale between experiments operating at the SpS and at the LHC or RHIC accelerator and provides an additional test bench to study the energy evolution in a region poorly-constrained from experimental data. The corresponding rapidity of the centre-of-mass frame with respect to the laboratory ranges in:

$$y_{CM} \sim arcsinh(\sqrt{E_N/2M_N\,c^2}) \in [3.8, 4.8] \tag{I.4.4}$$

which means that, being the laboratory frame LHCb instrumented region $\eta \in [2, 5]$, $y^* \in [-2.8, 0.2]$ rapidity values in the centre-of-mass frame can be accessed[1]. Being this related to the Feynman-x according to:

$$x_F \simeq \frac{2}{\sqrt{s_{NN}}}\sqrt{(Mc^2)^2 + (p_{\text{T}}c)^2} \cdot sinh(y^*) \simeq x_1 - x_2, \tag{I.4.5}$$

particles carrying a large fraction of the target nucleon momentum can be observed with SMOG. A comparison between the accessible Q^2 and target Bjorken-x values with pHe (left) and pp (right) collisions with a 6.5 TeV proton beam energy is shown in Fig. I.4.4. With SMOG, the high-x region, poorly constrained by previous experiments, is accessible at a lower energy scale than in pp. The unique kinematic coverage of SMOG is clear from Fig. I.4.5, comparing the fixed-target LHCb acceptance with those of other experiments.

The physics measurements that can be performed in the unique LHCb fixed-target kinematic regime cover different fields of interest, which can be grouped into:

- measurements of hadron production and spectra that are key inputs to tune the theoretical models used in generators. Some of these are also of interest for the CRs community, such as for the prompt antiproton production in pHe collisions, described in Sec. I.4.2.2, which significantly contributed to constrain the violation of the cross-section scaling at high energies, as discussed in Sec. I.2.2;

- constraints of the nuclear PDFs at high-x, where they are modified by the EMC effect, *i.e.* a different distribution for the parton momenta when in free or in bound states and where the contribution of a nucleon intrinsic charm could be significant. This is also relevant for neutrino astronomy, where the contribution from decays of charm hadrons produced in atmospheric showers, the most important background, are poorly constrained.

[1] In the ultra-relativistic hypothesis, the rapidity and the pseudorapidity can be assumed to be the same.

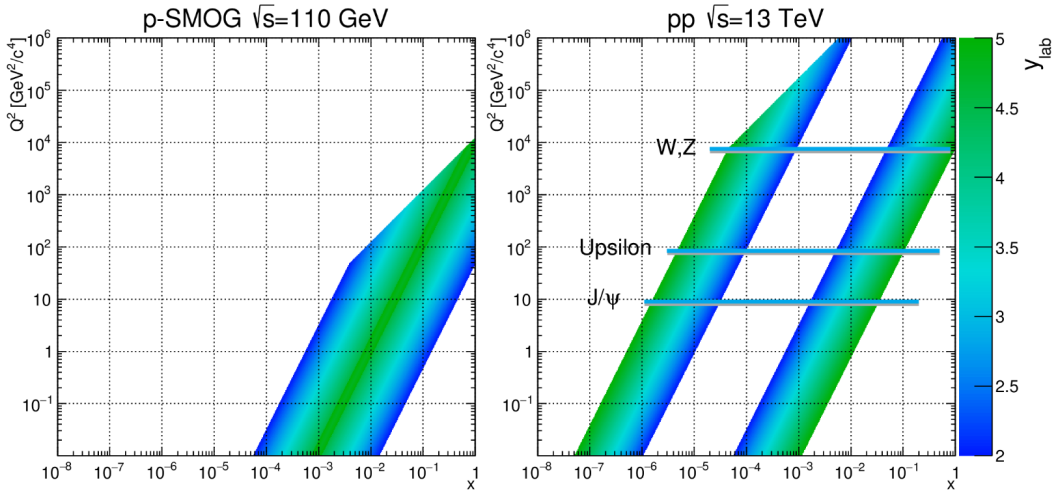


Figure I.4.4: Comparison between the kinematic region accessible to the LHCb experiment in the Q^2-target Bjorken-x plane for pHe (left) and pp (right) collisions with a 6.5 TeV proton beam energy. The high-x and low-Q^2 region, poorly constrained by previous experiments, is accessible at a lower energy scale with SMOG than in pp. Figure from Ref. [121].

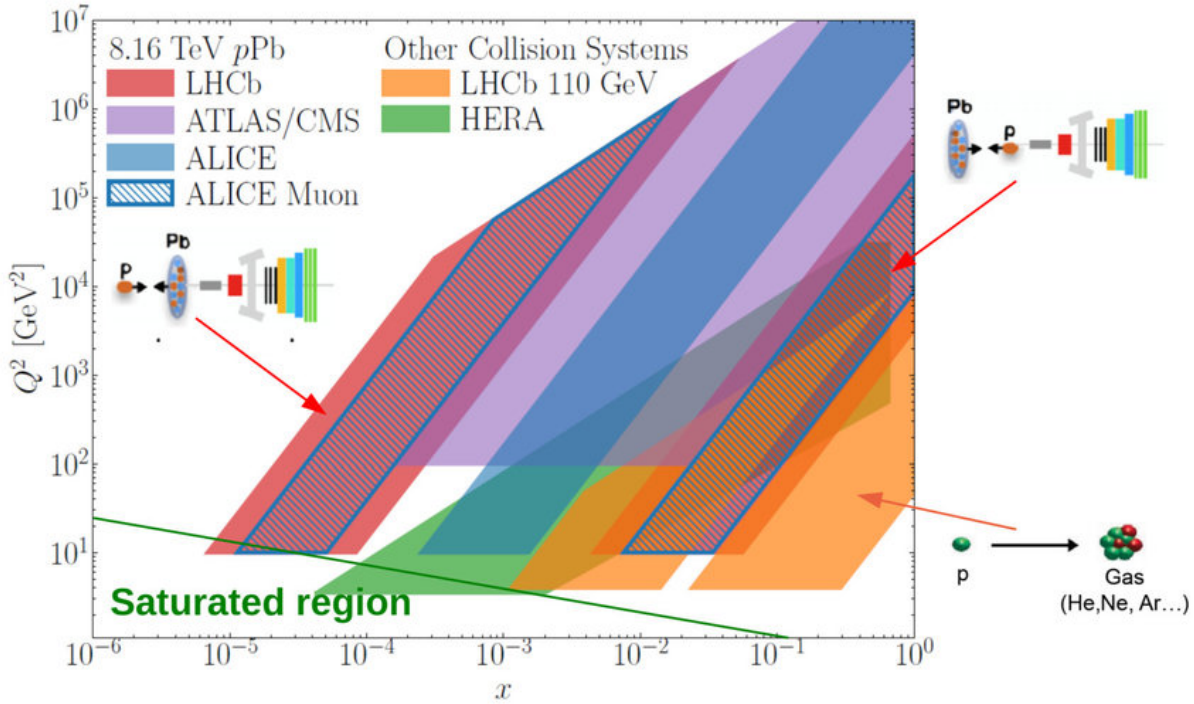


Figure I.4.5: Comparison between the Q^2-target Bjorken-x accessible regions to LHCb in its collider and fixed-target configurations with other experiments. Figure from Ref. [123].

- studies of cold nuclear-matter effects in heavy-flavour production.

A more detailed scrutiny of the physics opportunities of the LHCb fixed-target programme, especially with the upgrade of SMOG system is presented in Sec. IV.2.3.

I.4.2.2 The prompt antiproton production measurement

Among the measurements performed in Run2 with the SMOG system, the prompt antiproton production in pHe collisions at $\sqrt{s_{\rm NN}} = 110$ GeV [61] is described in this section in more details since its extension, described in Part II, has been performed during my *Ph.D.* programme on the same data sample. The cross-section is measured as:

$$\frac{d^2\sigma}{dpdp_{\rm T}}(p, p_{\rm T}) = \frac{(1 - f_{\rm bkg})^{i,j} \cdot N_{\overline{p}}^{i,j}}{\Delta p^i \Delta p_T^j \cdot \int \mathcal{L}\, dt \cdot \varepsilon^{i,j}} \tag{I.4.6}$$

being i, j indices running on the momentum-transverse momentum intervals considered for the analysis and:

- $N_{\overline{p}}$ the number of selected antiproton candidates;
- $f_{\rm bkg}$ the background fraction;
- $\Delta p \cdot \Delta p_{\rm T}$ the kinematic bin width;
- ε the total experimental efficiency.

Antiproton candidates are selected in the data sample with a 6.5 TeV proton beam colliding onto injected helium atoms, collected in 2016 and corresponding to $\sqrt{s_{\rm NN}} = 110$ GeV, within the geometric-kinematic phase-space:

$$\mathrm{PV}_z \in [-700, 100]\,\mathrm{mm} \qquad p \in [12, 110]\,\mathrm{GeV}/c \qquad p_{\rm T} \in [0.4, 4]\,\mathrm{GeV}/c. \tag{I.4.7}$$

The PV_z region was chosen requiring a good reconstruction efficiency for the antiprotons and the electrons, selected for the luminosity determination, the kinematic one imposing an efficient identification of the antiprotons in the RICH system. Data are collected with a minimum-bias trigger, *i.e.* only requiring:

- at the L0 level, any activity in the SPD detector;
- at the HLT1 level, at least one reconstructed track in the VELO detector.

Tracks associated to negatively-charged particles are required to be prompt, *i.e.* consistent with the pHe collision vertex, through the selection cut $\chi^2_{\rm IP} < 12$, being the $\chi^2_{\rm IP}$ of a track defined as the χ^2 difference of the PV fit considering or ignoring that track. This is related to the impact parameter but also takes into account the precision of its measurement and is thus less prone to variations with z in the region upstream the nominal interaction point. The $\chi^2_{\rm IP} < 12$ selection requirement selects both antiprotons directly produced at the primary vertex and from short-range decays, while the contamination from weak anti-hyperon decays, $\overline{\Lambda} \to \overline{p}\pi^+$ and $\overline{\Sigma}^- \to \overline{p}\pi^0$, is considered as a background and removed. A maximum of five tracks in the event with negative pseudorapidity is imposed. This is needed to remove in the sample a contamination from large-occupancy events caused by

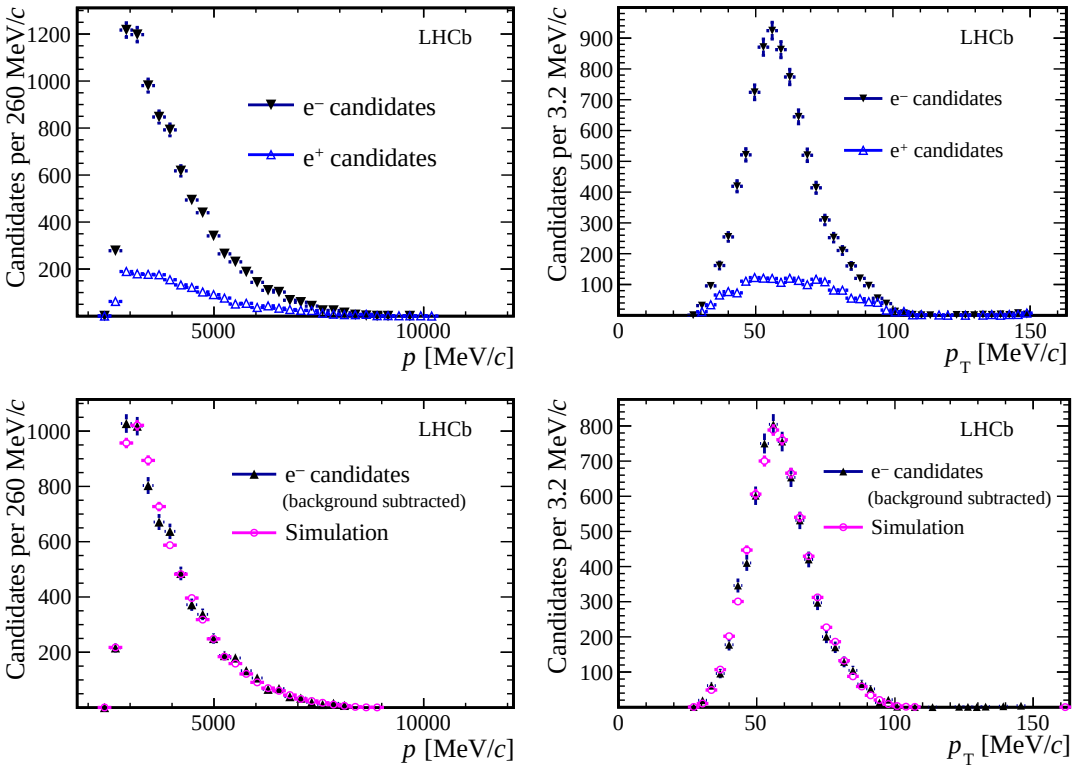


Figure I.4.6: Total electron and positron yield (top) as a function of the momentum (left) and transverse momentum (right) and comparison with the simulation after the background subtraction (bottom). Figure from Ref. [61].

high-pseudorapidity (up to $\eta \sim 9$) particles produced far upstream the nominal interaction point and interacting with the beam-pipe material. An extended shower of particles is produced and the large combinatorics induces the reconstruction of many fake tracks with negative pseudorapidity. The contamination from collisions on the LHC residual vacuum, dominated by hydrogen but also presenting beam-induced heavier nuclei, is also studied. The analysis is repeated on a sample acquired with the same pump configuration than the main one, but with no injected gas and the resulting fraction of measured events, $(0.60 \pm 0.09)\%$, is subtracted from the total.

The negatively-charged promptly-produced selected particles contain a mixture of pions, kaons and antiprotons, while all other species are negligible. To discriminate the antiprotons, a fit to the bidimensional $\mathrm{DLL}_{p,\pi}$-$\mathrm{DLL}_{p,K}$ distribution in pHe data is performed. Templates for these variables are drawn for all particle categories on the pHe simulation and on calibration channels in data and combined to be fitted with the relative abundances as free parameters. More details on the employed PID strategy are provided in Part III, where an alternative approach to PID analyses of fixed-target samples I developed during my *Ph.D.* programme employing machine-learning techniques is described.
The integrated luminosity $\int \mathcal{L}\, dt$ of the analysed pHe sample is measured considering the elastic scatterings between LHC protons and an atomic helium electron as:

$$\int \mathcal{L}\, dt = \frac{(1 - f_{bkg}) \cdot N_e}{Z_{He} \cdot \sigma_{pe} \cdot \varepsilon}, \tag{I.4.8}$$

Table I.4.1: Systematic uncertainties for the prompt antiproton production measurement in pHe collisions at $\sqrt{s_{\rm NN}} = 110$ GeV. Table from Ref. [61].

Statistical	
$\overline{p}$ yields	$0.5 - 11\%$ ($< 2\%$ for most bins)
Luminosity	$1.5 - 2.3\%$
Correlated systematic	
Luminosity	6.0%
Event and PV selection	0.3%
PV reconstruction	$0.4 - 2.9\%$
Tracking	$1.3 - 4.1\%$
Non-prompt background	$0.3 - 0.5\%$
Target purity	0.1%
PID	$3.0 - 6.0\%$
Uncorrelated systematic	
Tracking	1.0%
IP cut efficiency	1.0%
PV reconstruction	1.6%
PID	$0 - 36\%$ ($< 5\%$ for most bins)
Simulated sample size	$0.4 - 11\%$ ($< 2\%$ for most bins)

being N_e the number of reconstructed and selected electron candidates, $Z_{He} = 2$ the helium atomic number, σ_{pe} the pe^- elastic scattering cross-section and ε the total experimental efficiency. Top plots of Fig. I.4.6 show the total electron yield as a function of the momentum (left) and transverse momentum (right), reconstructed in the detector and selected with the isolation and kinematics of the candidate track. The residual background is expected to be dominated by soft diffractive events, whose observed products in the detector acceptance are symmetric in charge. The background-subtracted yield, shown in the bottom plots of Fig. I.4.6, shows an excellent agreement with the simulation. The obtained result is:

$$\int \mathcal{L}\, dt = 484 \pm 7 \pm 29\, \mu\text{b}^{-1}, \tag{I.4.9}$$

with the leading contribution to the uncertainty due to the reconstruction of the electron track in the LHCb detector. As listed in Tab. I.4.1, the luminosity measurement is the leading uncertainty on the cross-section and represents in general one of the limitations of the Run2 SMOG programme, as discussed in Part IV. The next-to-leading contribution is due to the PID strategy, suffering from the poor available statistics of reconstructed and selected calibration channels in data. The results for the antiproton production cross-section as a function of the momentum in three different $p_{\rm T}$ intervals are presented in Fig. I.4.7. The obtained uncertainties are much lower than the spread of the theoretical models and in excess with respect to EPOS-LHC, the generator used to produce the simulation.

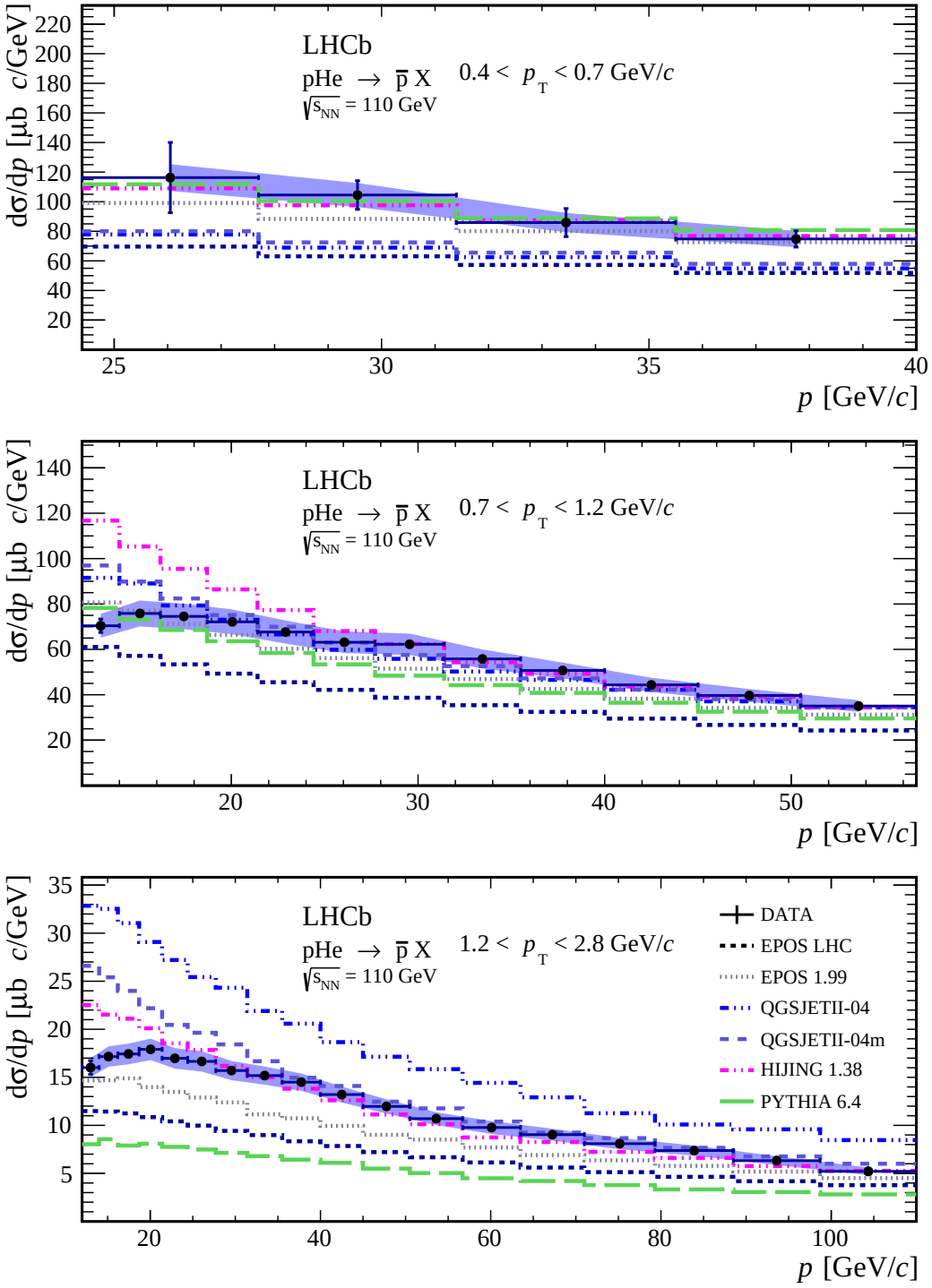


Figure I.4.7: Antiproton production cross-section in pHe as a function of the momentum in three bins of the transverse momentum. The LHCb data, in black, are compared with different theoretical models and the uncertainty are much lower than their spread. Figure from Ref. [61].

Part II

The non-prompt antiproton production measurement

Among the experimental inputs that have been provided to improve the modelling of the antiproton abundance from collisions between primary cosmic rays and the interstellar medium, LHCb in its fixed-target configuration performed the first measurement of $\sigma(p\text{He} \to \overline{p}_{\text{prompt}}X, \sqrt{s_{\text{NN}}} = 110\,\text{GeV})$. The production in anti-hyperon decays, *i.e.* $\overline{\Lambda} \to \overline{p}\pi^+$ with the $\overline{\Lambda}$ particle produced promptly or in a $\overline{\Xi}$ decay and $\overline{\Sigma}^- \to \overline{p}\pi^0$, was considered as a background in this first result. A measurement dedicated to this component has been performed during my *Ph.D.* programme, according to two complementary approaches. The former, inclusive, selects all reconstructed antiprotons produced in the pHe collisions and exploits the different impact parameter distribution between the promptly- and decay-produced. Templates for all antiproton and background categories are drawn and their combination is fitted to the collected data with the antiproton abundances as free parameters. By correcting for the respective experimental efficiencies, the detached-to-prompt antiproton ratio, where most of the systematic uncertainties cancel out, is measured and compared to the most common Monte-Carlo generators used in cosmic ray physics. A complementary approach, exclusive, is followed addressing the antiprotons produced in $\overline{\Lambda}_{\text{prompt}} \to \overline{p}\pi^+$ decays, the dominant contribution among all the detached ones. The cross-section of this process is measured and divided by the previous prompt results. The ratio between the results obtained with the inclusive and exclusive approaches is finally compared with the predictions for the fraction of detached antiprotons produced in $\overline{\Lambda}_{\text{prompt}}$ decays.

Chapter II.1 and II.2 present the inclusive and exclusive approaches and their results are compared and discussed in Chap. II.3. All the activities documented in this part have been performed as part of my *Ph.D.* programme and an article describing them is in preparation.

II.0.1 Analysis motivation

As discussed in Chapter I.2, the antiproton production cross-section is a key ingredient to the interpretation of the experimental data on the antiproton abundance in CRs and for indirect DM searches. After the measurement described in Sec. I.4.2.2, only considering the dominant prompt antiproton production, the anti-hyperon decays

$$\overline{\Lambda} \to \overline{p}\pi^+ \qquad \text{and} \qquad \overline{\Sigma}^- \to \overline{p}\pi^0, \tag{II.0.1}$$

are addressed in this measurement. No requirement on the anti-hyperon OVs is set, so they can be promptly-produced or result from cascade decays of excited states or antibaryons with higher mass ($\overline{\Xi}^+ \to \overline{\Lambda}\pi^+$, $\overline{\Xi}^0 \to \overline{\Lambda}\pi^0$ and $\overline{\Omega}^+ \to \overline{\Lambda}K^+$). The theoretical predictions for this contribution, only relying on scarce experimental data affected by large uncertainties, as previously illustrated in Fig. I.2.12, indicate an enhancement of the strangeness production that is common to other channels observed *e.g.* by the ALICE collaboration [124]. Its energy scale perfecty matches the one that can be explored by LHCb in its fixed-target configuration. The theoretical prediction for the detached-to-prompt antiproton production is presented in Fig. II.0.1 as a function of the antiproton momentum (top) and transverse momentum (bottom) within the kinematic phase-space accessible to LHCb, $p \in [12, 110]\,\text{GeV}/c$, $p_\text{T} \in [0.4, 4]\,\text{GeV}/c$, $\eta < 5$. Four generators included in the CRMC [125] package and among the most common for CR physics are considered: HIJING138 [126] (in blue), QGSJETII04 [11] (in yellow), EPOS-LHC [10] (in green) and EPOS199 [127] (in red). A prediction with the PYTHIA generator considering *pp* collisions is also added (in violet). All models predict the detached-to-prompt antiproton ratio to slightly depend on the transverse momentum, while a flatter evolution with the momentum is observed. The significant spread among the theoretical models confirms the need for a dedicated measurement of the quantity.

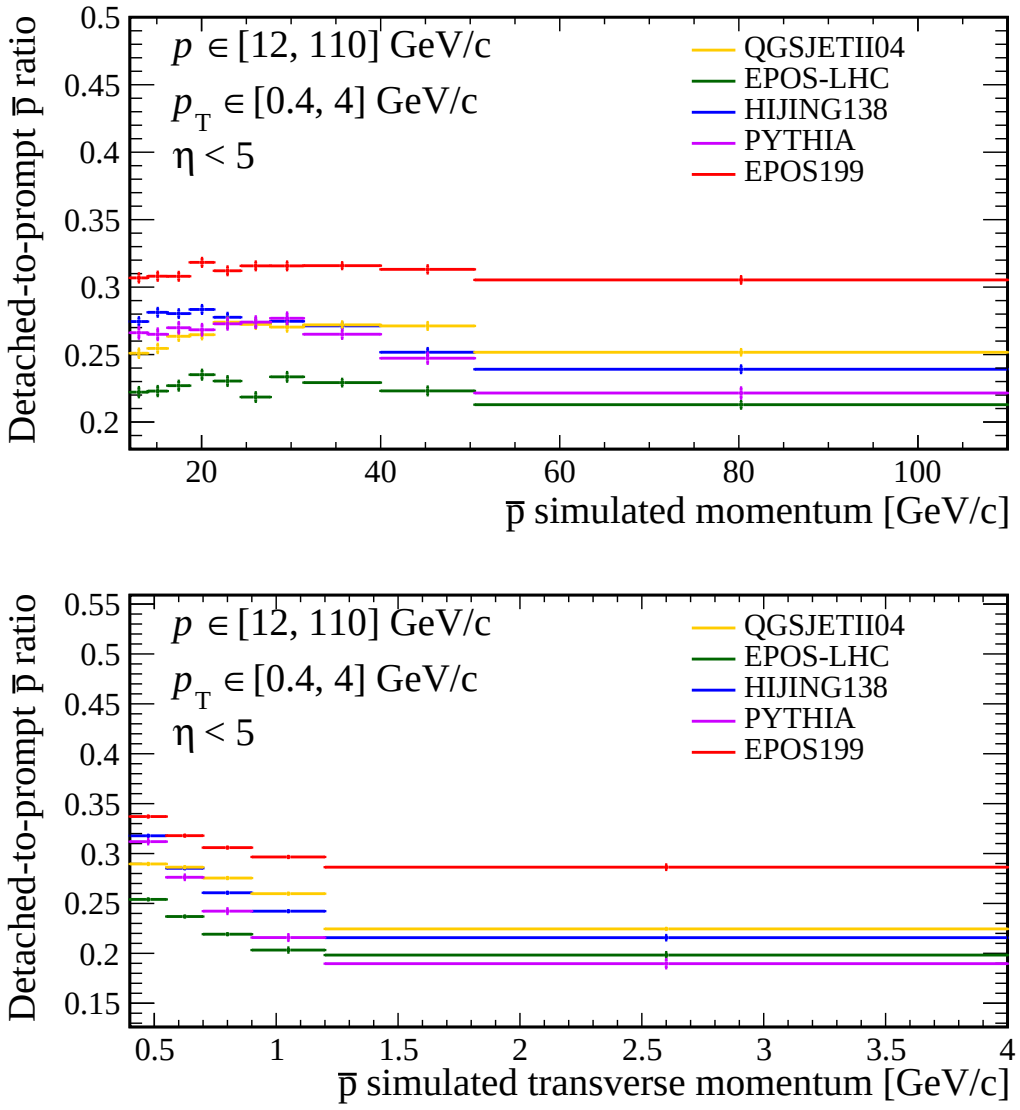


Figure II.0.1: Comparison among the predictions for the detached-to-prompt antiproton production as a function of its momentum (top) and its transverse momentum (bottom) in $p \in [12, 110]\,\mathrm{GeV}/c$, $p_{\mathrm{T}} \in [0.4, 4.]\,\mathrm{GeV}/c$, $\eta < 5$ according to the EPOS199 (red), EPOS-LHC (green), HIJING138 (blue), QGSJETII04 (yellow) and PYTHIA (in violet) generators. A large spread among the predicted values is found, which confirms the need for a dedicated measurement of the quantity.

II.1 Inclusive approach

II.1.1 Analysis strategy

The goal of the approach to the analysis presented in this chapter, where all the antiprotons produced in the pHe collisions are reconstructed in the detector and selected, is the measurement of the relative detached-to-prompt antiproton production:

$$R = \frac{\sigma(p\text{He} \to (\overline{\Lambda} \to \overline{p}\pi^+)\, X) + \sigma(p\text{He} \to (\overline{\Sigma}^- \to \overline{p}\pi^0)\, X)}{\sigma(p\text{He} \to \overline{p}_{\text{prompt}}\, X)}. \tag{II.1.1}$$

No requirements are set on the Origin Vertex (OV) of the $\overline{\Lambda}$ and $\overline{\Sigma}^-$ particles, so also the anti-hyperon excited states and cascade decays such as $\overline{\Xi}^+ \to \overline{\Lambda}\pi^+$, $\overline{\Xi}^0 \to \overline{\Lambda}\pi^0$ and $\overline{\Omega}^+ \to \overline{\Lambda}K^+$ are considered in the non-prompt component. With Eq. I.3.2:

$$R = \frac{N_{\text{det}}^{\text{Data}}/\varepsilon_{\text{det}}}{N_{\text{prompt}}^{\text{Data}}/\varepsilon_{\text{prompt}}}, \tag{II.1.2}$$

being $N_{\text{det (prompt)}}^{\text{Data}}$ the abundances of the two antiproton categories reconstructed and selected in the pHe data and $\varepsilon_{\text{det (prompt)}}$ the respective experimental efficiencies. In order to measure $N_{\text{det (prompt)}}^{\text{Data}}$, the large difference in the Impact Parameter (IP) distribution is exploited. To account for the precision of its measurement, which changes as a function of the z vertex coordinate, a related variable, called χ^2_{IP} and calculated for each track as the χ^2 difference for the fit to the Primary Vertex (PV) when considering or ignoring the track contribution, is used. Its simulated distribution for all the antiproton categories, defined on the simulated sample as explained in Sec. II.1.2, is presented in Fig. II.1.1 and three distinct peaks corresponding to prompt, detached and antiprotons produced in interactions with the detector material can be distinguished. A selection algorithm, discussed in Sec. II.1.3, is implemented to enhance the antiproton statistics over the other particle species and the χ^2_{IP} distribution for each antiproton and background category in the selected sample is modelled to obtain template distributions, as described in Sec. II.1.4. These are combined in a unique model with the antiproton relative contributions, $N_{\text{det (prompt)}}^{\text{Data}}$, determined with a fit procedure to the pHe data. Tails in the χ^2_{IP} reconstruction, whose inaccurate modelling in the simulation could induce a sizeable systematic bias, are checked for the dominant prompt antiprotons with a data-driven technique described in Sec. II.1.6.1. The efficiencies of the production of the simulated samples, the reconstruction of the antiprotons

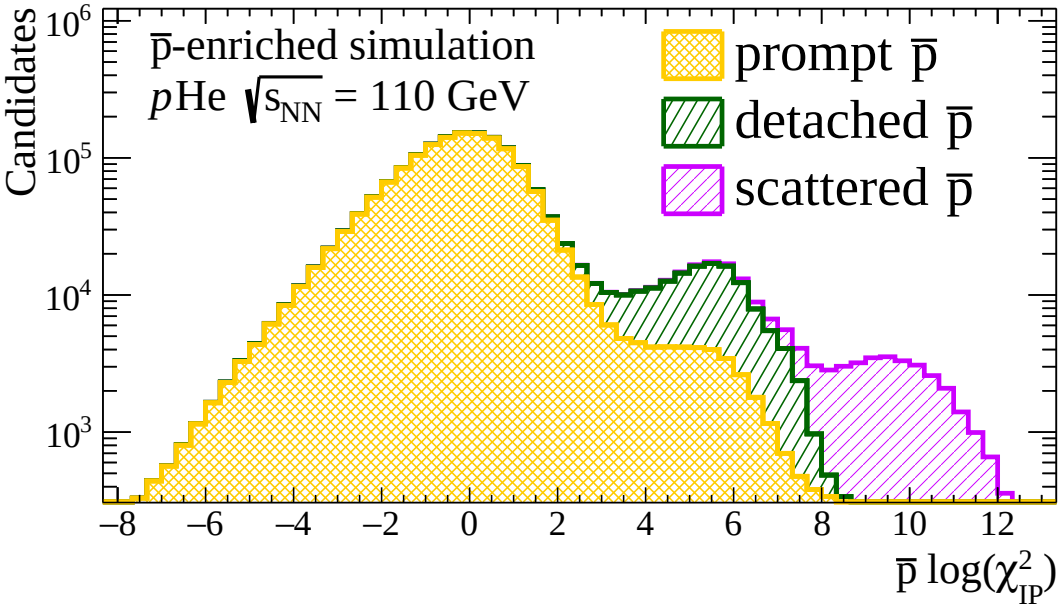


Figure II.1.1: Simulated $\log(\chi^2_{\text{IP}})$ distribution for antiprotons originating promptly (in orange), in anti-hyperon decays (in green) and in hadronic interactions with the detector material (in magenta). Three well distinct peaks appear.

in the detector and the selection algorithm that is applied to enhance their statistics are determined with the simulated sample and scrutinised in Sec. II.1.5. By measuring a ratio between two antiproton yields, most of the uncertainty is expected to cancel out. The residual contributions due to the antiproton identification or reconstruction are measured in Secs. II.1.6.4 and II.1.6.5, respectively. The simulated relative production yields of $\bar{\Lambda}$, $\bar{\Sigma}^-$ and $\bar{\Xi}$ baryons are checked with data-driven techniques in Secs. II.1.6.2 and II.1.6.3, respectively. The R results, obtained with Eq. II.1.2, are finally presented in Sec. II.1.7 and are the first experimental input in the pHe system for the $\Delta_{\bar{\Lambda}}$ quantity introduced in Sec. I.2.2.

II.1.2 Simulated samples

Two simulated samples have been produced for this analysis. Since the generator employed to simulate pp data at LHCb, Pythia [79], does not support proton-nucleus interaction, EPOS-LHC, specialised on soft hadronic production and nuclear effects, is used instead. It does not generate the production of charm quarks, whose antiproton production can be demonstrated to be negligible. The charm production in fixed-target pHe data has been measured by LHCb as $\sigma(c\bar{c}) = (144 \pm 12 \pm 4)\,\mu$b/nucleon [122]. Considering the fragmentation $f(c \to X \to \bar{p}) \leq 15\%$ (*e.g.* $f(c \to \Lambda_c^+) \simeq (5.7 \pm 0.6 \pm 0.3)\%$ [128]) and that in the prompt antiproton production measurement [61] the total inelastic cross-section is measured as $\sigma_{\text{inel}}^{\text{LHCb}}(p\text{He}, \sqrt{s_{\text{NN}}} = 110\,\text{GeV}) \simeq 127$ mb, the antiproton multiplicity from charm is expected to be below:

$$\frac{144 \cdot 10^{-3} \cdot 4 \cdot 0.15}{127} = 7 \cdot 10^{-4}, \tag{II.1.3}$$

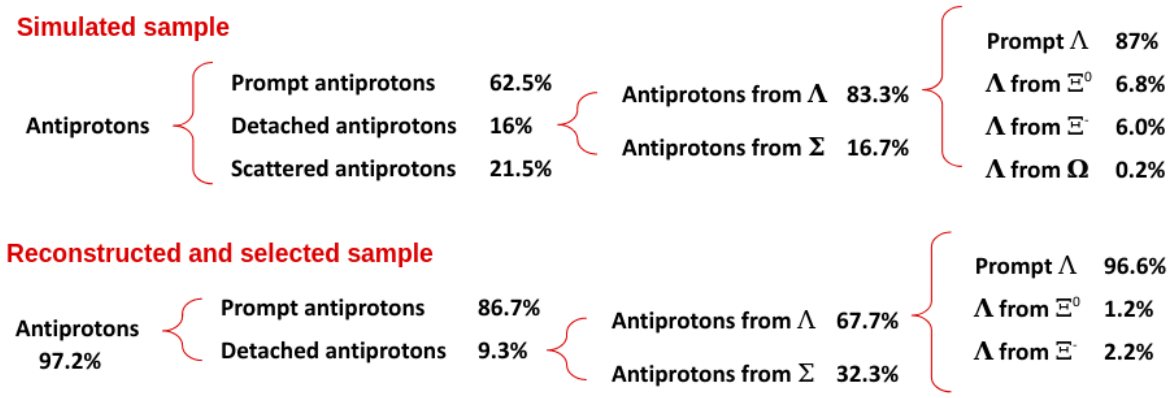


Figure II.1.2: Abundances of the antiproton categories in the simulated sample before (top) and after (bottom) the reconstruction of tracks in the detector and the application of the selection algorithms.

to be compared with that measured for prompt antiprotons of ~ 0.3.
The other simulation steps are instead common to pp data: the particle decays are described by EvtGen [114] and their interactions with the detector material by the Geant4 toolkit [115, 116]. The gas pressure evolution with the z coordinate, and consequently the pHe collision vertices, is assumed to be uniform in $z \in [-1000, 300]$ mm, which is a reasonable approximation considering that the gas is extracted by two pumps located in $z = \pm 20$ m. The two samples considered for the analysis are:

- a minimum-bias one, common to the prompt antiproton production measurement, with 11 million simulated pHe collisions. A realistic description of the antiproton and background category relative abundances to the real data is provided, but the statistics is found to be insufficient to draw smooth template $\chi^2_{\rm IP}$ distributions for detached and scattered antiprotons;

- a sample with 18 million pHe collisions enriched with high-energy antiprotons (denoted as $\overline{p}$-enriched in the following). Generated minimum-bias events are accepted only when in the collision at least one antiproton with $p_{\rm T} > 300\,{\rm MeV}/c$ and $\eta \in [2.0, 5.5]$ is produced. The resulting antiproton statistics is such that a precise description of the template distributions for all antiproton is achieved, but the antiproton-to-background ratio is biased.

A sample with the same selections as the latter but with the disabled propagation of the generated particles through the detector material is added. By comparing the generated and the simulated sample, provided that the number of considered events is the same, the detector effect can be separately studied. For all the described samples, the antiproton categories are defined as:

- prompt, if the distance between the simulated pHe vertex and the antiproton OV is below 10μ m. Both antiprotons directly originating in the hadronization of the quarks produced in the pHe collision and in a heavier particle short-range decay, ruled by the strong or the electromagnetic interaction, are thus included in this category;

Table II.1.1: Quark composition, mass and lifetime values of the anti-hyperon species considered in the analysis. Values from Ref. [6].

Particle	$\overline{\Lambda}$	$\overline{\Sigma}^-$	$\overline{\Xi}^+$	$\overline{\Xi}^0$	$\overline{\Omega}^+$
Valence quarks	$\overline{u}\overline{s}\overline{d}$	$\overline{u}\overline{u}\overline{s}$	$\overline{u}\overline{s}\overline{s}$	$\overline{d}\overline{s}\overline{s}$	$\overline{s}\overline{s}\overline{s}$
Mass [MeV/c^2]	1115.683	1189.37	1321.71	1314.9	1672.5
Lifetime [10^{-10} s]	2.63	0.802	1.639	2.90	0.821

Table II.1.2: Selection algorithm applied to the pHe data and simulated sample to enhance the antiproton statistics over the other charged particle species.

Event	$N_{SPD} > 0$, $N_{VELO} > 0$, $N_{\text{trk}}(\eta < 0) < 5$
PV	$z \in [-700, 100]$ mm $\sqrt{[(x\text{-}(0.81\,\text{mm}\text{-}z \cdot 490))^2 + (y\text{-}(\text{-}0.22\,\text{mm}\text{-}z \cdot 45))^2]} < 0.4$ mm
$\overline{p}$	$\chi^2/\text{ndf} < 4$, $GhostProb < 0.15$ $p \in [12, 100]$ GeV/c, $p_{\text{T}} \in [0.4, 4]$ GeV/c, $\eta < 5$ $\text{DLL}_{p,\pi} > 20$, $\text{DLL}_{p,K} > 10$

- detached, when produced in anti-hyperon decays, which, being ruled by the weak interaction, mostly happen farther than 10μm from the PV. The parent particle, a $\overline{\Lambda}$ or a $\overline{\Sigma}^-$, is required to be consistent with the PV or, for the former, to be produced in a weak decay of a $\overline{\Xi}^+$, a $\overline{\Xi}^0$ or a $\overline{\Omega}$ particle. This excludes anti-hyperons produced in an interaction with the detector material;
- scattered, for all other antiprotons produced in secondary interactions with the detector material.

The abundances of all the antiproton categories in the simulated sample before the event reconstruction in the detector and the selection to enhance the antiproton statistics are presented in the top part of Fig. II.1.2. Prompt antiprotons are dominant, while the detached ones mostly come from promptly-produced $\overline{\Lambda}$ decays, followed by the $\overline{\Sigma}^-$ ones and from the double-weak $\overline{\Xi}$ decays. The contribution from $\overline{\Omega}$ decays, below 1% even before the event reconstruction in the detector, will be neglected in the following. The quark composition, the mass and the lifetime values for all the anti-hyperons are listed in Tab. II.1.1.

II.1.3 Antiproton selection

In this section, the selection algorithm summarised in Tab. II.1.2 and discriminating the antiprotons from the other charged-particle species is described. The event-level preselection requires:

- the minimum-bias trigger selections discussed in Sec. I.4.2.2, *i.e.* any activity in the SPD detector ($N_{SPD} > 0$) at the hardware level and at least one reconstructed track in the VELO ($N_{VELO} > 0$) for the HLT1;

Table II.1.3: Percent abundances of the antiproton and background particles in the simulated sample after the application of the selection requirements listed in Tab. II.1.2 except those on the DLL variables.

Particle	Abundance [%]	Particle	Abundance [%]	Particle	Abundance [%]
π^-	64.0	$\overline{p}$	23.7	K^-	9.5
Ghost	1.6	e^-	0.98	$\overline{\Sigma}^-$	0.14

- one reconstructed PV with $z \in [-700, 100]$ mm and consistent with the beam geometry. Such a z range was chosen in the prompt antiproton analysis imposing a good experimental efficiency for the reconstruction of the antiprotons and of the electrons, used to measure the pHe luminosity. Also this measurement, where the focus is on antiprotons originating after the flight of a long-living particle, profits from accepting vertices far upstream the nominal interaction point;
- less than 5 reconstructed backward tracks in the event ($N_{\text{trk}}(\eta < 0) < 5$). As discussed in Sec. I.4.2.2, this is needed to remove the contamination from high-occupancy events produced by a particle interacting with the beam-pipe material. The cut is expected to be independent on the antiproton category and its efficiency cancels in the ratio R.

After the pre-selection, the negatively-charged particles reconstructed with good quality combining the information from all tracking detectors are filtered with thresholds on the track fit χ^2/ndf and on the probability they are reconstructed combining energy deposits of several particles (*GhostProb*), both introduced in Sec. I.3.2.1.2. In order to impose the direct identification of the antiprotons in the RICH detectors, the same kinematic phase-space as the prompt antiproton production measurement:

$$p \in [12, 100]\,\text{GeV}/c \qquad p_{\text{T}} \in [0.4, 4]\,\text{GeV}/c \qquad \eta < 5 \tag{II.1.4}$$

is set. The selected sample contains a mixture of antiprotons and other particles, as indicated in Tab. II.1.3. The dominant contamination comes from pions and kaons, while the large difference in production rate and the kinematic selection mostly suppress electrons and muons. A residual contamination from ghost tracks, *i.e.* reconstructed from energy deposits from different particles, is found.
In order to distinguish the antiprotons from the other charged particles, the PID variables introduced in Sec. I.3.2.2.1 are considered. Their distribution for all selected particles (top) and for the antiprotons only (bottom) in the minimum-bias simulated sample are presented in Fig. II.1.3. The dashed red lines indicate the cuts that are applied and included in Tab. II.1.2. The fit to the χ^2_{IP} that is used to count prompt and detached antiprotons is indeed not sensitive to the particle species and the contribution from MISIDentified (MISID) particles could bias the results. Assuming the residual background contamination is small with respect to the antiprotons, it can be fixed in the fit to its predicted abundance in the minimum-bias simulated sample. The relative fraction of MISID particles is studied in intervals of the antiproton momentum and transverse momentum with limits:

$$p\text{: } [12,\ 14,\ 16.2,\ 18.7,\ 21.4,\ 24.4,\ 27.7,\ 31.4,\ 40,\ 50.5,\ 110]\ \text{GeV}/c$$
$$p_{\text{T}}\text{: } [0.4,\ 0.55,\ 0.7,\ 0.9,\ 1.2,\ 4]\ \text{GeV}/c.$$

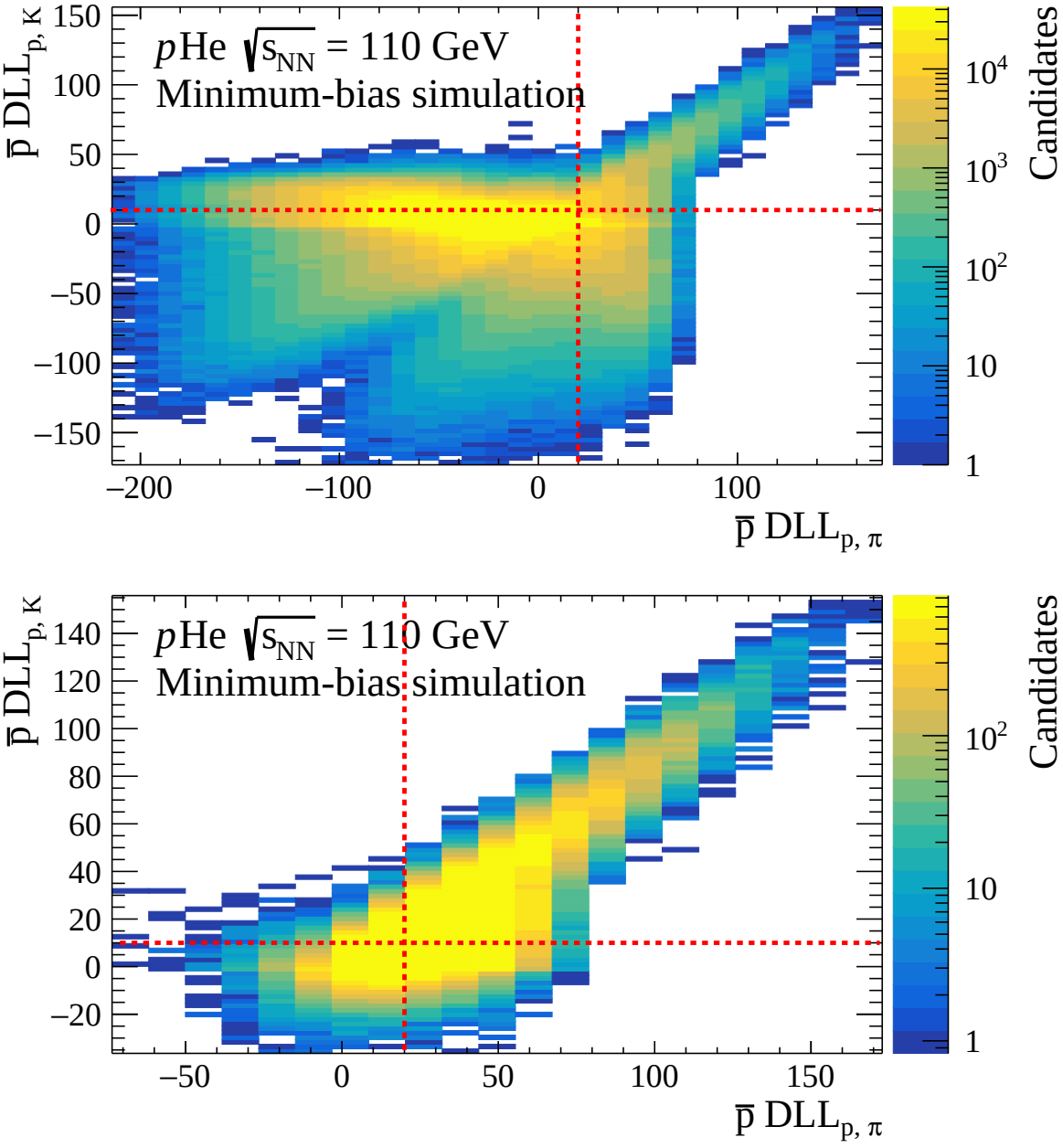


Figure II.1.3: Distributions of the $\mathrm{DLL}_{p,\pi}$ and $\mathrm{DLL}_{p,K}$ variables in the minimum-bias simulated sample for all particles selected by the cuts in Tab. II.1.2 except the PID ones (top) and for only those with a $\overline{p}$ simulated identity (bottom). Red dashed lines show the PID thresholds applied in the selection algorithm.

All bins where the fraction of MISID particles in the minimum-bias simulated sample exceeds 20% of the total selected particles are removed from the analysis. This validates the fit procedure, but, on the other hand, a sizeable PID inefficiency is introduced. However, this is expected to be poorly dependent on the antiproton $\chi^2_{\rm IP}$ and, consequently, the related systematic uncertainty, discussed in Sec. II.1.6.4, is expected to be small.
After the reconstruction of the pHe collision in the detector and the application of the selection algorithm, the abundances of the antiproton categories are indicated in the bottom part of Fig. II.1.2. By comparing with the values in the simulation (on top), the main dependence of the reconstruction and selection efficiencies from the OV_z emerges. Prompt antiprotons, whose OV_z coincides by definition with the selected PV_z range, are more efficiently reconstructed. Among the detached ones, antiprotons coming from promptly-produced $\overline{\Lambda}$ decays are dominant and the other follows with proportions that depend on the relative production yield and on the lifetime of their (grand)parent antibaryon, listed in Tab. II.1.1.

II.1.4 Template fit

After the application of the selection algorithm, the $\chi^2_{\rm IP}$ distributions for all antiproton and MISID categories are organised in histograms with 75 bins for $\overline{p}\,\chi^2_{\rm IP} \in [-10, 15]$ in each momentum-transverse momentum interval defined above. Two strategies are considered to obtain templates for the $\chi^2_{\rm IP}$ distributions:

- a histogram model, where the binned distribution is directly used (via the RooHistPdf class in RooFit [129]);
- a parametric model, where it is parametrised with the sum of Gaussian functions.

The former makes no assumptions on the shape of the $\chi^2_{\rm IP}$ distribution, but statistical fluctuations in poorly-populated kinematic interval are large and, as a consequence, the template can present spike structures. The latter, ensuring a continuous description of the $\chi^2_{\rm IP}$ binned distribution, mitigates this effect and is used to estimate a systematic uncertainty due to the choice of the parametrization. For the detached antiprotons, despite the possibility to separately define on simulation antiprotons from $\overline{\Lambda}$, $\overline{\Sigma}^-$ and $\overline{\Xi}$ decays, a unique template is built according to the predicted relative abundances. The $\chi^2_{\rm IP}$ region covered by the three categories is indeed similar, making the distinction in the fit to the pHe data challenging. On the opposite, the higher statistics makes the templates drawing more precise. The relative contribution from the three anti-hyperon species, constrained by the isospin symmetry, is fixed to the EPOS-LHC prediction. Inaccuracies for the simulated $\overline{\Lambda}$-to-$\overline{\Sigma}^-$ and $\overline{\Lambda}$-to-$\overline{\Xi}$ ratios, possibly causing a distortion of the template for detached antiprotons, are studied in Secs. II.1.6.2 and II.1.6.3, respectively.

The templates for the antiproton and background categories integrating over the whole considered kinematic phase-space are firstly considered. For the parametric model, the number of Gaussian functions for each template is determined with an iterative procedure. The fit model is the sum of an increasing number of Gaussian functions, limited to a maximum of 15 for prompt antiprotons, which are dominant and to 9 for all other categories (except the electrons, always modelled as a simple Gaussian). For

each category and in each kinematic interval, the models are applied to the $\chi^2_{\rm IP}$ binned distributions and their parameters are determined in a maximum-likelihood fit. The likelihood function is:

$$\mathcal{L} = \prod_{i=1}^{N_{\rm bins}} \frac{\lambda_i^{n_i}}{n_i!} e^{-\lambda_i} \tag{II.1.5}$$

$$\lambda_i = \sum_{j=1}^{N_{\rm Gauss}} \frac{w_j}{\sqrt{2\pi}\sigma_j} \cdot exp\Big[- \frac{(x_i - \mu_j)^2}{2\sigma_j^2} \Big],$$

being x_i the $\chi^2_{\rm IP}$ centre value for the bin i, $N_{\rm bins}$ and $N_{\rm Gauss}$ the number of bins in the histogram and of considered Gaussian functions and μ_j, σ_j and w_j the mean value, the width and the relative weight in the sum of the Gaussian j. The parameters are initialised as:

- for all μ_j values, according to a uniform distribution in $[-10, 15]$;
- for all σ_j values, equal to 3;
- for all w_j values, to $1/N_{Gauss}$.

Then, their values minimising $-2ln(\mathcal{L})$ are iteratively determined. Since the $\chi^2_{\rm IP}$ distribution can present more than one peak structure, for each $N_{\rm Gauss}$ value a second iterative procedure initialising the Gaussian means with 10 different seeds randomly chosen between 1 and 500 is performed. The fit procedure with the best χ^2/ndf and where no parameter has collapsed on its minimum or maximum allowed value in the fit (which is typically a hint for fit failure) is chosen. In all fit attempts, $\sigma_j > 0.5$ is initially required in order to remove spikes in the fit function. Then, if the best fit χ^2/ndf is larger than 2.5, further fit procedures decreasing the allowed minimum width by 0.05 at each iteration are performed.
Figs. II.1.4 and II.1.5 show the obtained results in the $\overline{p}$-enriched simulated sample for the antiproton and background categories, respectively, according to the histogram (left) and parametric (right) models and integrating over the full kinematic phase-space. For the latter, the fit pull distribution is also presented and demonstrates a good quality of the parametrization. An unexpected structure at $log(\chi^2_{\rm IP}) \sim 5$ is observed in the $\chi^2_{\rm IP}$ distribution for prompt antiprotons. Its origin and the contribution to the systematic uncertainty due to its possible imprecise simulation are described in Sec. II.1.6.1.
The templates drawing is then repeated as a function of the antiproton reconstructed momentum and transverse momentum considering the binning scheme indicated above. In the high-momentum or high-transverse momentum regions, the available statistics in the simulation for some category is insufficient to draw a smooth template. To mitigate for this, a minimum number of 100 candidates is required to draw the template and, in case the criterion is not met, those from the neighbour bins are added up to that limit. Since a smooth evolution of the $\chi^2_{\rm IP}$ with the kinematics is expected, this improves the precision of the parametrization of the distributions without introducing a significant distortion. The same fit procedure explained above is repeated for each category in all kinematic intervals, with some examples in random bins for prompt and detached antiprotons illustrated in Fig. II.1.6.

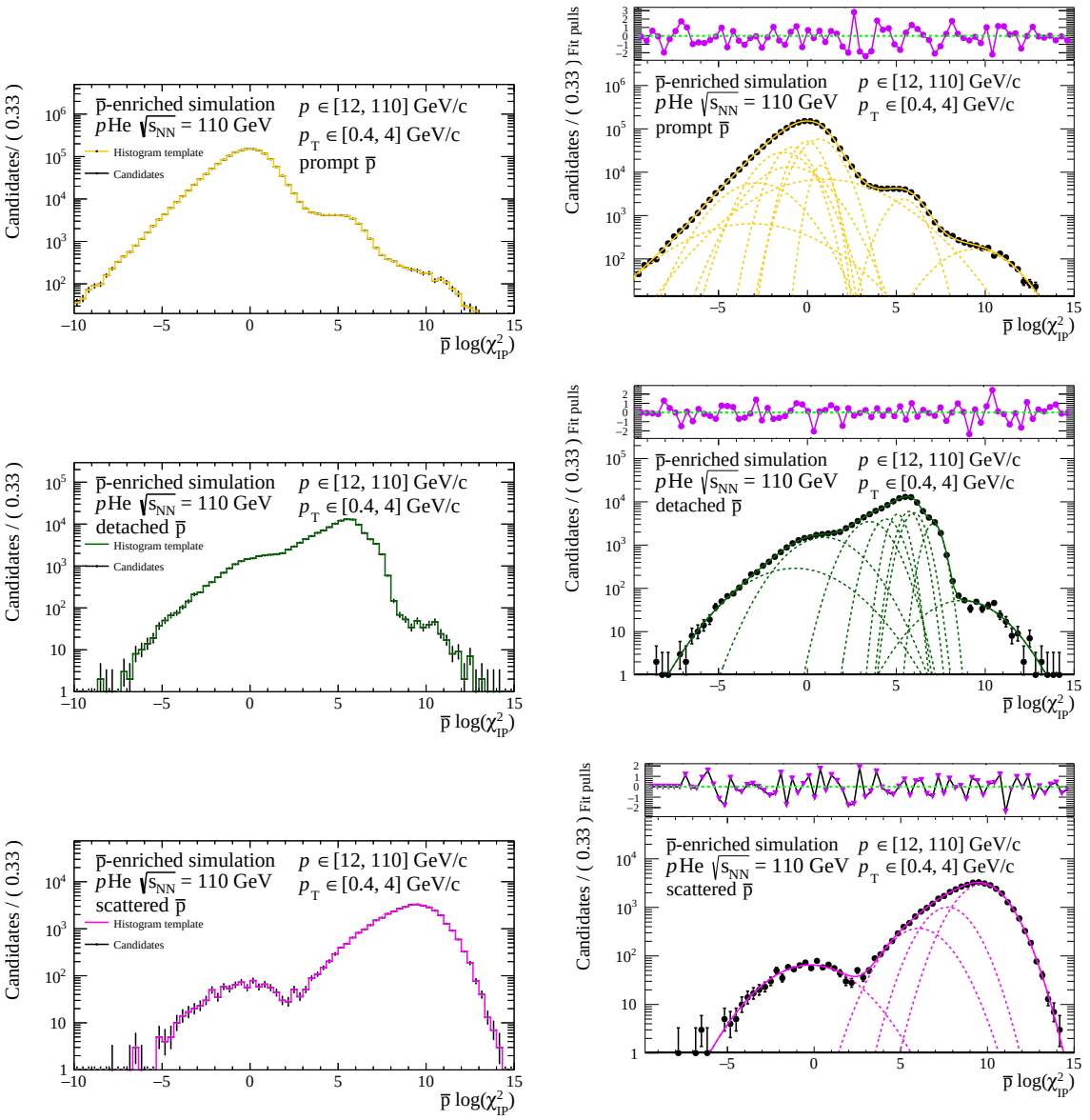


Figure II.1.4: Modelling of the χ^2_{IP} distributions for the prompt (top), detached (middle) and scattered (bottom) antiproton categories in the $\overline{p}$-enriched simulated sample according to the histogram (left) and parametric (right) model.

The template distributions for all antiproton and background categories are then composed in a unique fit model. As a sanity check of the parametrization with the two models, the fit is firstly performed on the minimum-bias and $\overline{p}$-enriched simulated samples, considering the respective templates and fixing the MISID contamination, about 3% of the total selected particles after the selection algorithm implementation, to the respective predictions. For this fit, the expectation values λ_i that are considered in the likelihood function are:

$$\lambda_i = \sum_p f_p \cdot \lambda_i^p \tag{II.1.6}$$

being p an index running over the particle species and f_p their relative abundances. The free parameters of the fit are $f_{\mathrm{prompt}\ \overline{p}}$, $f_{\mathrm{det}\ \overline{p}}$ and $f_{\mathrm{scattered}\ \overline{p}}$. The expectation is to

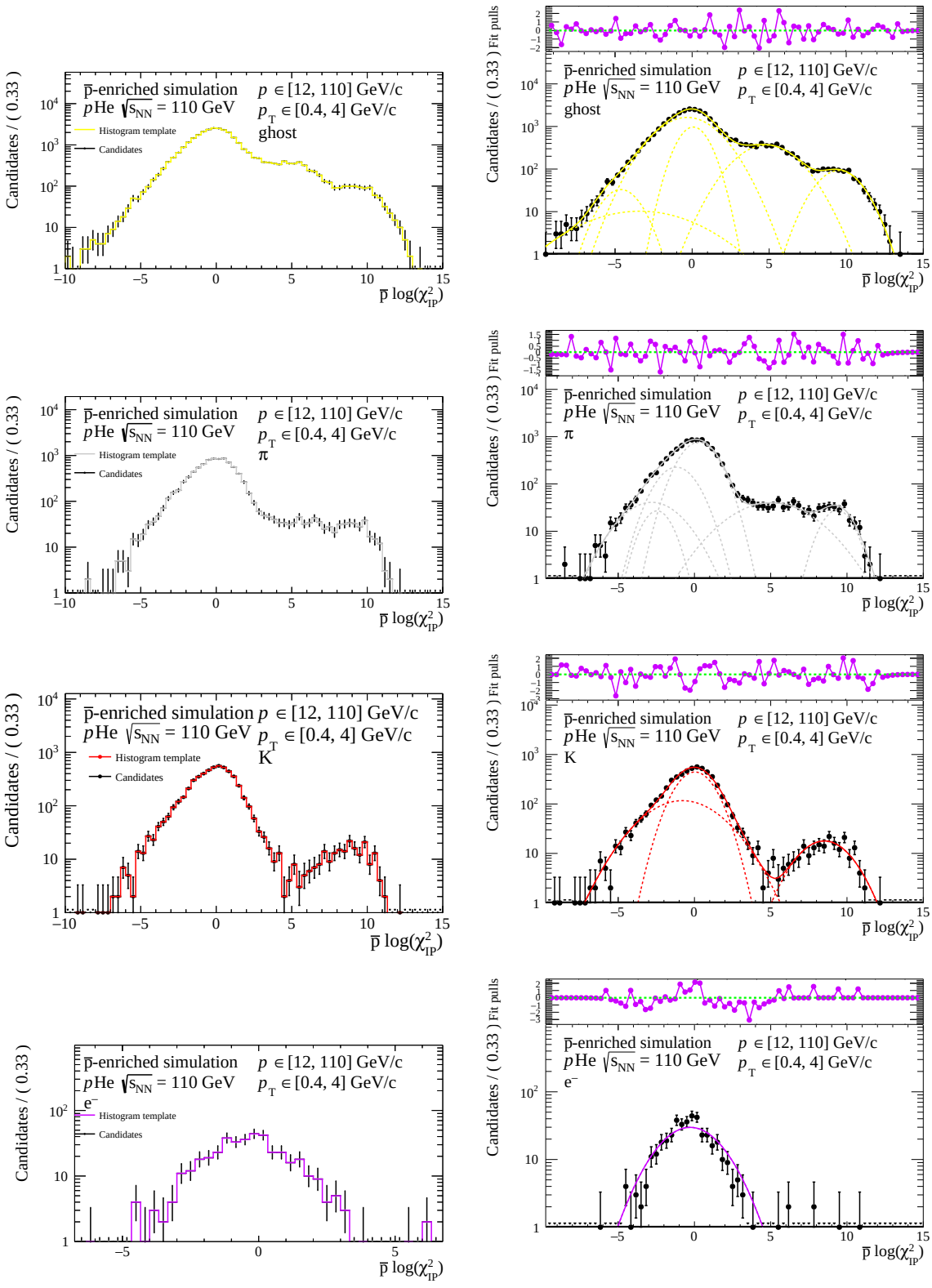


Figure II.1.5: Modelling of the χ^2_{IP} distributions for the (from top to bottom) ghost, pion, kaon and electron background categories in the $\overline{p}$-enriched simulated sample according to the histogram (left) and parametric (right) model.

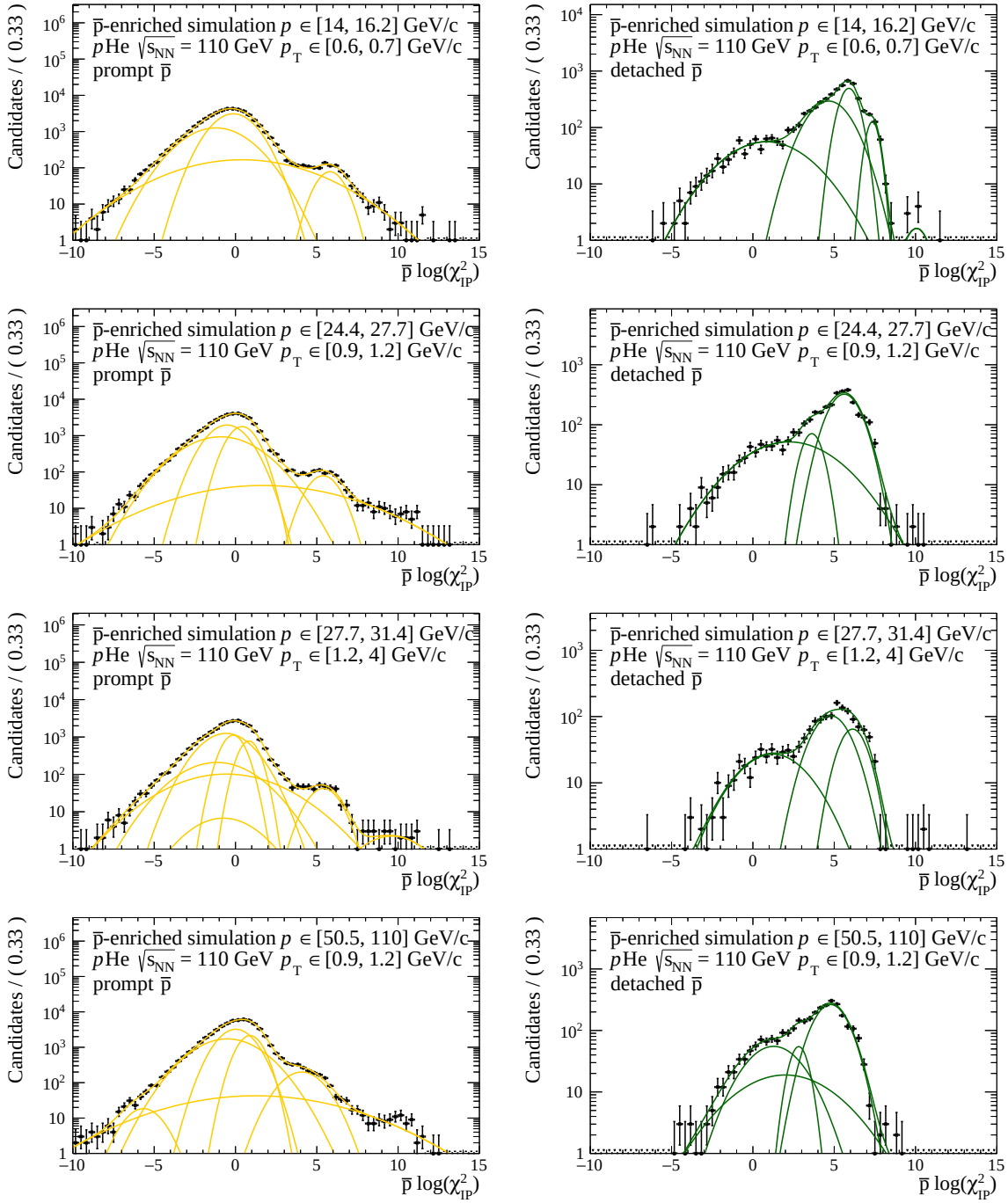


Figure II.1.6: Parametric models for the χ^2_{IP} distributions of prompt (left) and detached (right) antiprotons in the indicated kinematic bins in the $\overline{p}$-enriched simulated sample.

Table II.1.4: Prediction according to the minimum-bias and the $\overline{p}$-enriched simulated samples and fitted prompt and detached antiproton abundances for all fits applied over the samples integrated over the full kinematic phase-space.

		prompt $\overline{p}$	detached $\overline{p}$
MB-MC prediction		93670	8005
Fit to MB-MC	Hist. model	93672.1 ± 329.4	8004.5 ± 128.2
	Par. model	93689.4 ± 329.4	7957.1 ± 128.1
$\overline{p}$-MC prediction		1549312	133433
Fit to $\overline{p}$-MC	Hist. model	1549348 ± 1304	133455 ± 514
	Par. model	1549358 ± 1304	133441 ± 514
MB-MC prediction (scaled to data)		1048328	89596
Fit to data	Hist. model	1026376 ± 1095	128247 ± 485
	Par. model	1026425 ± 1095	128247 ± 485

reproduce in the two samples the respective predicted antiproton abundances.
Figs. II.1.7 and II.1.8 show the obtained results for the minimum-bias and the $\overline{p}$-enriched simulated samples integrating over the full kinematic phase-space according to the histogram (top) and parametric (bottom) models. In all plots, a good quality of the fit is achieved and the results in two models are found to agree, validating the parametrization strategies. The fitted numbers of prompt and detached antiprotons, listed for both simulated samples in Tab. II.1.4, are divided to get the detached-to-prompt antiproton yield ratio R_{raw}. With the histogram model,

$$R_{\text{raw}}^{\text{MB-MC}} = \frac{N_{\text{det}}^{\text{MB-MC}}}{N_{\text{prompt}}^{\text{MB-MC}}} = 0.0854 \pm 0.0014 \tag{II.1.7}$$

for the minimum-bias and

$$R_{\text{raw}}^{\overline{p}\text{-MC}} = \frac{N_{\text{det}}^{\overline{p}\text{-MC}}}{N_{\text{prompt}}^{\overline{p}\text{-MC}}} = 0.0861 \pm 0.0003 \tag{II.1.8}$$

for the $\overline{p}$-enriched simulated sample. Within the experimental uncertainty, the results in the two samples agree. This proves that the selection cuts that are applied to the generation of the $\overline{p}$-enriched sample do not induce bias in the result. Providing this a larger statistics than the minimum-bias one, it will be used as the main simulation for this analysis.
To verify that the predictions are reproduced in all the considered intervals of the antiproton momentum and transverse momentum, the fit is iterated. The results for the prompt (top) and detached (middle) antiproton fitted abundances and for their ratio (bottom) are presented in Fig. II.1.9 projected onto the momentum (left) and transverse momentum (right) axes. In all plots, the predictions in the $\overline{p}$-enriched simulated samples are shown in green and superimposed to the results with the histogram (parametric) model in red (blue). Over the full kinematic phase-space, both models give consistent results with the prediction, validating the fit procedure.

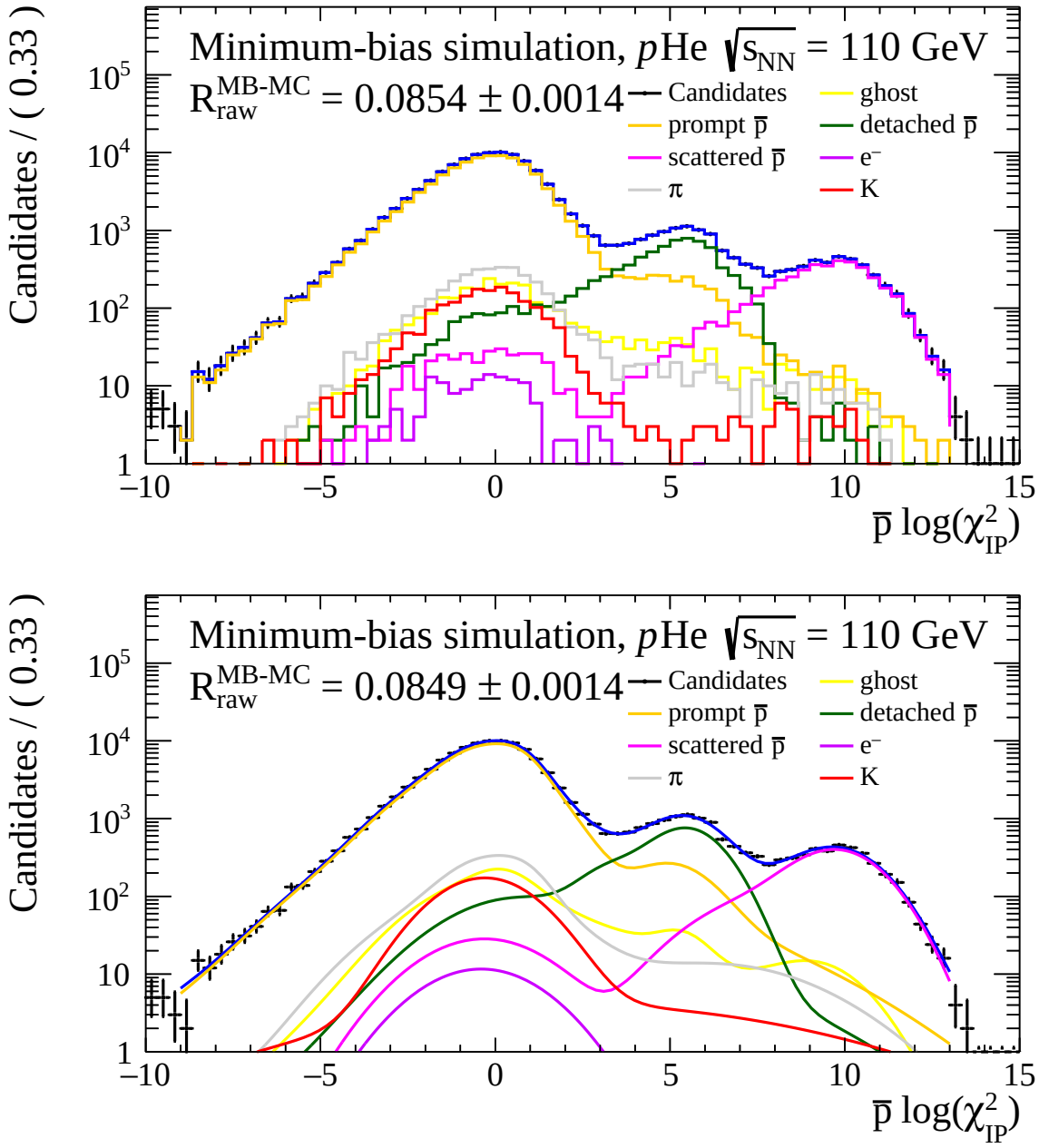


Figure II.1.7: $\chi^2_{\rm IP}$ distribution for all selected particles in the minimum-bias simulated sample integrated over the considered kinematic phase-space. The distribution is modelled with the templates drawn on the minimum-bias simulation according to the histogram (top) and the parametric (bottom) models for all background and antiproton categories. The relative abundances of the former are fixed to the minimum-bias prediction, the latter are left as free parameters of the fit. The resulting detached-to-prompt yield ratio $R_{\rm raw}^{\rm MB\text{-}MC}$ is indicated.

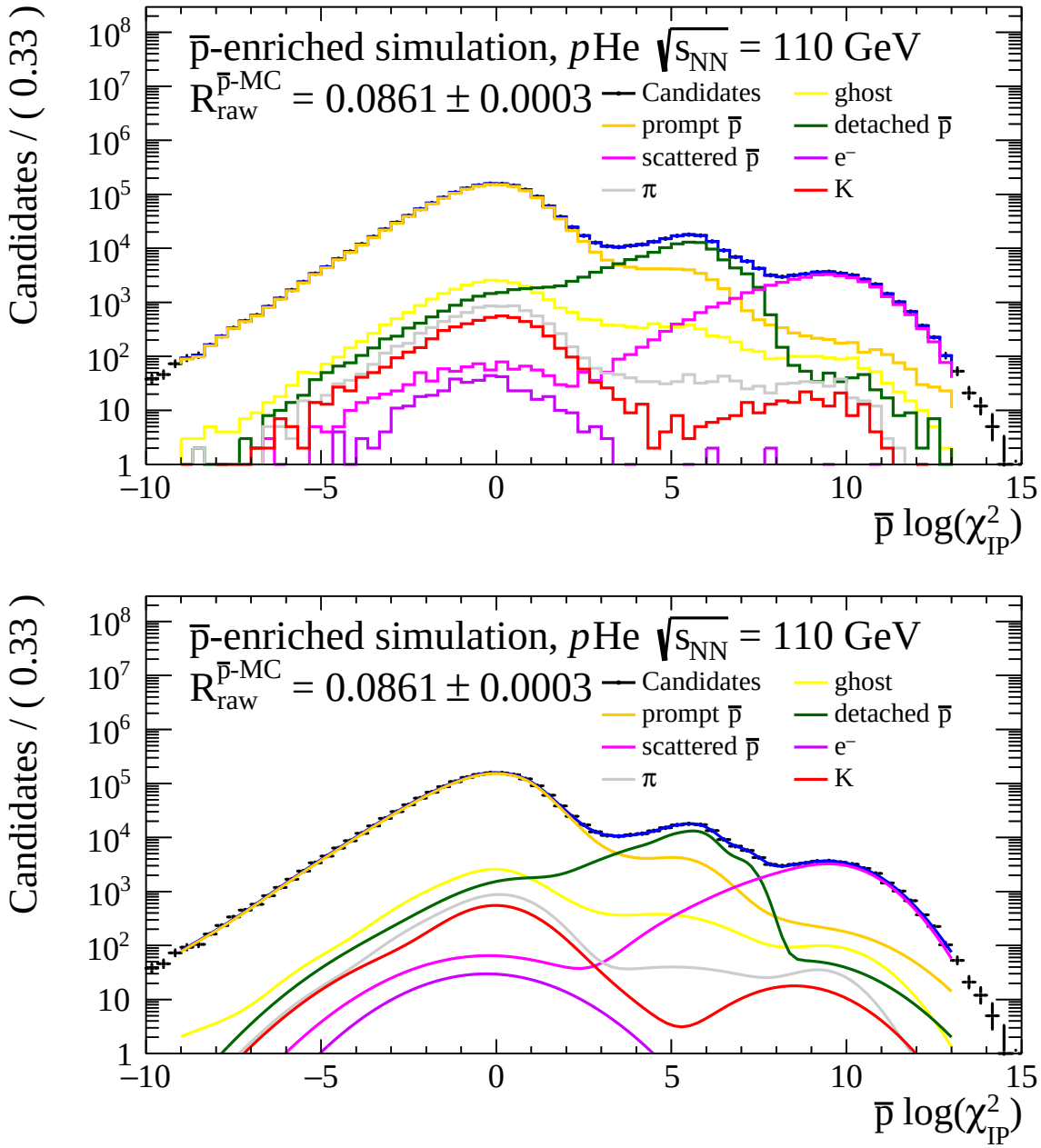


Figure II.1.8: χ^2_{IP} distribution for all selected particles in the $\overline{p}$-enriched simulated sample integrated over the considered kinematic phase-space. The distribution is modelled with the templates drawn on the $\overline{p}$-enriched simulation according to the histogram (top) and the parametric (bottom) models for all background and antiproton categories. The relative abundances of the former are fixed to the $\overline{p}$-enriched prediction, the latter are left as free parameters of the fit. The resulting detached-to-prompt yield ratio $R^{\overline{p}\text{-MC}}_{\mathrm{raw}}$ is indicated.

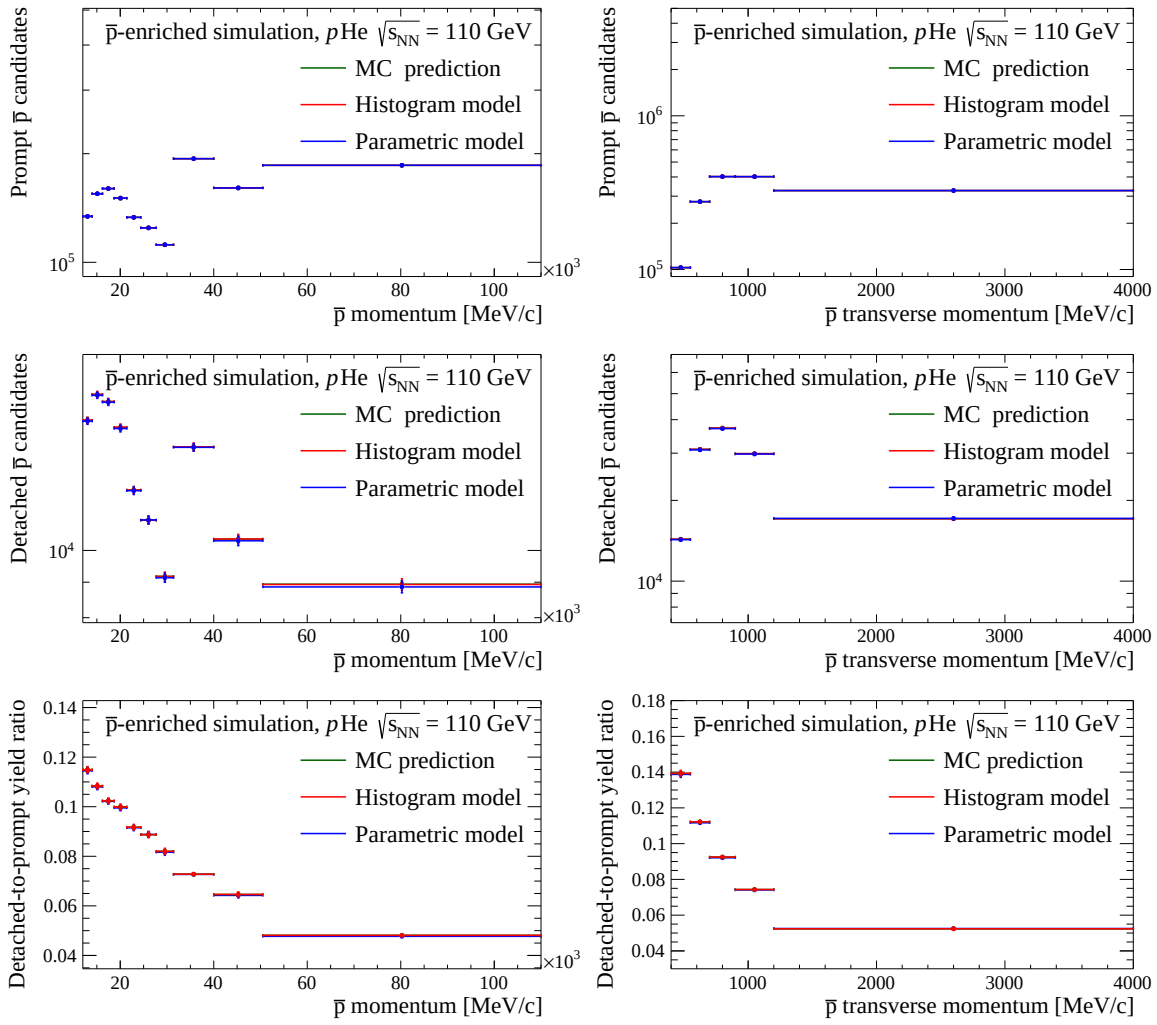


Figure II.1.9: Validation of the fit procedure on the $\overline{p}$-enriched simulated sample: comparison between the predictions (in green) and the fit results according to the histogram (in red) and parametric (in blue) models for the (top) prompt and (middle) detached antiproton abundances and for their yield ratio (bottom) as a function of the antiproton momentum (left) and transverse momentum (right).

The fit to the data sample is then performed. The abundances of the antiproton and MISID categories are initialised according to the predictions in the minimum-bias simulation, scaled by the difference in statistics with the data. The templates are drawn on the $\overline{p}$-enriched simulated sample, less prone to imperfections because of the higher statistics. Figure II.1.10 shows the $\chi^2_{\rm IP}$ distribution for the total pHe data sample and the same three-peak structure corresponding to prompt, detached and scattered antiprotons as in simulation is visible. Only the last peak shows a different evolution than on simulation as a consequence of the larger statistics available in data and of some simplifications in the description of the detector material in the GEANT4 simulation. Considering that for $log(\chi^2_{\rm IP}) > 11$ the scattered-to-detached antiproton ratio is below ~ 0.01, the fit is limited to $\overline{p}$ $log(\chi^2_{\rm IP}) \sim 11.5$, the endpoint of the detached template. Biases in the fit due to the scattered tail are thus excluded. Top (bottom) plot in Fig. II.1.10 shows the fit results according to the histogram (parametric) models integrating over the full kinematic

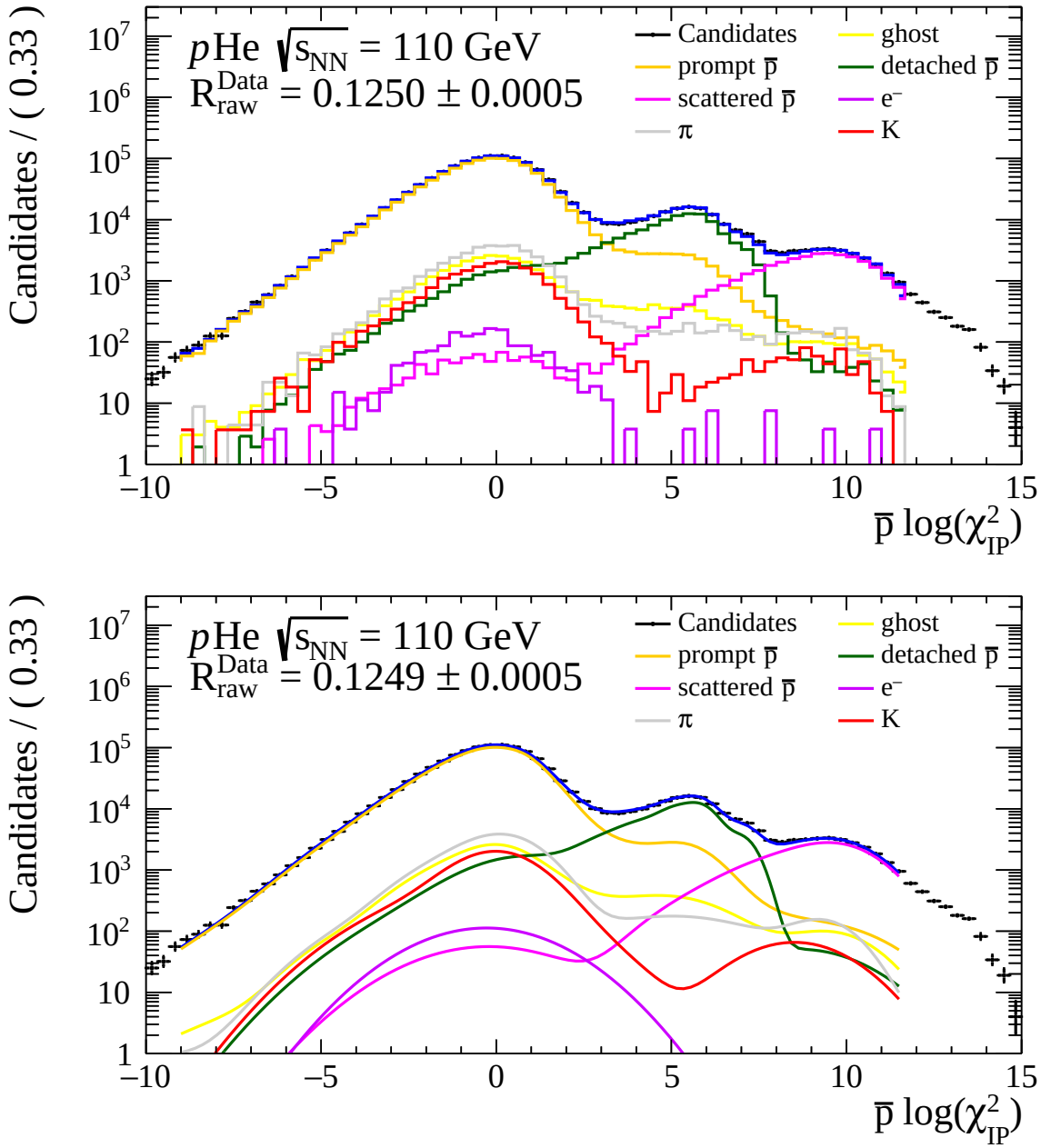


Figure II.1.10: χ^2_{IP} distribution for all selected particles in the data sample integrated over the considered kinematic phase-space. The distribution is modelled with the templates drawn on the $\overline{p}$-enriched simulated sample according to the histogram (top) and the parametric (bottom) models for all background and antiproton categories. The relative abundances of the former are fixed to the prediction in the minimum-bias simulated sample, the latter are left as free parameters of the fit. The resulting detached-to-prompt yield ratio $R_{\mathrm{raw}}^{\mathrm{Data}}$ is indicated

phase-space. The yields for prompt and detached antiprotons obtained from the fits, listed in Tab. II.1.4, are divided to calculate the raw detached-to-prompt yield ratio:

$$R_{\rm raw}^{\rm Data} = \frac{N_{\rm det}^{\rm Data}}{N_{\rm prompt}^{\rm Data}}(histogram) = 0.1250 \pm 0.0005 \tag{II.1.9}$$

with the histogram and

$$\frac{N_{\rm det}^{\rm Data}}{N_{\rm prompt}^{\rm Data}}(parametric) = 0.1249 \pm 0.0005 \tag{II.1.10}$$

with the parametric model. The $R_{\rm raw}^{\rm Data}\,(histogram) - R_{\rm raw}^{\rm Data}\,(parametric)$ difference, estimating the systematic uncertainty due to the choice of a parametrization, is found to be small with respect to the statistical contribution. On the opposite, when comparing with the predicted detached-to-prompt antiproton yield ratio in Eq. II.1.8, an excess by a factor $\sim 45\%$ is measured.
The fit is repeated in the considered intervals of the antiproton momentum and transverse momentum and a comparison with the expectations is performed as described above. Figures. II.1.11 and II.1.12 show the predictions (top) and the obtained result in the two models (middle and bottom) for prompt and detached antiproton raw yields, respectively, in the intervals of the antiproton momentum and transverse momentum. The resulting detached-to-prompt yield ratio is presented in Fig. II.1.13 according to the histogram (top) and parametric (bottom) model. All plots are then projected onto the momentum (left) and transverse momentum (right) axes and shown superimposed in Fig. II.1.14. In this case, an evident difference with respect to the expectations is always observed. Results for the two models are mostly consistent, with some larger deviations at low-transverse momentum. This is a consequence of the uncertainty in the chosen parametrization for antiprotons and for backgrounds, representing in these kinematic regions a larger fraction of the total. The effect is measured as the difference in the obtained results with the two models and is presented in Fig. II.1.15. In most of the bins the found difference due to the parametrization is minor with respect to the statistical one.

II.1.5 Experimental efficiency

As indicated in Eq. II.1.2, the R measurement requires the detached-to-prompt $\overline{p}$ yield ratio to be corrected by the respective experimental efficiencies:

$$\varepsilon_{\rm det} = \varepsilon_{\rm det}^{\rm sim} \cdot \varepsilon_{\rm det}^{\rm rec\ \&\ sel} = \frac{N_{\rm det}^{\rm sim}}{N_{\rm det}^{\rm gen}} \cdot \frac{N_{\rm det}^{\rm rec\ \&\ sel}}{N_{\rm det}^{\rm sim}} \tag{II.1.11}$$

$$\varepsilon_{\rm prompt} = \varepsilon_{\rm prompt}^{\rm sim} \cdot \varepsilon_{\rm prompt}^{\rm rec\ \&\ sel} = 1 \cdot \frac{N_{\rm prompt}^{\rm rec\ \&\ sel}}{N_{\rm prompt}^{\rm sim}} \tag{II.1.12}$$

being:

- $N_{\rm det\,(prompt)}^{\rm rec\ \&\ sel}$ the number of detached (prompt) reconstructed and selected antiprotons;
- $N_{\rm det\,(prompt)}^{\rm sim}$ the number of simulated detached (prompt) antiprotons;

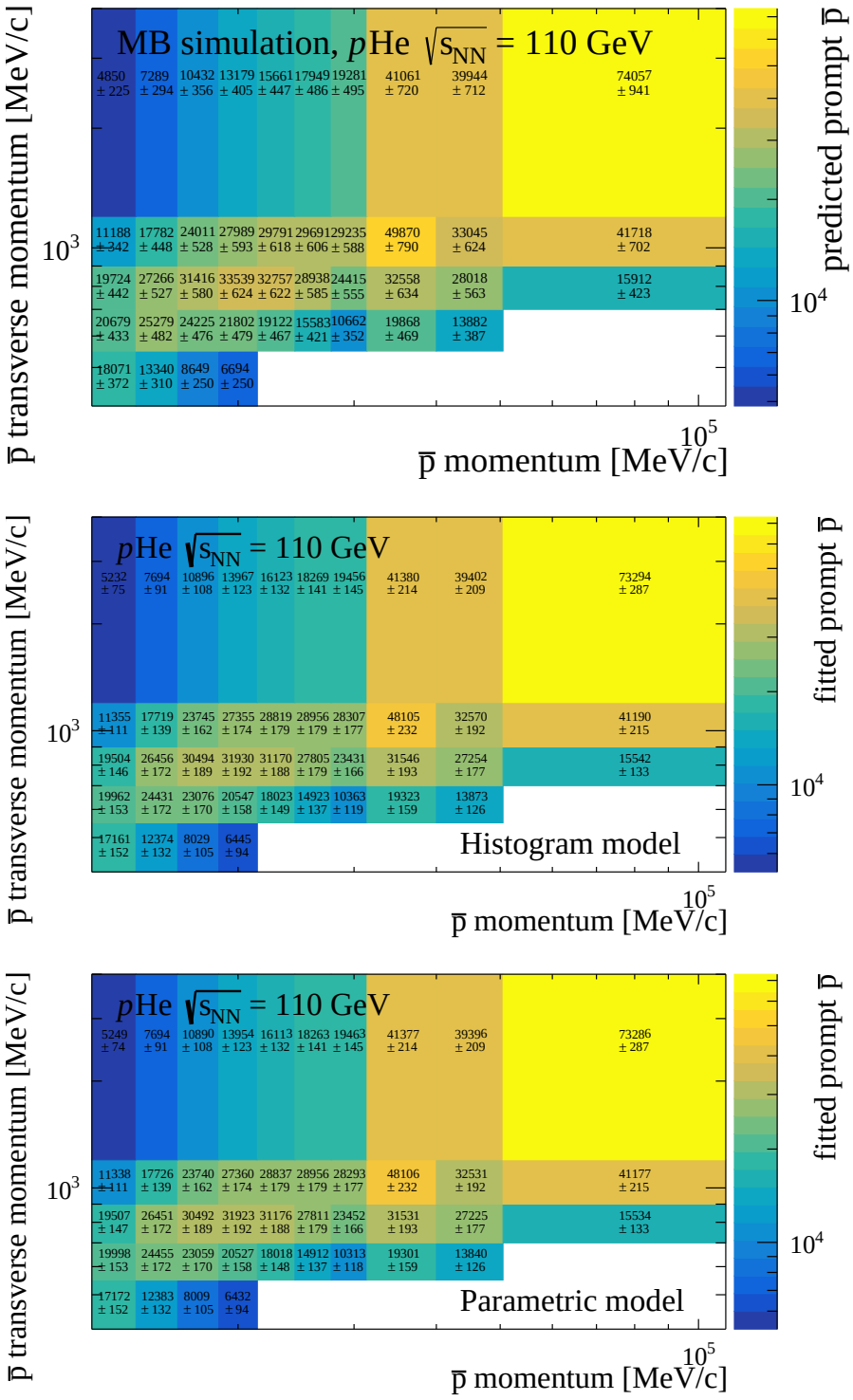


Figure II.1.11: Abundance of the prompt antiprotons in momentum and transverse momentum bins predicted by the minimum-bias simulation (top) and obtained from the template fit to the data sample according to the histogram (middle) and parametric (bottom) models.

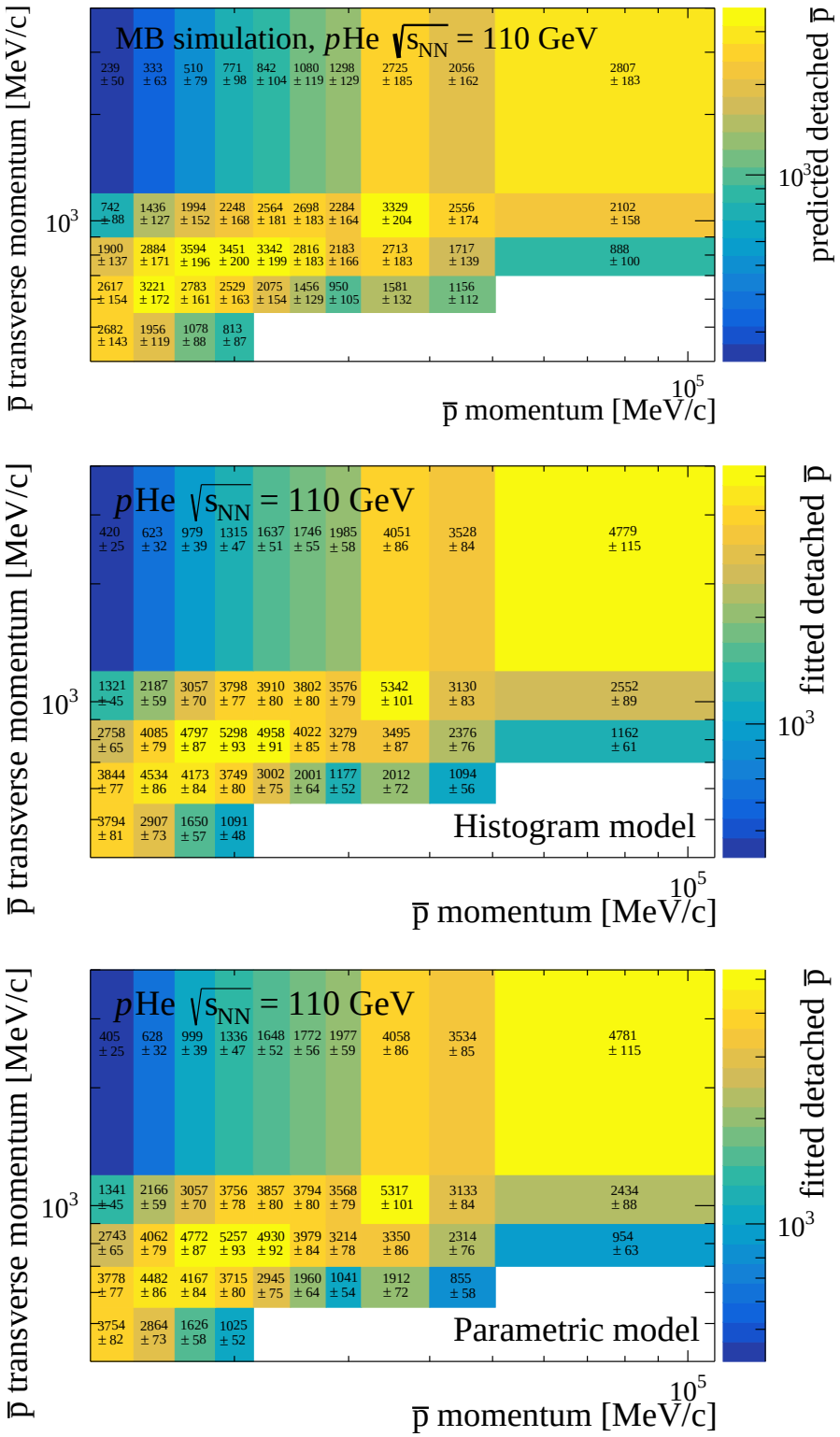


Figure II.1.12: Abundance of the detached antiprotons in momentum and transverse momentum bins predicted by the minimum-bias simulation (top) and obtained from the template fit to the data sample according to the histogram (middle) and parametric (bottom) models.

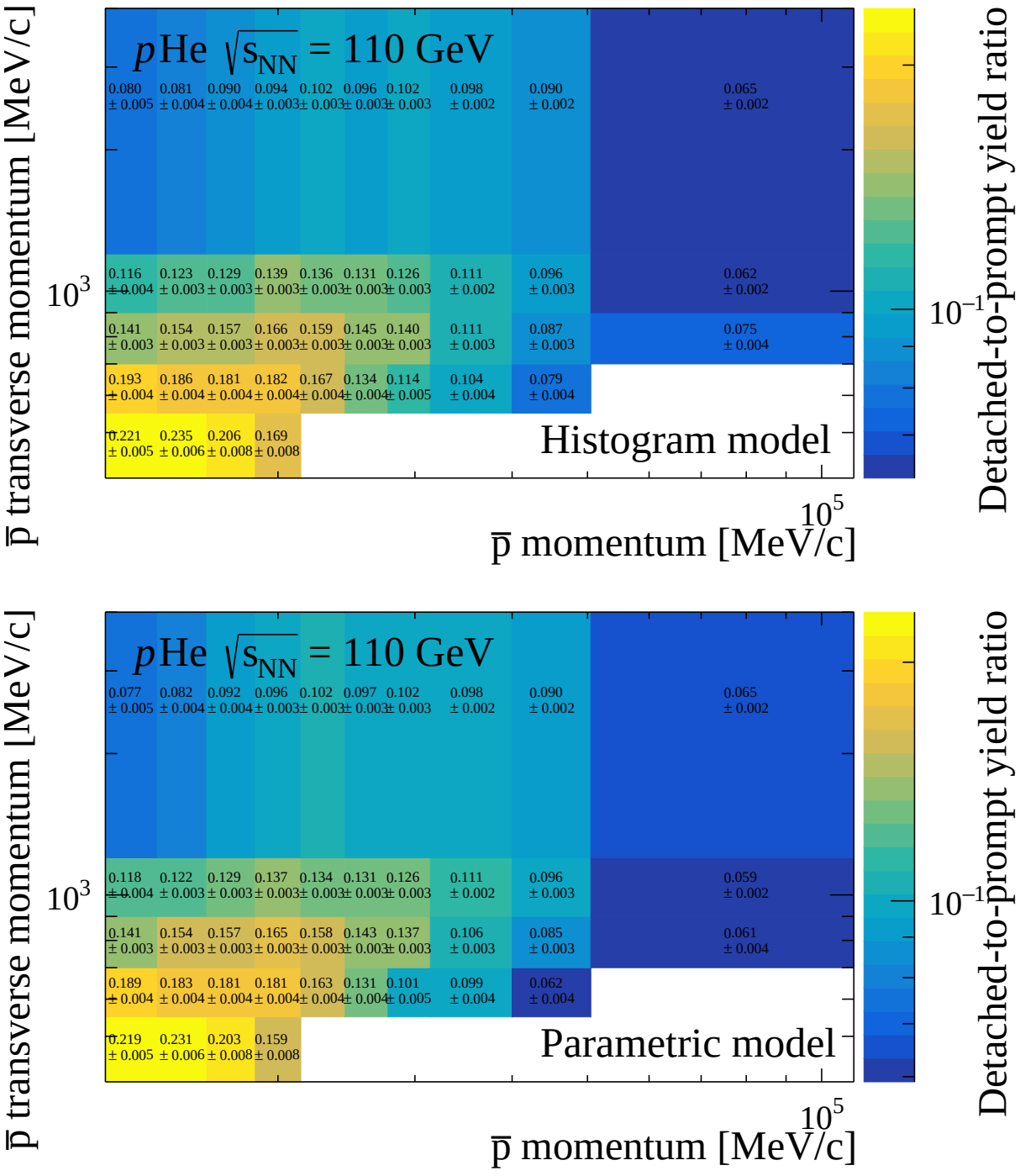


Figure II.1.13: Detached-to-prompt antiproton yield ratio in data as a function of the antiproton momentum and transverse momentum as resulting from the fit with the templates drawn with the parametric (top) or histogram (bottom) model.

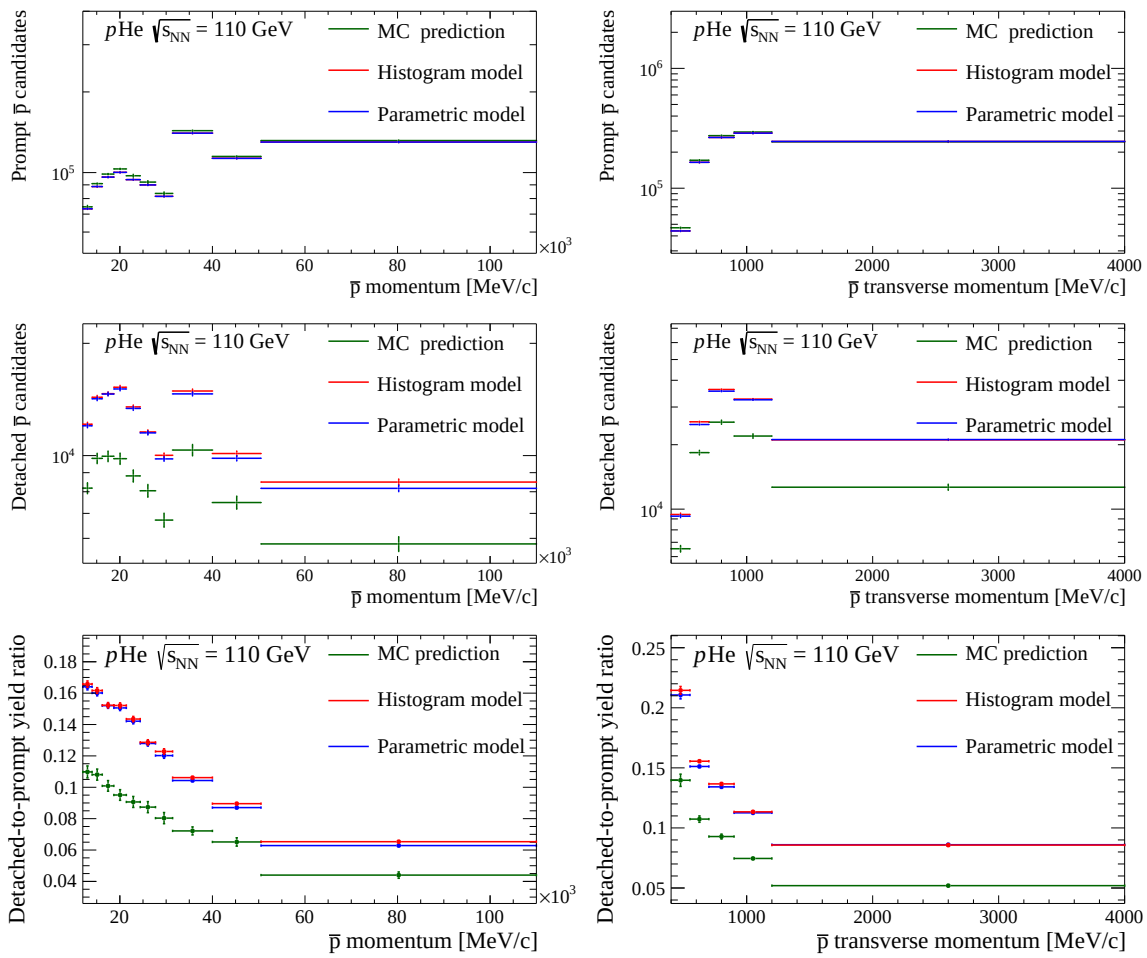


Figure II.1.14: Comparison between the prediction (in green) and the fit results according to the histogram (in blue) and parametric (in red) model for (top) prompt and (middle) detached antiproton abundances and for their ratio as a function of the antiproton momentum (left) and transverse momentum (right).

- $N_{\text{det}}^{\text{gen}}$ the number of generated detached antiproton candidates, including those that are excluded from the simulated sample because of a parent particle interaction with the detector material or its leaving the LHCb detector before decaying. By definition, generated and simulated prompt antiprotons coincide instead.

A further efficiency term due to the selection requirements that are applied to produce the $\overline{p}$-enriched simulated sample, discussed in Sec. II.1.2, should be considered. Anyhow, the imposed thresholds are looser with respect to the offline selection and do not contribute to the inefficiency, as confirmed by the observed consistency between the R_{raw} measurement in the minimum-bias and the $\overline{p}$-enriched samples. The efficiency for the generation of the sample is thus 1. The evaluation of the simulation and reconstruction and selection efficiencies are discussed in Sec. II.1.5.1 and II.1.5.2, respectively.

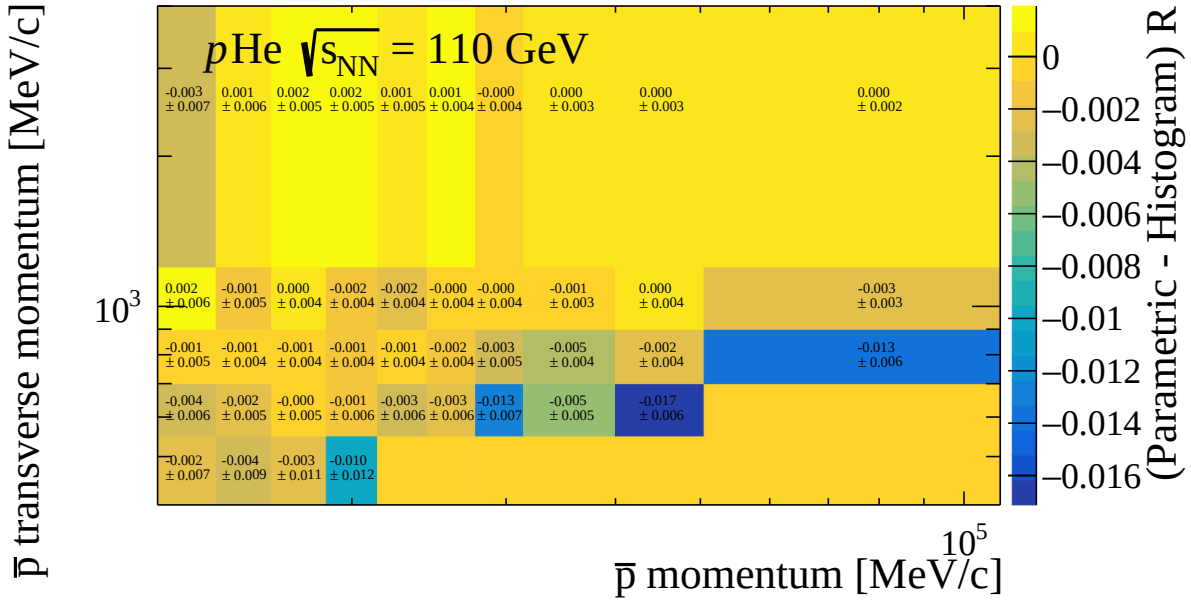


Figure II.1.15: Difference between the obtained results for the detached-to-prompt antiproton yield ratio in pHe data with the parametric and histogram models as a function of the antiproton momentum and transverse momentum.

II.1.5.1 Simulation efficiency

A first term contributing to the efficiency for detached antiprotons is due to the generated candidates that are not included in the simulated sample since, before its decay, the anti-hyperon:

- is involved in a hadronic interaction with the detector material;
- leaves the LHCb detector region.

By definition, both effects are null for prompt antiprotons, while

$$\varepsilon_{\rm det}^{\rm sim} = \frac{N_{\rm det}^{\rm sim}}{N_{\rm det}^{\rm gen}}. \tag{II.1.13}$$

Generated candidates are counted in the sample produced with the same selection requirements as the $\overline{p}$-enriched simulated sample, but disabling the particles propagation through the LHCb detector. Provided that the generated and simulated samples are statistically equivalent, *i.e.* that the same number of pHe collisions is considered, the ratio of their detached antiprotons gives $\varepsilon_{\rm det}^{\rm sim}$. The procedure is validated on prompt antiprotons, with the results in the considered intervals of the antiproton momentum and transverse momentum (top) and projected onto the two axes (bottom) presented in Fig. II.1.16. In all bins, a value consistent with one within the statistical uncertainty is found. For the detached antiprotons, the results are presented in Fig. II.1.17. The efficiency at low momenta, where hadronic interactions with the detectors are dominant, is near to one and then steeply decreases because of the anti-hyperon decays outside of the LHCb detector region. This is expected, considering the antiproton momentum

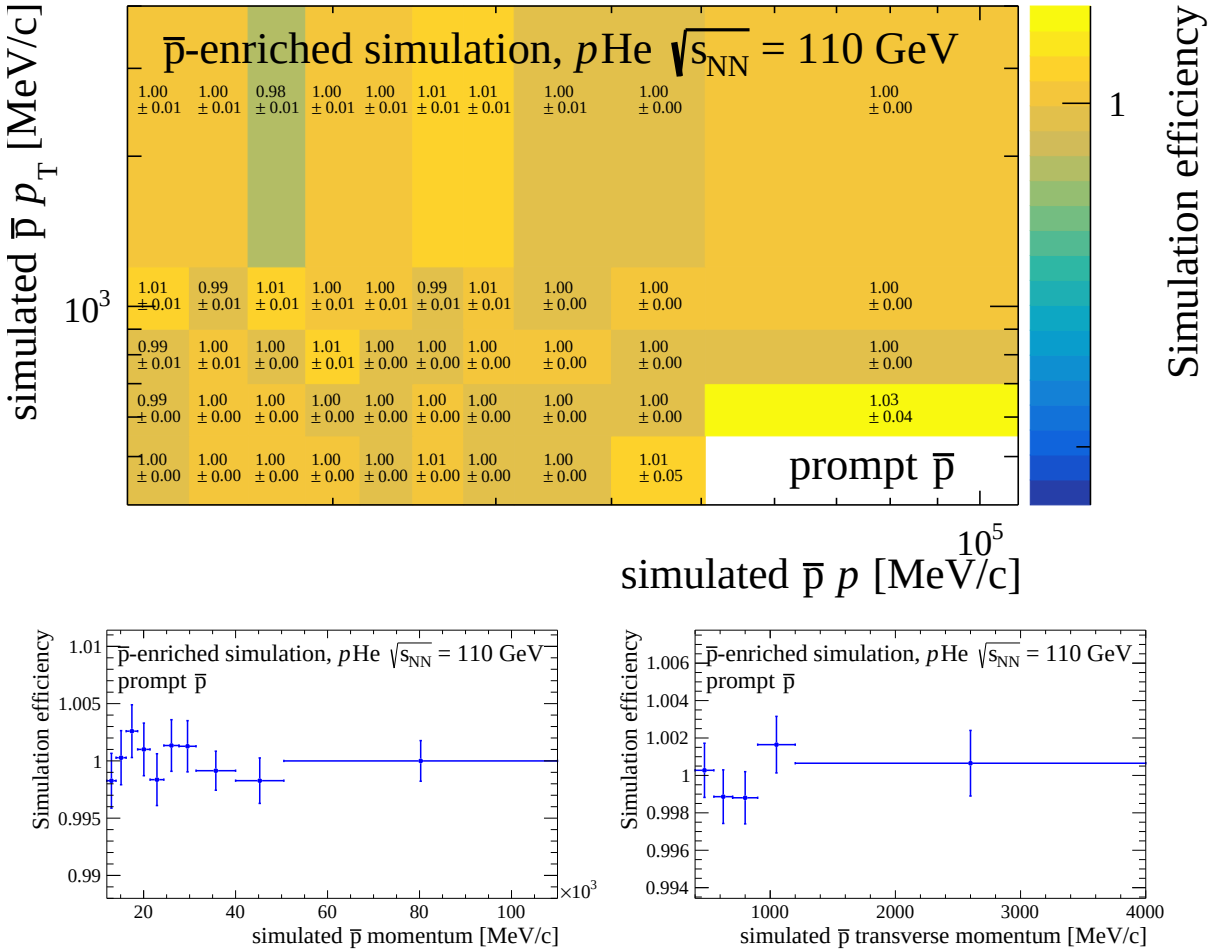


Figure II.1.16: Simulation efficiency for prompt antiprotons in the $\overline{p}$-enriched simulated sample in bins of their momentum and transverse momentum (top) and projected onto the momentum (bottom left) and transverse momentum axis (bottom right).

correlation with the anti-hyperon time of flight.
The simulation efficiency is then separately measured for the antiprotons produced in $\overline{\Lambda}$ and $\overline{\Sigma}^-$ decays. Also, as illustrated in Fig. II.1.2, the $\overline{\Lambda}$ baryons produced promptly or in $\overline{\Xi}^+$ or $\overline{\Xi}^0$ decays are distinguished because of the expected different antiproton OV_z distribution. For the separate anti-hyperon categories, the statistical fluctuations are such that a complete bidimensional study can only be performed for promptly-produced $\overline{\Lambda}$ particles, with the results presented in Fig. II.1.18. For all other categories, the dependence on the transverse momentum is neglected and only the efficiency evolution with the momentum is studied in Fig. II.1.19. For $\overline{\Sigma}^-$ particles, $\varepsilon^{\mathrm{sim}}_{\overline{\Sigma}^-} \simeq 1$ is found, being their lifetime the lowest among all baryons; for the other species, a steep decreasing with the momentum is found. As expected, a clear dependence of the simulation efficiency, particularly for its component due to the baryons leaving the detector, on the anti-hyperon lifetime is found.

An independent strategy to measure the fraction of anti-hyperon decays occurring outside the LHCb detector (indicated as decay-in-LHCb efficiency in the following) is performed to cross-check the evolution with the momentum. This is based on the decay

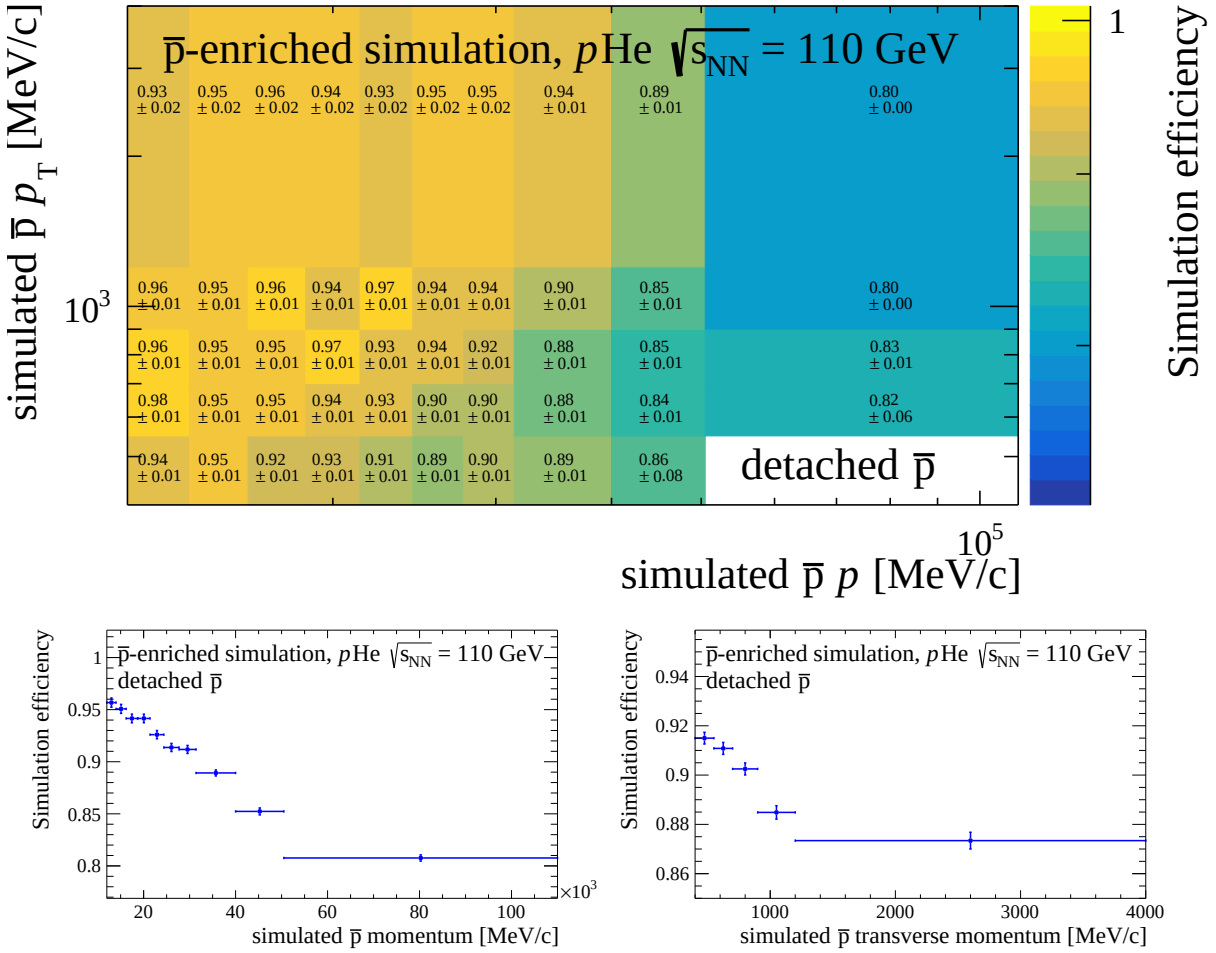


Figure II.1.17: Simulation efficiency for detached antiprotons in the $\overline{p}$-enriched simulated sample in bins of their momentum and transverse momentum (top) and projected onto the momentum (bottom left) and transverse momentum axis (bottom right).

time distribution, defined as:

$$t\,[\,\mathrm{ns}] = \frac{|OV - PV|[mm]\ \cdot\ M[\,\mathrm{MeV}/c^2]}{p[\,\mathrm{MeV}/c]\ \cdot\ 300} \tag{II.1.14}$$

being OV and PV the antiproton origin and the primary vertices, M the decaying anti-hyperon nominal mass and p its momentum. For antiprotons originating in $\overline{\Xi}$ decays, an effective decay time is considered, being OV-PV the distance covered by two decays in cascade and including the $\overline{\Lambda}$ mass and momentum in the formula. All quantities are considered before the event reconstruction in the detector, so the decay time distributions are expected to follow the theoretical evolutions: an exponential for prompt $\overline{\Lambda}$ and $\overline{\Sigma}^-$ decays and the convolution of two exponential functions for the $\overline{\Xi}$ baryons. As illustrated in the decay-time histograms of Fig. II.1.20, this is the case for only low-momentum antiprotons (left). With the increasing of the momentum (right), a fraction of the generated anti-hyperon candidates leaves the detector region and is removed from the sample. This is measured by comparing the decay time distributions with the expected theoretical

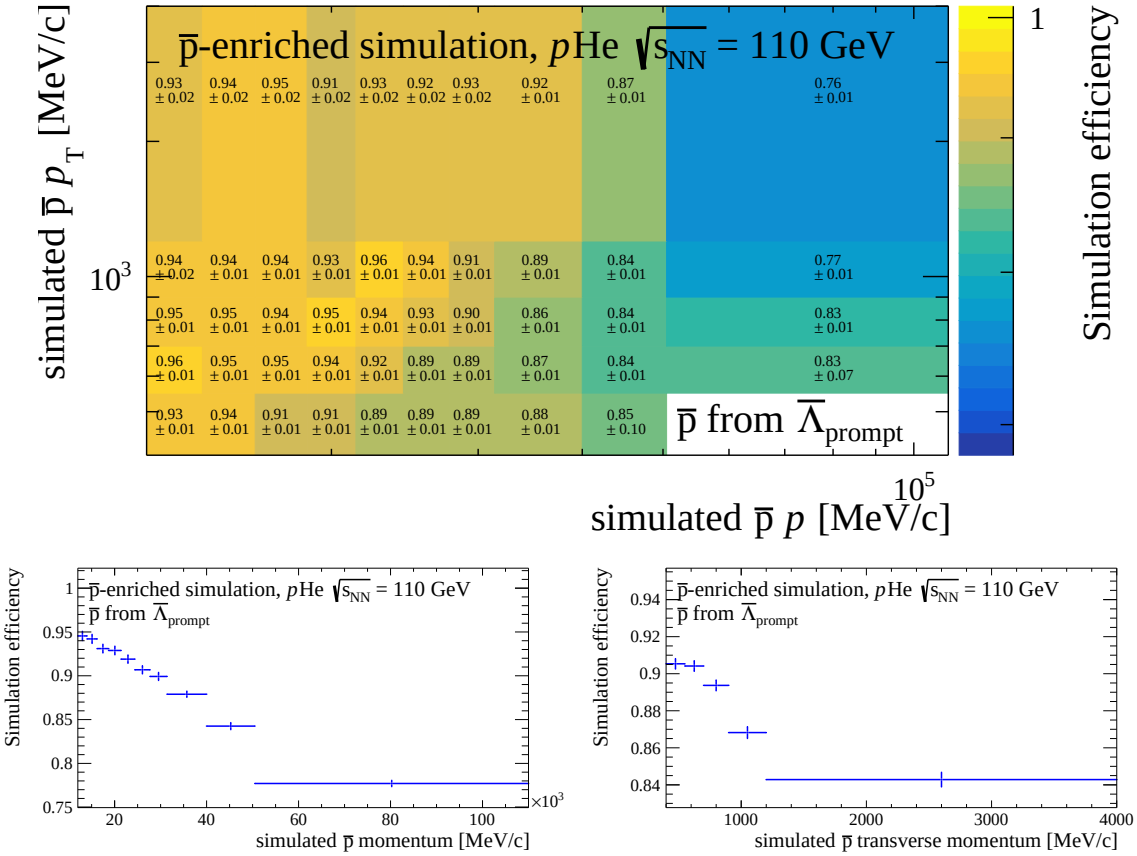


Figure II.1.18: Simulation efficiency for antiprotons from promptly-produced $\overline{\Lambda}$ baryons in the $\overline{p}$-enriched simulated sample in bins of their momentum and transverse momentum (top) and projected onto the momentum (bottom left) and transverse momentum axis (bottom right).

curves as:

$$\varepsilon^{\text{decay-in-LHCb}}_{\overline{\Lambda}_{\text{prompt}},\,\overline{\Sigma}^{-}} = \frac{N^{\text{MC}}_{\text{sim}\,\overline{\Lambda}_{\text{prompt}},\,\overline{\Sigma}^{-}}}{\int a \cdot e^{-t/\tau_{\overline{\Lambda}_{\text{prompt}},\,\overline{\Sigma}^{-}}}\,dt}$$

$$\varepsilon^{\text{decay-in-LHCb}}_{\overline{\Xi}^{+},\,\overline{\Xi}^{0}} = \frac{N^{\text{MC}}_{\text{sim}\,\overline{\Xi}^{+},\,\overline{\Xi}^{0}}}{\int a \cdot |e^{-t/\tau_{\overline{\Xi}^{+},\,\overline{\Xi}^{0}}} - e^{-t/\tau_{\overline{\Lambda}}}|\,dt} \tag{II.1.15}$$

being:

- $N^{\text{MC}}_{\overline{\Lambda}_{\text{prompt}},\,\overline{\Sigma}^{-},\,\overline{\Xi}^{+},\,\overline{\Xi}^{0}}$ the number of $\overline{p}$ originating in a $\overline{\Lambda}_{\text{prompt}}$, $\overline{\Sigma}^{-}$, $\overline{\Xi}^{+}$, $\overline{\Xi}^{0}$ decay;
- $\tau_{\overline{\Lambda}_{\text{prompt}},\,\overline{\Sigma}^{-},\,\overline{\Xi}^{+},\,\overline{\Xi}^{0}}$ the baryon lifetimes, fixed to the respective nominal values;

In order to determine the normalization parameter a, a fit to the binned distribution is performed, as illustrated in Fig. II.1.20 for each anti-hyperon category in a low- (left) and high-momentum (right) kinematic interval. Since the distortion of the distribution for high decay-time values could bias the fit result, it is limited to a maximum value t_{max}. This is iteratively increased up to 55% of the histogram end-point, chosen as a reasonable

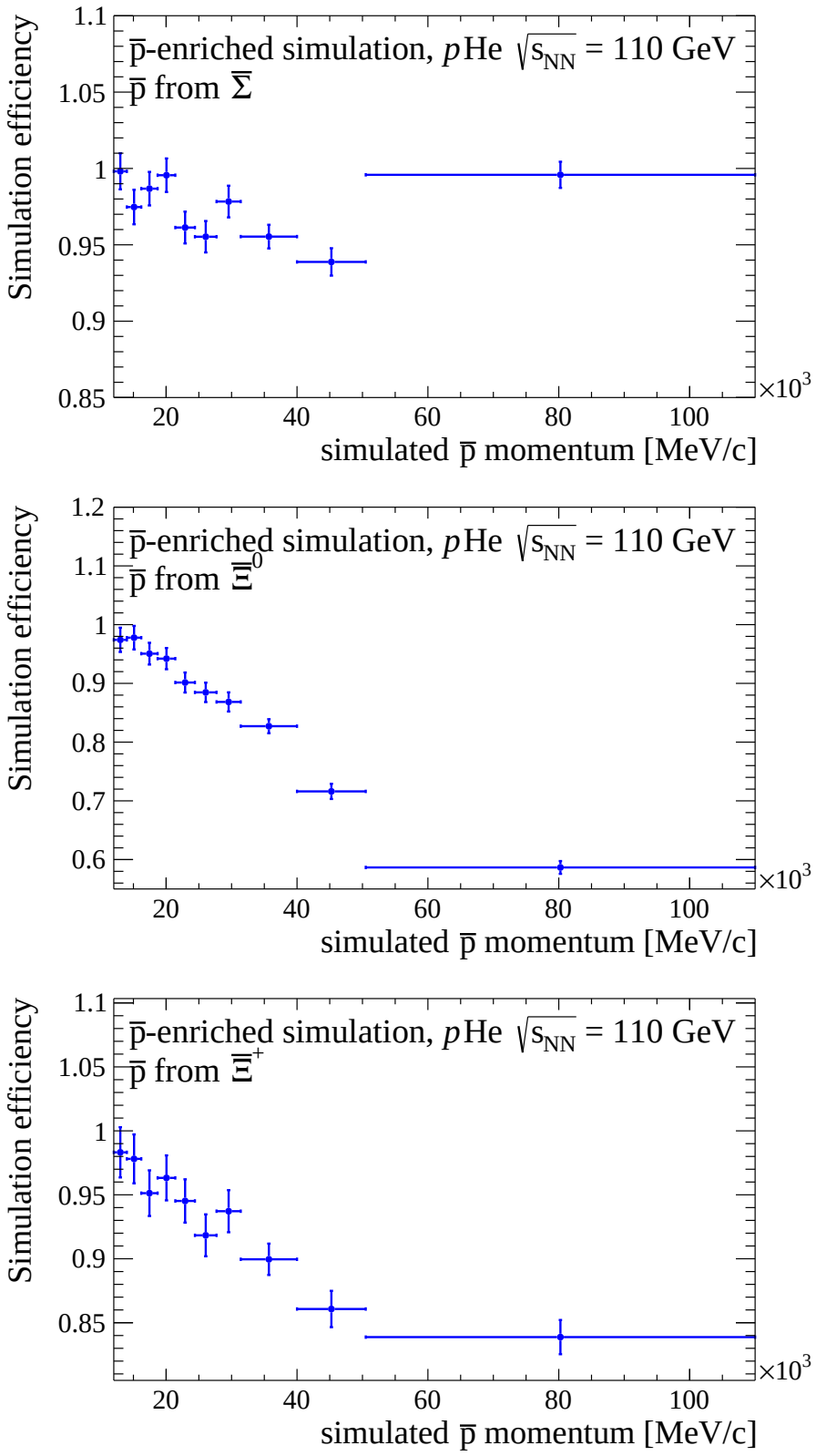


Figure II.1.19: Simulation efficiency for antiprotons originating in $\overline{\Sigma}^-$ (top), $\overline{\Xi}^0$ (middle) and $\overline{\Xi}^+$ (bottom) decays as a function of their momentum in the $\overline{p}$-enriched simulated sample. A clear dependence of the efficiency from the baryon lifetime emerges.

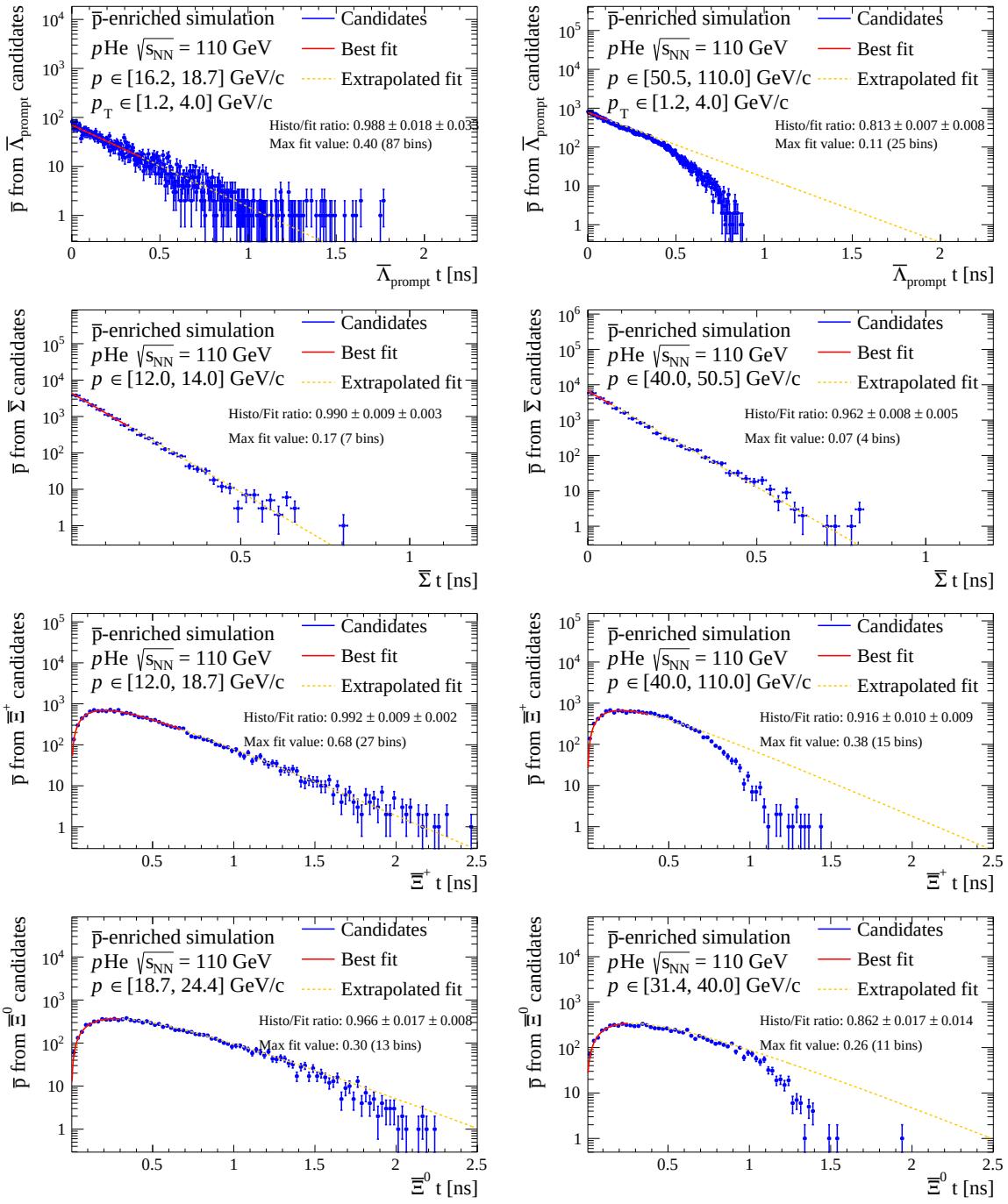


Figure II.1.20: Decay time distributions for (from top to bottom) $\overline{\Lambda}_{\rm prompt}$, $\overline{\Sigma}^{-}$, $\overline{\Xi}^{+}$ and $\overline{\Xi}^{0}$ baryons in the $\overline{p}$-enriched simulated sample in randomly chosen low- (left) and high-momentum (right) kinematic intervals. The employed models are an exponential function for the fist two anti-hyperons, the difference of two exponentials for the others. The ratio between the histogram and function integrals gives the decay-in-LHCb efficiency.

guess and, for each fit attempt, the χ^2/ndf is measured. The integral of the best fit function over the full decay time range and the related uncertainty are used to measure the simulation efficiency, according to Eq. II.1.15. The contribution to the uncertainty due to the choice of t_{max} is also added in square as the largest $\varepsilon^{\text{decay-in-LHCb}}$ variation from the result obtained repeating the fit with $t_{\text{max}} = (1 \pm 0.1) \cdot t^{\text{best}}_{\text{max}}$.
The obtained results for $\overline{\Lambda}_{\text{prompt}}$ (top) and for $\overline{\Xi}^{+}$ decays (bottom) are presented as the green curves of Fig. II.1.21 and compared to the total simulation efficiency (in blue). For both anti-hyperon categories, the decay-in-LHCb efficiency is observed to be larger than the simulation efficiency, which is expected as it accounts only one of the contributions of the latter, and mostly follows the same evolution with the momentum. Being the interactions with the detector material almost flat with the kinematics, as shown in Sec. II.2.5.2 of the exclusive approach to the analysis, the two presented approaches give consistent results.

II.1.5.2 Reconstruction and selection efficiencies

The efficiencies of the event reconstruction in the detector and of the applied selection algorithm are measured for detached (prompt) antiprotons as:

$$\varepsilon^{\text{rec \& sel}}_{\text{det (prompt)}} = \frac{N^{\text{rec \& sel}}_{\text{det (prompt)}}}{N^{\text{sim}}_{\text{det (prompt)}}}. \tag{II.1.16}$$

The candidates are counted as a function of the reconstructed antiproton momentum and transverse momentum at the numerator, of the simulated ones at the denominator. The efficiency measurement thus also takes into account the migrations among neighbour kinematic intervals because of the resolution effects in the momentum reconstruction.
The results for all antiproton categories when integrating over the full kinematic phase-space are listed in Tab. II.1.5. Large differences are found, motivating, as previously discussed, the different relative simulated and reconstructed abundances of Fig. II.1.2. For the reconstruction, the efficiency mostly depends on the number of energy deposits left by the particle in the tracker detectors, which is a function of the track pseudorapidity and ϕ starting angle, of the particle momentum, determining its propagation in the magnetic field and of the event geometry, *i.e.* the pHe collision and the antiproton OV. As an example, prompt antiprotons are produced at the PV and typically cross a large fraction of the VELO detector before entering the magnetic field. The simulated OV_z distribution for the detached ones, instead, extends by up to many metres from the nominal interaction point. Only considering the particles that can be reconstructed combining the information from the full tracking system, *i.e.* the range $\text{OV}_z < 600\,\text{mm}$, a lower efficiency is expected for detached antiprotons because of the particles with $\text{OV}_z \in [100, 600]\,\text{mm}$. This is confirmed in Fig. II.1.22, presenting the reconstruction and selection efficiencies for prompt (top) and detached (middle) antiprotons in bins of the antiproton momentum and transverse momentum and compared projected onto the two axes (bottom). Being the different efficiency values due to geometric effects, known to be accurately described by the simulation, no large systematic effects are expected. A possible residual contribution and a study of the efficiency dependences are discussed in Sec. II.1.6.5.

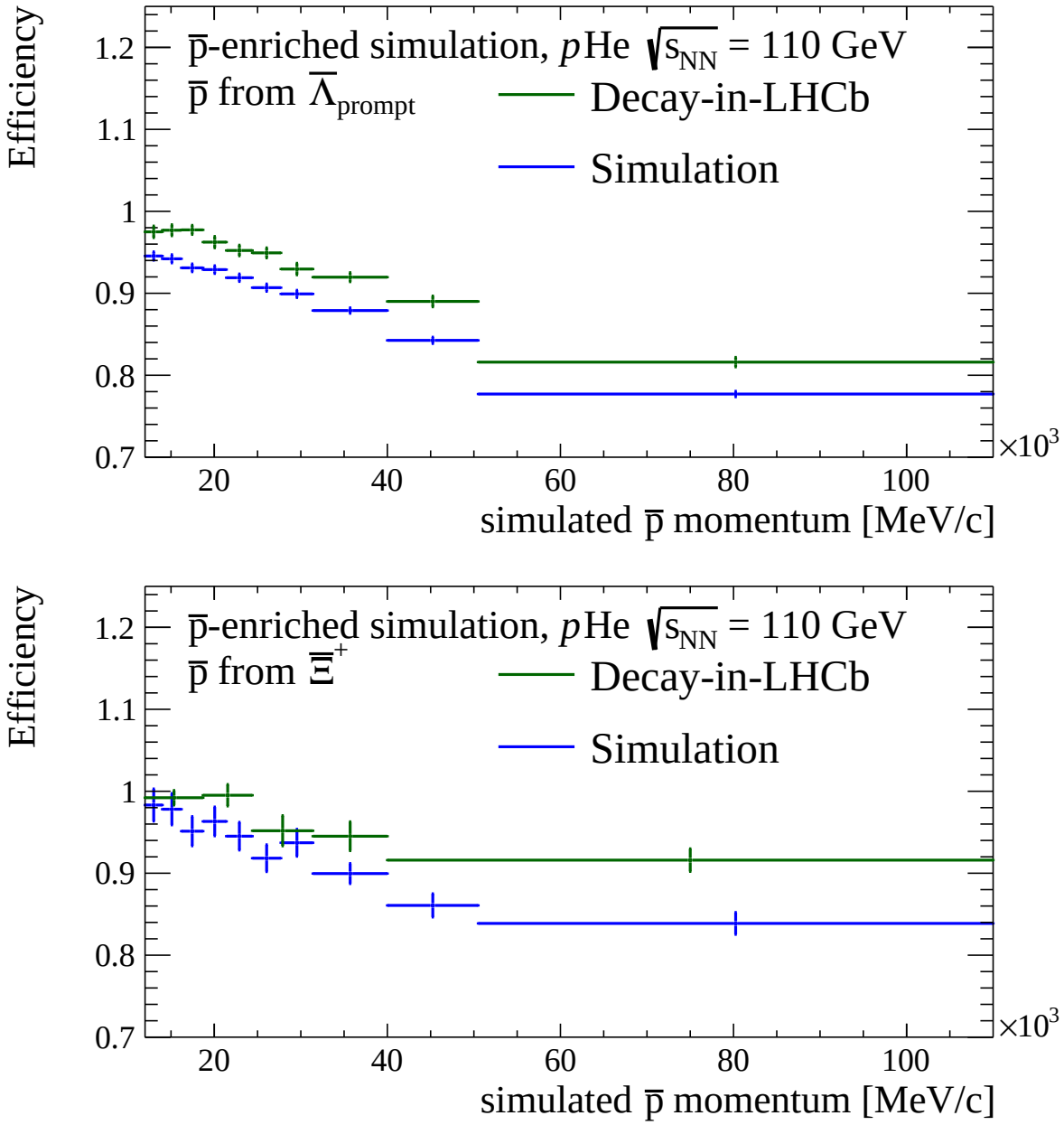


Figure II.1.21: Comparison of the simulation (in blue) and decay-in-LHCb (in green) efficiency results for antiprotons originating in promptly-produced $\bar{\Lambda}$ (top) and $\bar{\Xi}^+$ (bottom) decays as a function of the antiproton momentum in the $\bar{p}$-enriched simulated sample. The obtained values and their evolutions with the momentum confirm the two measurement strategies.

Table II.1.5: Reconstruction and selection efficiencies for all antiproton categories in the $\bar{p}$-enriched simulated sample when integrating over the full considered kinematic phase-space.

$\bar{p}$ category	efficiency [%]	$\bar{p}$ category	efficiency [%]	$\bar{p}$ category	efficiency [%]
prompt	34.59 ± 0.02	detached	11.70 ± 0.03	prompt $\bar{\Lambda}$	10.57 ± 0.03
$\bar{\Sigma}^-$	22.28 ± 0.10	$\bar{\Xi}^+$	3.00 ± 0.07	$\bar{\Xi}^0$	2.11 ± 0.07

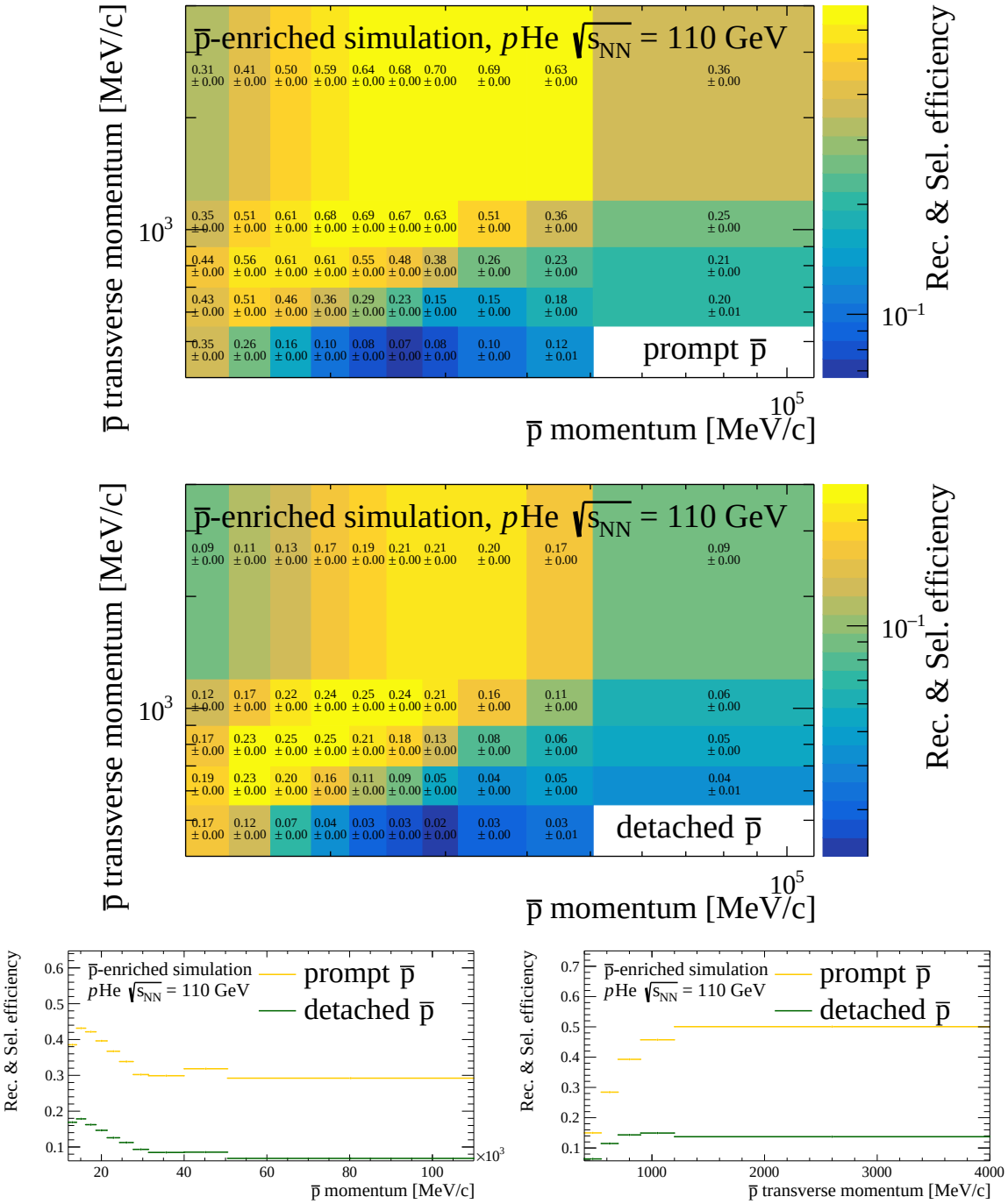


Figure II.1.22: Reconstruction and selection efficiencies for prompt (top) and detached (middle) antiprotons in the $\overline{p}$-enriched simulated sample in intervals of the antiproton kinematics. Bottom plots show the direct comparisons of the obtained values projected onto the momentum (left) and transverse momentum (right) axes.

II.1.6 Systematic uncertainties

The contributions to the measurement systematic uncertainty are scrutinised in this section with simulation-based or data-driven studies. Section II.1.6.1 focuses on the observed structure at $log(\chi^2_{\rm IP}) \sim 5$ in the template distribution for prompt antiprotons, investigating its origin and the contribution to the uncertainty due to its possible imprecise modelling in the simulation. Section II.1.6.2 and II.1.6.3 analyse the uncertainty related to imprecise assumptions of the EPOS-LHC generator for the $\overline{\Lambda}$-to-$\overline{\Sigma}^-$ and the $\overline{\Lambda}$-to-$\overline{\Xi}^+$ abundance ratios, respectively. Finally, Secs. II.1.6.4 and II.1.6.5 evaluate the PID and the reconstruction different performances between prompt and detached antiprotons.

II.1.6.1 Prompt antiproton $\chi^2_{\rm IP}$ template

As presented in Fig. II.1.4, the prompt antiproton $\chi^2_{\rm IP}$ distribution, other than a dominant peak around zero, presents a second structure for $log(\chi^2_{\rm IP}) > 3$. Being prompt antiprotons defined as those produced consistently with the pHe collision vertex, such a structure is unexpected. Moreover, overlapping this with the $\chi^2_{\rm IP}$ region mostly occupied by detached antiprotons, its imprecise simulation could bias the result of the fit. The origin of the candidates in this peaking structure is discussed in Sec. II.1.6.1.1 employing the simulated sample and is demonstrated to be due to a deviation of the prompt antiproton following a scattering with the RF foil detector material. The contribution to the systematic uncertainty is estimated in Sec. II.1.6.1.2 by comparing the simulated template with one drawn on a sample of antiprotons coming from $\overline{\Lambda}(1520) \to \overline{p}K^+$ decays reconstructed and selected in the 2017 pNe data.

II.1.6.1.1 High-$\chi^2_{\rm IP}$ prompt antiprotons

The origin of prompt antiproton candidates with $log(\chi^2_{\rm IP}) > 3$ (called tail in the following) is studied employing the $\overline{p}$-enriched simulated sample. Since the $\chi^2_{\rm IP}$ is not related to the particle species but to the track and vertex reconstructions only, the PID selection requirements are released for this study and all prompt particles are considered in order to enhance the selected sample size.
The reconstruction quality for prompt particles with $log(\chi^2_{\rm IP}) \in [-4, 2]$ (in green) and with $log(\chi^2_{\rm IP}) \in [3, 8]$ (in magenta) is firstly compared, with the results for the track (top left) and vertex (top right) fit χ^2/ndf and for the resolution on the x (bottom left) and y (bottom right) momentum components presented in Fig. II.1.23. With respect to the dominant $\chi^2_{\rm IP}$ peak, high-$\chi^2_{\rm IP}$ prompt particles are reconstructed with lower accuracy and with a larger difference with respect to the simulated momentum. The reconstruction quality requirements applied in the selection algorithm could be possibly tightened to mitigate the high-$\chi^2_{\rm IP}$ peaking structure, but this could induce a bias between detached and prompt antiprotons, considering that the former are typically reconstructed combining a lower number of energy deposits, as discussed above. Rather, a variable that is found to discriminate the prompt candidates in the two $\chi^2_{\rm IP}$ regions is:

$$p_{\rm T}\,\chi^2 = \sum_{i\,=\,x,\,y} \frac{(p_i^{\rm rec} - p_i^{\rm sim})^2}{\sigma^2_{p_i}} \tag{II.1.17}$$

being

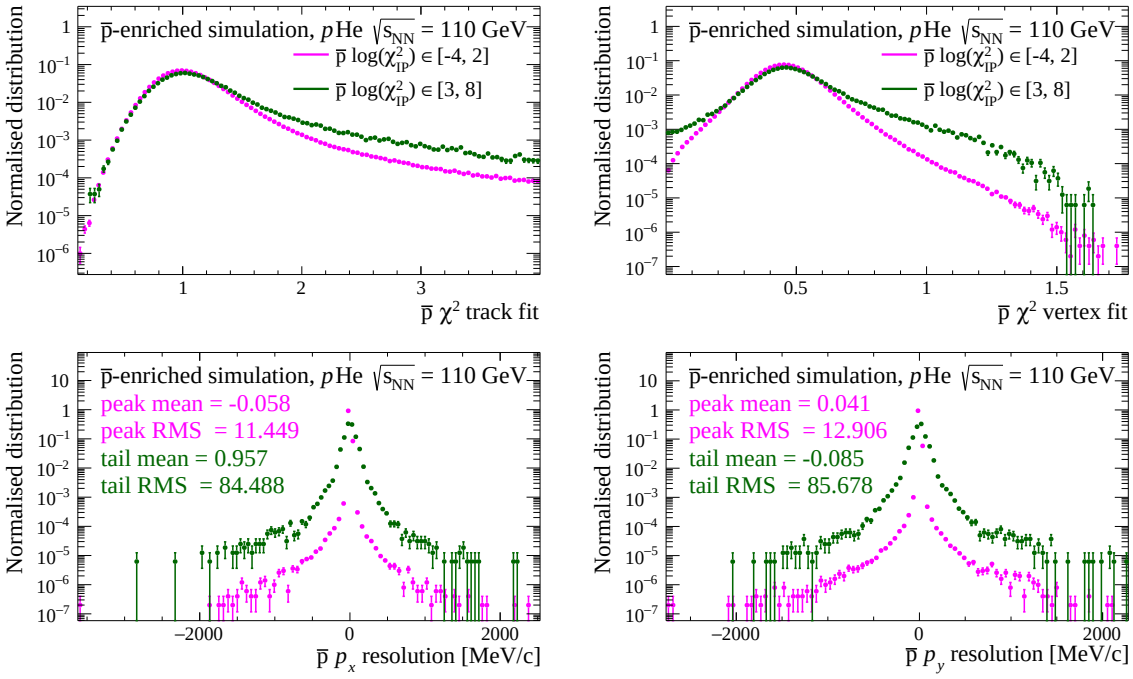


Figure II.1.23: Comparison between the χ^2/ndf distributions for the track (top left) and vertex (top right) reconstruction and between the resolutions on the x and y momentum components for low- (in green) and high-$\chi^2_{\rm IP}$ (in purple) prompt particles in the $\overline{p}$-enriched simulated sample.

- $p^{\rm rec\,(sim)}_{x\,[y]}$ the reconstructed (simulated) x $[y]$ momentum component;
- $\sigma_{p_{x\,(y)}}$ the resolution on the x (y) momentum, measured as the RMS of the resolution histograms for all prompt particles as $\sigma_{p_x} = 18.9\,{\rm MeV}/c$ ($\sigma_{p_y} = 20.0\,{\rm MeV}/c$).

The correlation between $p_{\rm T}\,\chi^2$, quantifying the reconstruction quality on the transverse momentum, and $log(\chi^2_{\rm IP})$ is illustrated in Fig. II.1.24. While the particles in the dominant $\chi^2_{\rm IP}$ peak region all populate $p_{\rm T}\,\chi^2 < 4$, a structure with $p_{\rm T}\,\chi^2$ up to 100 is observed in the high-$\chi^2_{\rm IP}$ region. Within the bidimensional plot, the two categories are defined as:

$$\begin{aligned} \text{peak candidates} &: log(\chi^2_{\rm IP}) \in [-4, 2]\ \&\ p_{\rm T}\,\chi^2 < 4 \\ \text{tail candidates} &: log(\chi^2_{\rm IP}) \in [3, 8]\ \&\ p_{\rm T}\,\chi^2 \in [4, 100] \end{aligned} \qquad \text{(II.1.18)}$$

The angular distributions for peak and tail candidates are compared in Fig. II.1.25 (with $\theta = \sqrt{p_x^2 + p_y^2}/p_z$ and $\phi = atan(|p_y|/p_x)$). For both categories low θ values are dominant and the ϕ distribution presents a depletion in $\phi \in [0, 0.5]$ rad due to the magnetic field effect. Outside this interval, the ϕ distribution is uniform for particles with low-$\chi^2_{\rm IP}$, while peaks at the ϕ boundary regions for the high-$\chi^2_{\rm IP}$ ones. By comparing this distribution with the geometry of the VELO detector, shown in Fig. I.3.6, tail events are demonstrated to accumulate where the two VELO halves are joint and the aluminium RF foils are located. Together with the larger inaccuracy of the momentum reconstruction with respect to low-$\chi^2_{\rm IP}$ particles, this suggests that the high-$\chi^2_{\rm IP}$ prompt particles are reconstructed after a scattering with the RF foils, deviating their momentum before the track detection in the VELO detector. Since the effect has a potential large influence on the result, cross-checks are needed to validate its extent.

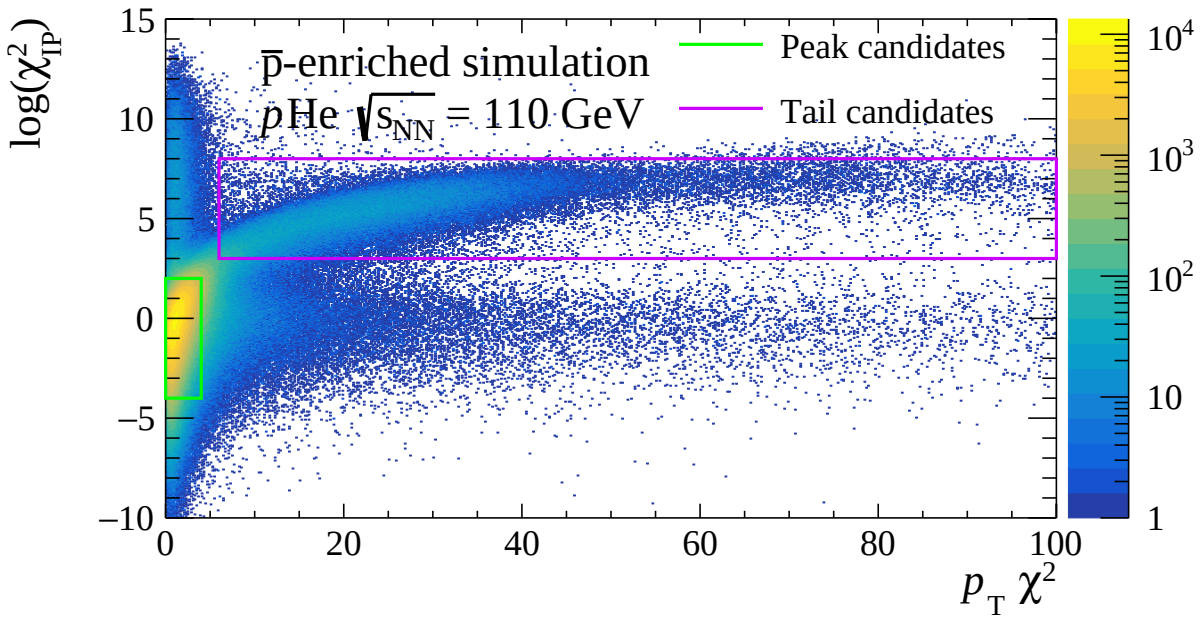


Figure II.1.24: Correlation for prompt particles in the $\overline{p}$-enriched simulated sample between the $\log(\chi^2_{\rm IP})$ and the $p_{\rm T}$ χ^2 variables. The low- and high-$\chi^2_{\rm IP}$ categories can be clearly distinguished and are highlighted by a green and a violet box, respectively.

The quality of the detector material simulation and its possible impact on the $\chi^2_{\rm IP}$ values is studied by considering in the pHe data and simulated samples the distribution of the scattered particle origin vertices. These are selected with the requirements listed in Tab. II.1.6: the same trigger requirements as the analysis are applied, but a larger region is accepted for the PV$_z$. Only events with no tracks in the backward direction are accepted and, as in the studies above, the PID selections are removed in order to increase the available statistics. All particles presenting $log(\chi^2_{\rm IP}) > 8$ and a reconstructed OV with $\chi^2/{\rm ndf} < 1.5$ are selected for this study. Figures. II.1.26 and II.1.27 compare the normalised distributions for the selected particle origin vertices. The structure of the VELO detector can be clearly seen. In the comparison between data and simulation, some differences arise, such as a structure at all z coordinates and with $|y| \simeq 5\,$mm, only present in the latter. This is expected as a consequences of the simplified detector structure that is assumed in the simulation. The effect of these differences on the $\chi^2_{\rm IP}$ is then studied and presented in Fig. II.1.28. The left plot compares the particle $\chi^2_{\rm IP}$ distributions in data and in simulation, the right one in two different y bins for the simulation. The observed consistency in both plots demonstrates that the differences in the detector material simulation do not induce distortions on the $\chi^2_{\rm IP}$ variable. Prompt particles at large $\chi^2_{\rm IP}$ are thus also to be expected in the data sample.

II.1.6.1.2 Template data-driven cross-check with $\overline{\Lambda}(1520) \to \overline{p}K^+$

In order to confirm that the high-$\chi^2_{\rm IP}$ prompt antiprotons are produced in a hadronic interaction with the RF foil material and to quantify their relative fraction in the template, which is needed to assign the related contribution to the systematic uncertainty, a data-

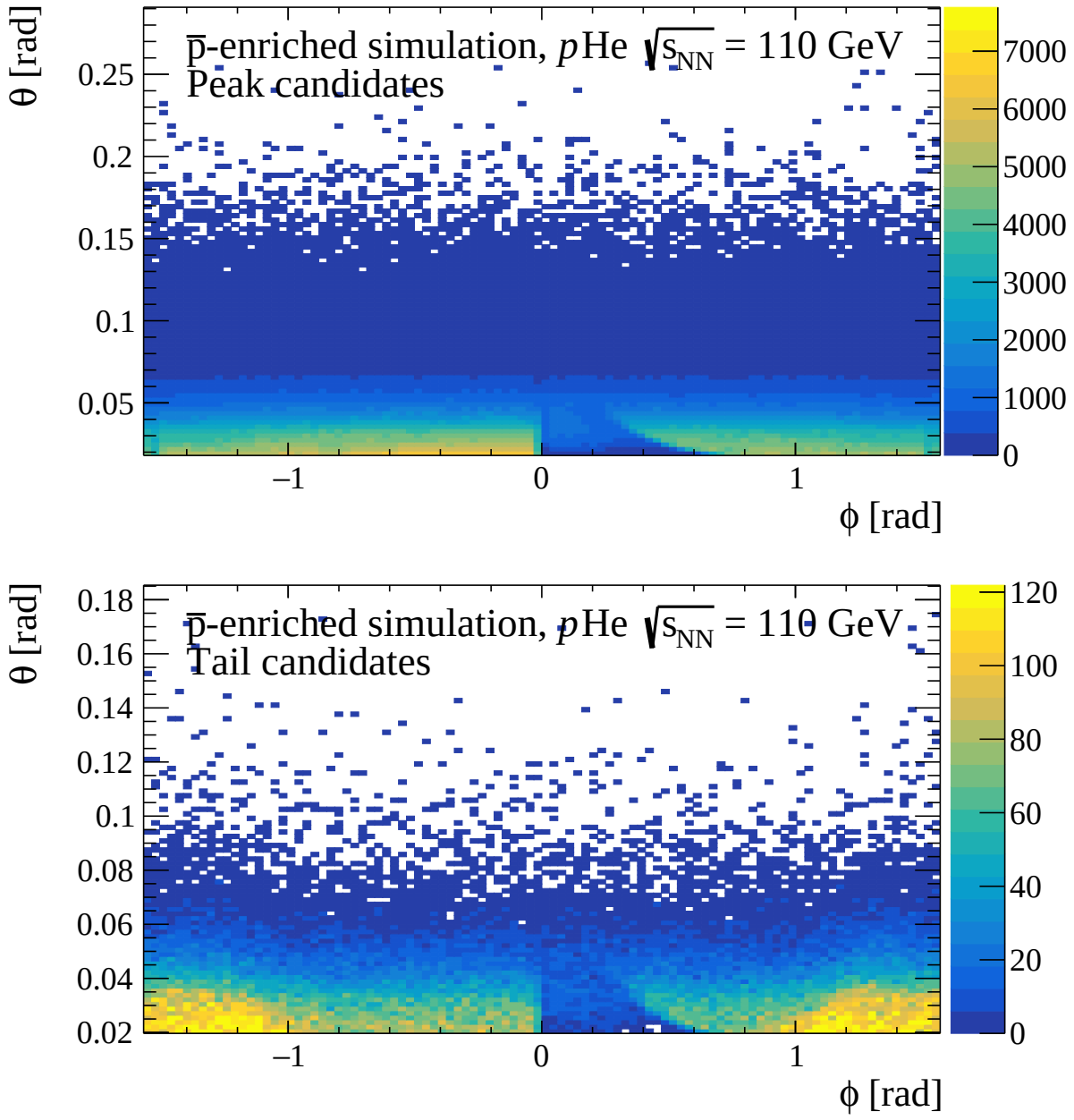


Figure II.1.25: Angular distribution for low- (top) and high-$\chi^2_{\rm IP}$ (bottom) prompt particles in the $\overline{p}$-enriched simulated sample. While in both cases low θ values are dominant, the ϕ distribution is uniform for the former and peaks at the boundary regions in the latter.

Table II.1.6: Applied cuts to pHe data and simulated samples for the selection of particles originating from interactions with the detector material.

Selection requirements
$N_{SPD} > 0$, $N_{VELO} > 0$, $N_{\rm trk}(\eta < 0) = 0$, $\mathrm{PV}_z \in [-1000, 300]$ mm OV $\chi^2/\mathrm{ndf} < 1.5$, $log(\chi^2_{\rm IP}) > 8$

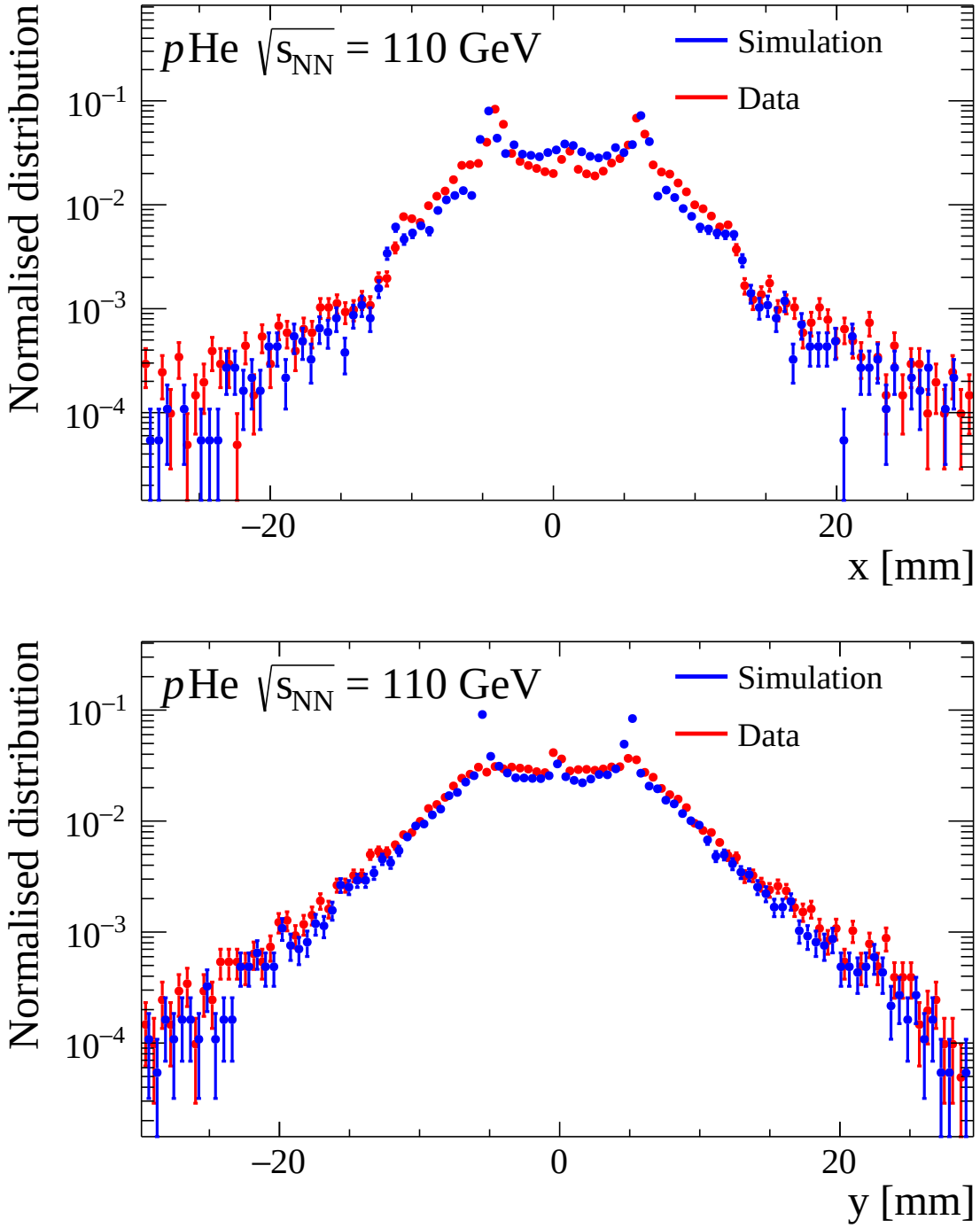


Figure II.1.26: Comparison between the normalised distributions for the origin vertex x (top) and y (bottom) coordinates for particles produced in hadronic interactions with the detector material in pHe data (in red) and in simulation (in blue). Two structures at $|y| \simeq 5\,\text{mm}$ are found to be present in the simulated but not in the data sample.

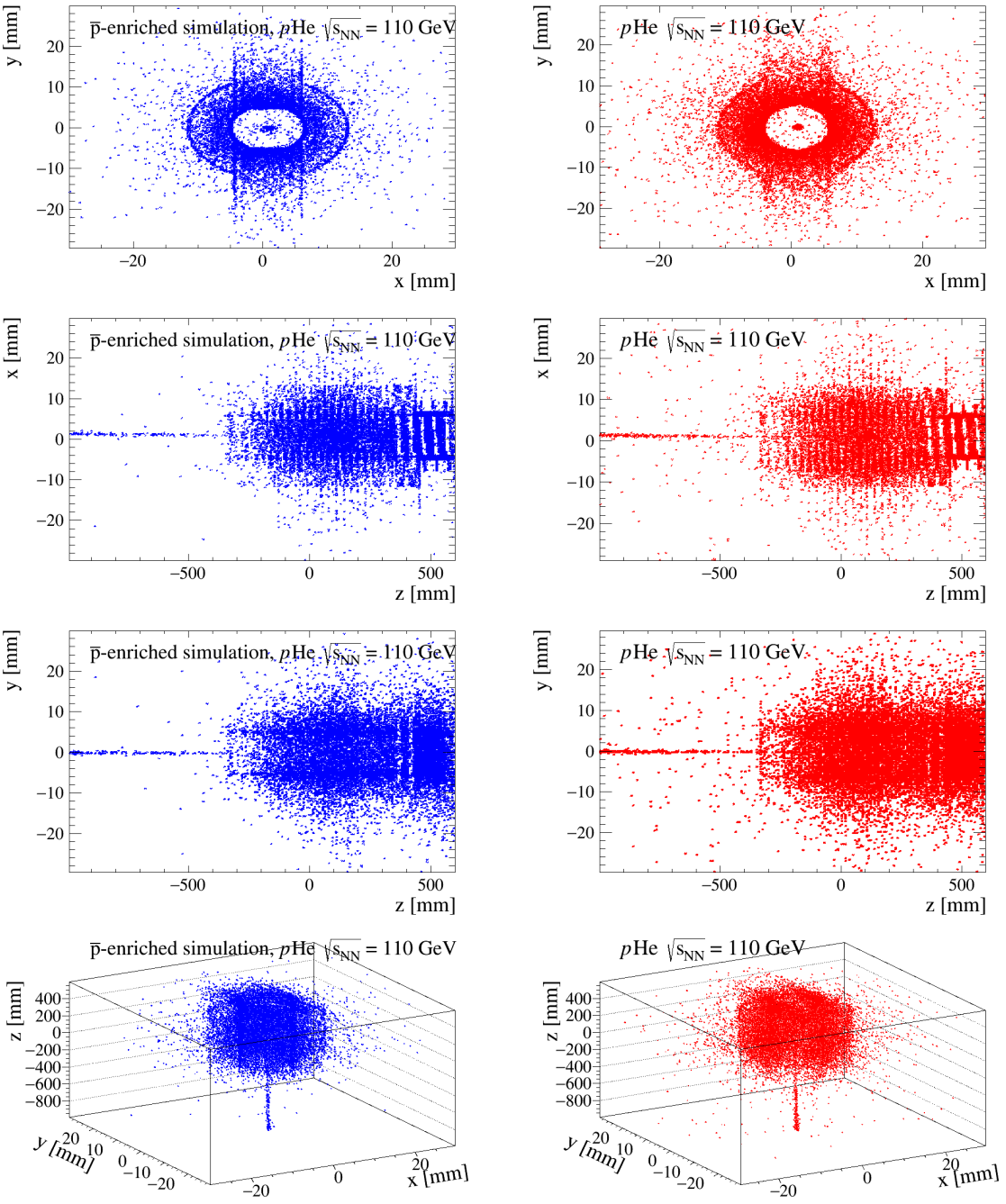


Figure II.1.27: Comparison between the distributions for (in order) the origin vertex *xy*, *xz*, *xz* and *xyz* coordinates of particles produced in hadronic interactions with the detector material in *p*He simulation (in left) and data (right).

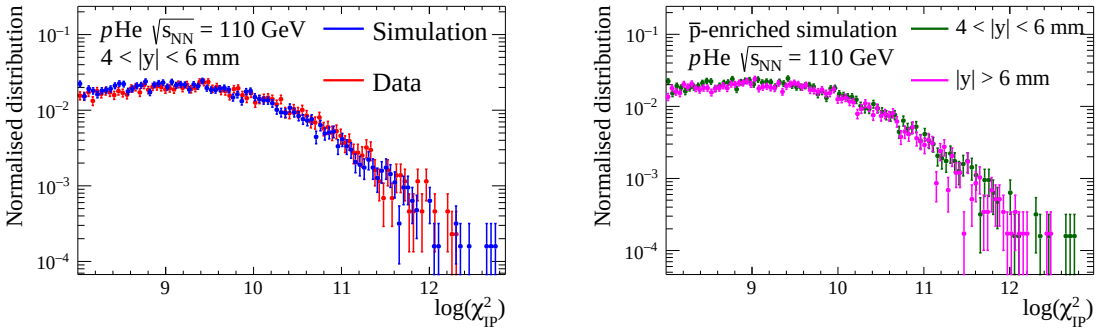


Figure II.1.28: Comparison between the $\chi^2_{\rm IP}$ distributions in pHe data and simulation restricted to the $4 < |y| < 6$ mm region (left) and between two different y regions in the simulation (right). The observed consistency between the distributions proves that the approximations on the detector material made on the simulation do not cause a distortion of the $\chi^2_{\rm IP}$ profile.

driven approach is performed. To select a sample of prompt antiprotons on data, the decay of a baryon with:

- a low enough mass to get the required statistics to study this suppressed effect;
- a strong decay with a good branching ratio towards the final state with an antiproton,

is considered. These characteristics would be offered by the $\overline{\Lambda}(1520) \to \overline{p}K^+$ decay but, unfortunately, the available statistics in the pHe sample is insufficient. The pNe sample collected in 2017, with two orders of magnitude more antiprotons on target as shown in Fig. I.4.3, is employed instead. To the interest of this cross-check, two aspects of the pNe data-taking differing from the pHe one need to be taken into account:

- the pNe sample was acquired simultaneously with the pp collisions employing the LHC bunches not colliding in LHCb. Because of the 25 ns spacing, protons can migrate between neighbour bunches and pp collisions can contaminate the nominal beam-empty configurations. The top Fig. II.1.29 presents indeed the PV$_z$ distribution for $\overline{\Lambda}$ (1520) candidates and the sum of a continuous flat evolution of the pNe collisions and a peak around the nominal pp interaction point is observed. To mitigate its extent, a selection cut requiring $N_{\rm trk}(\eta < 0) < 5$ is set. The residual pp contamination, considering that the origin of the $\overline{\Lambda}$ (1520) particle is not of interest for this study and that weights will be applied to the pNe sample to reproduce the pHe distributions of the relevant variables, is not expected to affect the results;
- the trigger acquisition of $\overline{\Lambda}$ (1520) candidates in the pNe sample differs from the pHe one. Indeed, given the increase in the output data rate arising from the simultaneous data-taking strategy, the minimum-bias trigger selection, employed for the pHe data sample, was only acquired in the pNe one for one event over 20. Most of the $\overline{\Lambda}$ (1520) particles are instead selected by inclusive lines with thresholds on the particle momentum and transverse momentum. The resulting kinematic spectra, as illustrated in bottom Fig. II.1.29, are the sum of two contributions;

The selection algorithm for $\overline{\Lambda}$ (1520) decays in the pNe data is described in Tab. II.1.7. In the preselection, applied during the LHCb data filtering, the signal decay is defined. Good-quality reconstructed particles with $p > 12\,{\rm GeV}/c$ and $p_{\rm T} > 380\,{\rm MeV}/c$ are considered

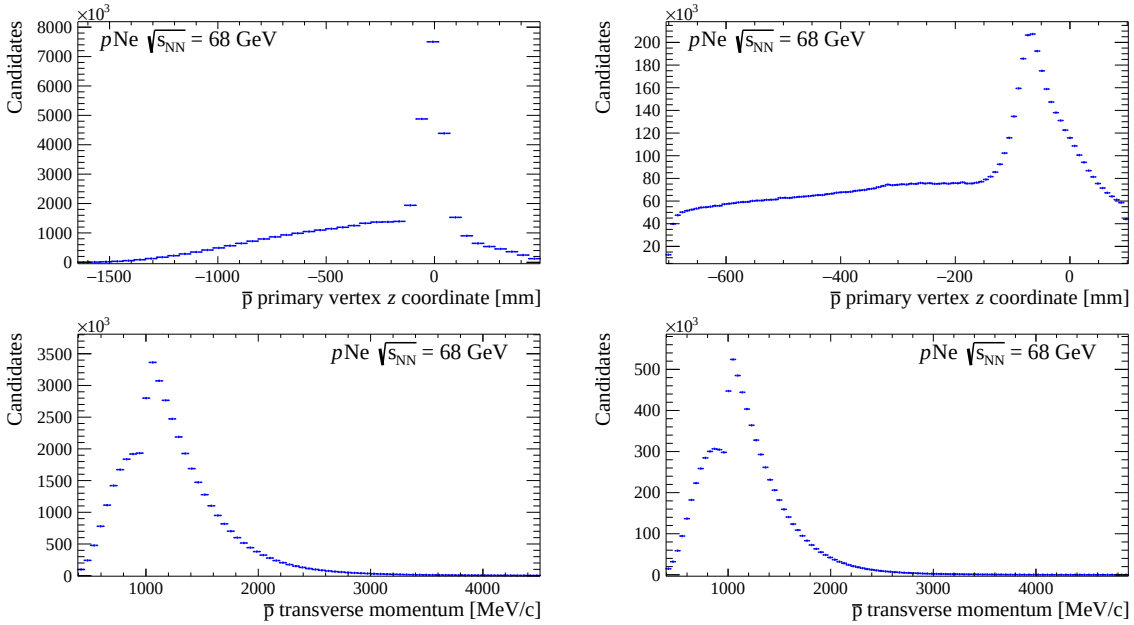


Figure II.1.29: Distributions of the antiproton candidates z coordinate of the primary vertex (top) and of transverse momentum (bottom) in the 2017 pNe data sample before (left) and after (right) the application of the offline selection cuts.

Table II.1.7: Selection algorithm applied to the 2017 pNe data for $\overline{\Lambda}(1520) \to \overline{p}K^+$ decays.

Selection	Particle	Selection cuts
Preselection	$\overline{p}$, K^+	$p > 12\,\text{GeV}/c$, $p_\text{T} > 380\,\text{MeV}/c$
		trk $\chi^2/\text{ndf} < 5, GhostProb < 0.15$
	$\overline{p}$	$\text{DLL}_{p,\pi} > 20$, $\text{DLL}_{p,K} > 10$
	K^+	$\text{DLL}_{K,\pi} > 10$, $\text{DLL}_{p,K} < 0$
	Combination	$M < 2\,\text{GeV}/c^2$
	$\overline{\Lambda}$ (1520)	$M < 1.56\,\text{GeV}/c^2$, $DV\ \chi^2/\text{ndf} < 16$, $\text{DV}_z < 2200\,\text{mm}$
Offline	Event	$N_\text{trk}(\eta < 0) < 5$
	$\overline{p}$, K^+	$\text{PV}_z \in [-700, 100]\,\text{mm}$, $\text{OV}_z \in [-700, 700]\,\text{mm}$
	$\overline{p}$	$p_\text{T} > 400\,\text{MeV}/c$, $\eta < 5$
	K^+	$\text{DLL}_{K,\pi} > 20$, $log(\chi^2_\text{IP}) < 3$
	Combination	$\lvert M - M(\overline{\Lambda}(1520))\rvert < 500\,\text{MeV}/c^2$
	$\overline{\Lambda}$ (1520)	$\lvert M - M(\overline{\Lambda}(1520))\rvert < 100\,\text{MeV}/c^2$

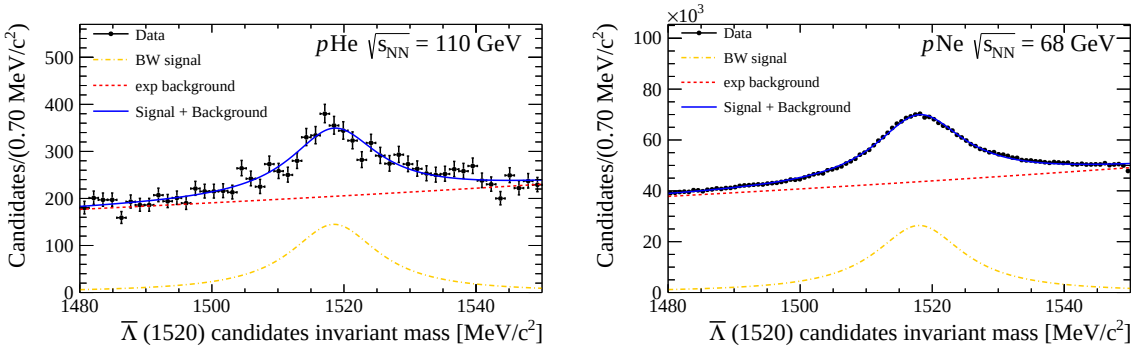


Figure II.1.30: $\overline{\Lambda}^*$ candidates invariant mass distribution in the pHe (left) and in the pNe (right) data samples. Both distributions are modelled with the sum of a Breit-Wigner for the signal and an exponential function for the background.

and tight cuts on the PID variables are set in order to select antiprotons and kaons. The two-track combinations are formed, the parent-particle invariant mass is evaluated according to Eq. I.3.5 and required to be lower than 1.56 GeV/c^2. The $\overline{\Lambda}$ (1520) Decay Vertex (DV) is required to be fitted with a good χ^2/ndf and to be placed well before the magnet z coordinate. In the offline selection step, a more detailed description of the signal decay is performed. A maximum of five reconstructed tracks with negative pseudorapidity is required, in order to mitigate the effect of the spurious pp collisions. Kinematic and geometric requirements are tightened up to match those in Sec. II.1.3. To further enhance the statistics without largely distorting the distributions for the antiproton, the kaon is required to be promptly-produced with $log(\chi^2_{\rm IP}) < 3$ and a tighter PID cut is set. The accepted invariant mass interval around the nominal $\overline{\Lambda}$ (1520) value is reduced, yet leaving enough background events to define a functional form for their parametrization.
In order to compare the prompt antiproton templates drawn in the pHe and in the pNe samples, mitigating the effect of the two different acquisition strategies, weights are applied to the former. The *sPlot* technique [130], evaluating a probability for each event to belong to a signal or a background hypothesis, is used. The invariant mass distributions of the $\overline{\Lambda}$ (1520) candidates in the pHe (left) and the pNe (right) samples are modelled with the sum of a Breit-Wigner for the signal and an exponential function for the background. The results are presented in Fig. II.1.30 and exemplify the different size of the selected statistics in the two samples. With the defined functional forms for the invariant mass of the signal and background $\overline{\Lambda}$ (1520) candidates, weights for the former are evaluated to be applied, according to the general prescriptions of the method, to all other variables not correlated with the mass. This is the case to a good extent of the kinematic and geometry distributions, that are compared normalised between the pHe and pNe samples and divided to obtain weights. Due to the scarce available statistics in the former sample, a multi-dimensional comparison is not possible and weights for the antiproton transverse momentum and for the number of hits in the SPD detector are only considered.

In order to draw the $\chi^2_{\rm IP}$ distribution for the antiprotons produced in $\overline{\Lambda}$ (1520) decays, the *sPlot* technique cannot be directly applied because of its correlation with the $\overline{\Lambda}$ (1520) invariant mass. This is demonstrated by Fig. II.1.31, presenting the $\overline{\Lambda}$ (1520) candidates invariant mass distributions in two far antiproton $\chi^2_{\rm IP}$ regions. For high-$\chi^2_{\rm IP}$ antiprotons, where an interaction with the detector material occurs, the resolution on the invariant mass worsens. A different strategy to draw the template is thus adopted. After the

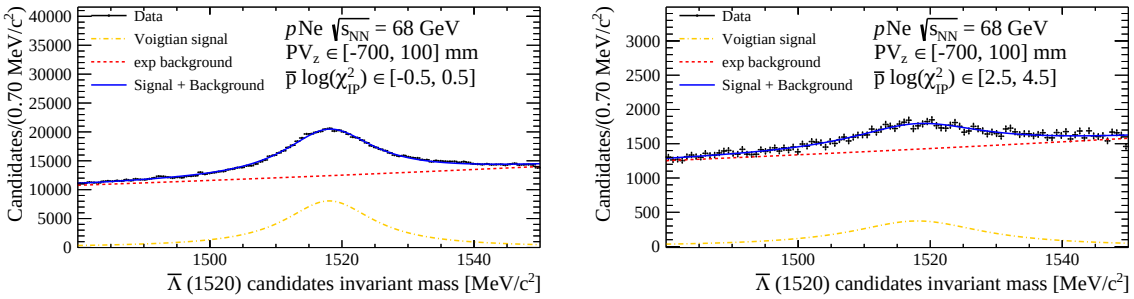


Figure II.1.31: $\bar{\Lambda}(1520)$ candidates invariant mass distribution in the pNe sample in two far $\bar{p}$ $\chi^2_{\rm IP}$ regions. Both are modelled with the convolution of a Breit-Wigner and a Gaussian distribution for the signal and an exponential function for the background. Because of the hadronic interactions with the detector material, a degradation of the invariant mass resolution for large $\chi^2_{\rm IP}$ values is found.

application of the weights, candidates are divided into intervals for the antiproton $\chi^2_{\rm IP}$ and a fit to the $\bar{\Lambda}$ (1520) invariant mass distribution is performed in each interval. By iterating over the full $\bar{p}$ $\chi^2_{\rm IP}$ range, the distribution for the fitted number of signal events provides the data-driven template to be compared with the pHe simulation. The employed model is a Voigtian distribution for the signal, the convolution of a Breit-Wigner and a Gaussian, and an exponential function for the background. The fitted number of signal events as a function of the $\bar{p}$ $\chi^2_{\rm IP}$ is shown as the blue marker points in Fig. II.1.32 and compared with the simulated template for prompt antiprotons in the pHe sample. An overall good consistency between the two distributions is found, confirming that the high-$\chi^2_{\rm IP}$ prompt antiproton structure is correctly modelled in the simulation. The template drawn with the pNe data is modelled according to the histogram model and substituted to the simulated pHe prompt antiproton one in the fit to the pHe data. The results are presented in Fig. II.1.33 and give for the detached-to-prompt $\bar{p}$ yield ratio:

$$R^{\rm Data}_{\rm raw} = 0.1305 \pm 0.0005. \qquad \text{(II.1.19)}$$

By comparing with the measurement in Eq. II.1.9, a 4.4% relative difference, assigned as the systematic uncertainty, is found.

II.1.6.2 Simulated $\bar{\Lambda}$-to-$\bar{\Sigma}^-$ ratio

Another source of uncertainty that is taken into account is a possible imprecise assumption for the relative $\bar{\Lambda}$-to-$\bar{\Sigma}^-$ production in the simulation, implicitly assumed in:

- the drawing of the $\chi^2_{\rm IP}$ template for detached antiprotons, obtained from the $\bar{\Lambda}$ and $\bar{\Sigma}^-$ decays weighted according to the predicted abundances in EPOS-LHC;
- the detached reconstruction and selection and the simulation efficiency evaluations, that are obtained from the combined samples of antiprotons originating in $\bar{\Lambda}_{\rm prompt}$, $\bar{\Sigma}^-$, $\bar{\Xi}^+$ and $\bar{\Xi}^0$ decays.

To measure on data the $\bar{\Lambda}$-to-$\bar{\Sigma}^-$ ratio, templates for the two anti-hyperon categories are separately drawn and included in the template combination to be fitted to data.

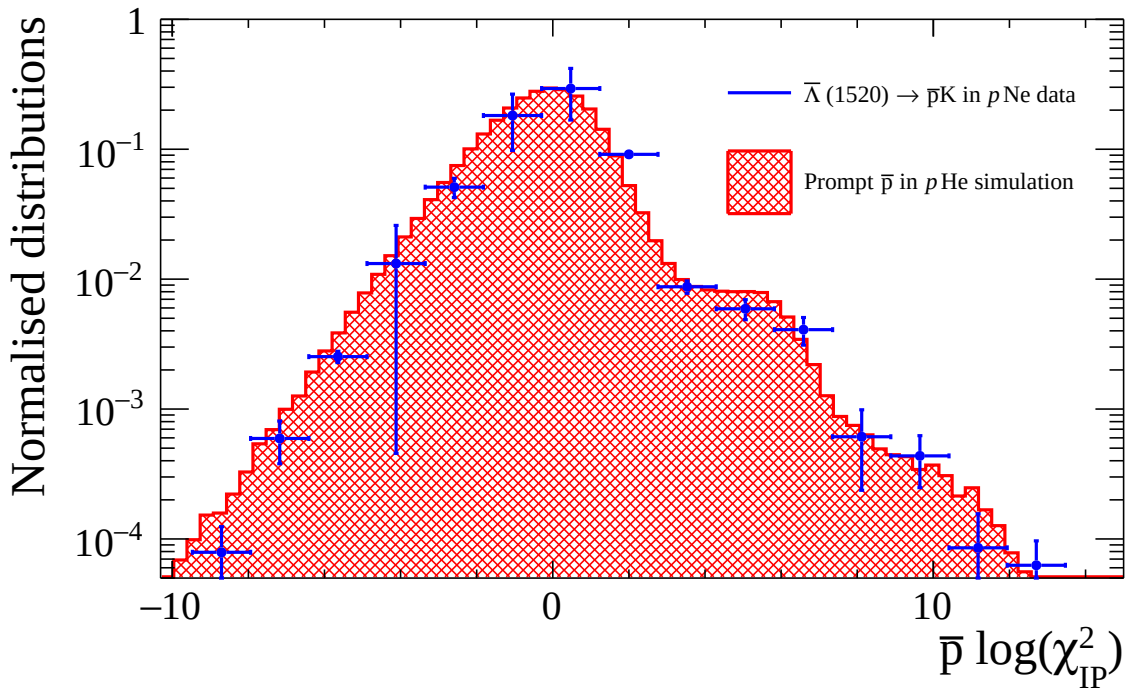


Figure II.1.32: Comparison between the normalised prompt antiproton $\chi^2_{\rm IP}$ template in the pHe simulation and that obtained from the pNe data as explained in the text. The overall match between the two confirms a correct modelling of high-$\chi^2_{\rm IP}$ antiprotons in the simulation.

Considering Eq. II.1.6, the free parameters for this fit are $f_{\rm prompt\bar{p}}$, $f_{\bar{p}\ \rm from\ \bar{\Lambda}}$, $f_{\bar{p}\ \rm from\ \bar{\Sigma}^-}$ and $f_{\rm scattered\bar{p}}$. Figure II.1.34 shows the result to the pHe data integrated over the full kinematic phase-space with the histogram (top) and parametric (bottom) models. The templates for the two anti-hyperon categories mostly cover the same $\chi^2_{\rm IP}$ region, as expected. The $\overline{\Sigma}^-$ template averages to slightly larger values than the $\overline{\Lambda}$ particle because of the higher difference between its mass and the sum of the final-state particles, causing on average a larger deviation of the antiproton with respect to the parent particle flight direction. The raw detached-to-prompt yield ratio is:

$$R^{\bar{\Lambda}+\bar{\Sigma}^-}_{\rm raw} = \frac{N^{\rm Data}_{\bar{\Lambda}} + N^{\rm Data}_{\bar{\Sigma}^-}}{N^{\rm Data}_{\rm prompt}} = 0.1255 \pm 0.0008 \qquad \text{(II.1.20)}$$

with the histogram and

$$R^{\bar{\Lambda}+\bar{\Sigma}^-}_{\rm raw} = \frac{N^{\rm Data}_{\bar{\Lambda}} + N^{\rm Data}_{\bar{\Sigma}^-}}{N^{\rm Data}_{\rm prompt}} = 0.1254 \pm 0.0008 \qquad \text{(II.1.21)}$$

with the parametric model. These are found to be consistent and, by comparing with the results in Eq. II.1.9 and II.1.10, $\sim 0.4\%$ larger than the value measured fixing the $\overline{\Lambda}$-to-$\overline{\Sigma}^-$ ratio to the EPOS-LHC expectation. Within the statistical uncertainty, they are still consistent. When comparing the $\overline{\Lambda}$-to-$\overline{\Sigma}^-$ ratio prediction in the minimum-bias simulated sample with the fit results, instead, a 18.8% difference is observed, as reported in Tab. II.1.8. This is consistent with the expected theoretical uncertainty on the quantity. The effect of the $\overline{\Sigma}^-$ underestimation on the measurement of the simulation and reconstruction and

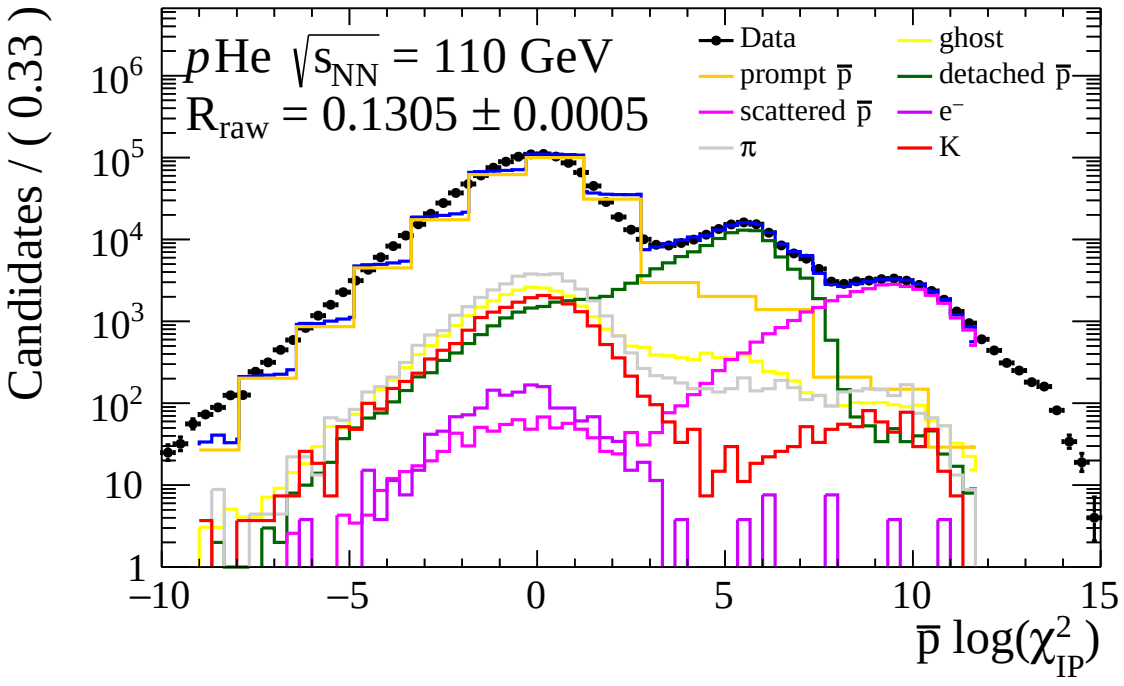


Figure II.1.33: $\chi^2_{\rm IP}$ distribution in pHe data modelled with the sum of all templates drawn on the pHe simulation except that for prompt antiprotons, obtained with the fit procedure applied to the pNe data explained in the text. The obtained $R_{\rm raw}$ value, indicated in the plot, is compared to Eq. II.1.9 to assign a systematic uncertainty due to the mis-modelling of the high-$\chi^2_{\rm IP}$ prompt antiprotons.

Table II.1.8: Prediction in the minimum-bias simulated sample and results for the fit applied to the data samples for antiproton abundances originating from $\bar{\Lambda}$ and $\bar{\Sigma}^-$ decays.

	$\bar{p}$ from $\bar{\Lambda}$	$\bar{p}$ from $\bar{\Sigma}^-$	$\bar{\Sigma}^-/\bar{\Lambda}$	$(\bar{\Sigma}^-/\bar{\Lambda})^{\rm Data}/(\bar{\Sigma}^-/\bar{\Lambda})^{\rm MC}$
MC prediction	60287	29309	0.486 ± 0.003	− − −−
Hist. model	81665 ± 588	47159 ± 570	0.577 ± 0.008	1.188 ± 0.018
Par. model	81731 ± 589	46999 ± 572	0.575 ± 0.008	1.183 ± 0.018

selection efficiencies is then studied. Their measurement is repeated introducing a weight to enhance the $\bar{\Sigma}^-$ statistics such that:

$$\varepsilon^{\rm sim}_{\rm det} = \frac{\Sigma_{i=\bar{\Lambda}_{\rm prompt},\,\bar{\Sigma}^-,\,\bar{\Xi}^+,\,\bar{\Xi}^0}\; w_i \cdot N^{\rm MC}_{{\rm gen}\,i} \cdot \varepsilon^{\rm sim}_i}{\Sigma_{i=\bar{\Lambda}_{\rm prompt},\,\bar{\Sigma}^-,\,\bar{\Xi}^+,\,\bar{\Xi}^0}\; w_i \cdot N^{\rm MC}_{{\rm gen}\,i}} \tag{II.1.22}$$

$$\varepsilon^{\rm rec\ \&\ sel}_{\rm det} = \frac{\Sigma_{i=\bar{\Lambda}_{\rm prompt},\,\bar{\Sigma}^-,\,\bar{\Xi}^+,\,\bar{\Xi}^0}\; w_i \cdot N^{MC}_{{\rm sim}\,i} \cdot \varepsilon^{\rm rec\ \&\ sel}_i}{\Sigma_{i=\bar{\Lambda}_{\rm prompt},\,\bar{\Sigma}^-,\,\bar{\Xi}^+,\,\bar{\Xi}^0}\; w_i \cdot N^{\rm MC}_{{\rm sim}\,i}}, \tag{II.1.23}$$

being $w_{\bar{\Sigma}^-} = 1.188$ and equal to unity elsewhere. Figures II.1.35 and II.1.36 show the resulting percent variation of the simulation and reconstruction and selection efficiencies, respectively. For the former, the effect is in most bins smaller than 0.2% and slightly

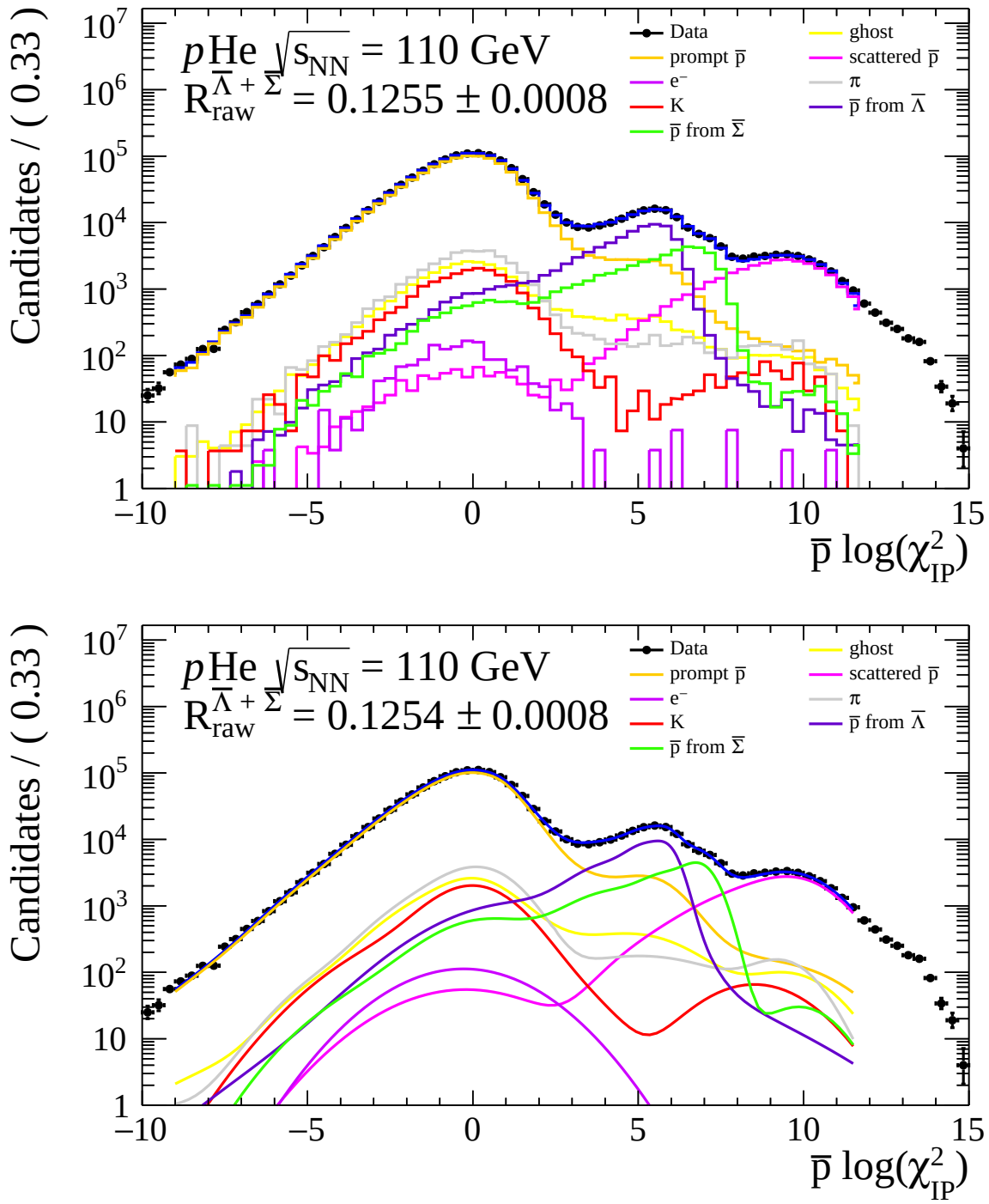


Figure II.1.34: Modelling of the χ^2_{IP} distribution in the pHe data sample according to the histogram (top) and the parametric (bottom) models with the abundances of prompt, scattered and $\bar{\Lambda}$- and $\bar{\Sigma}^-$-produced antiprotons as free parameters. The resulting R_{raw} values are indicated.

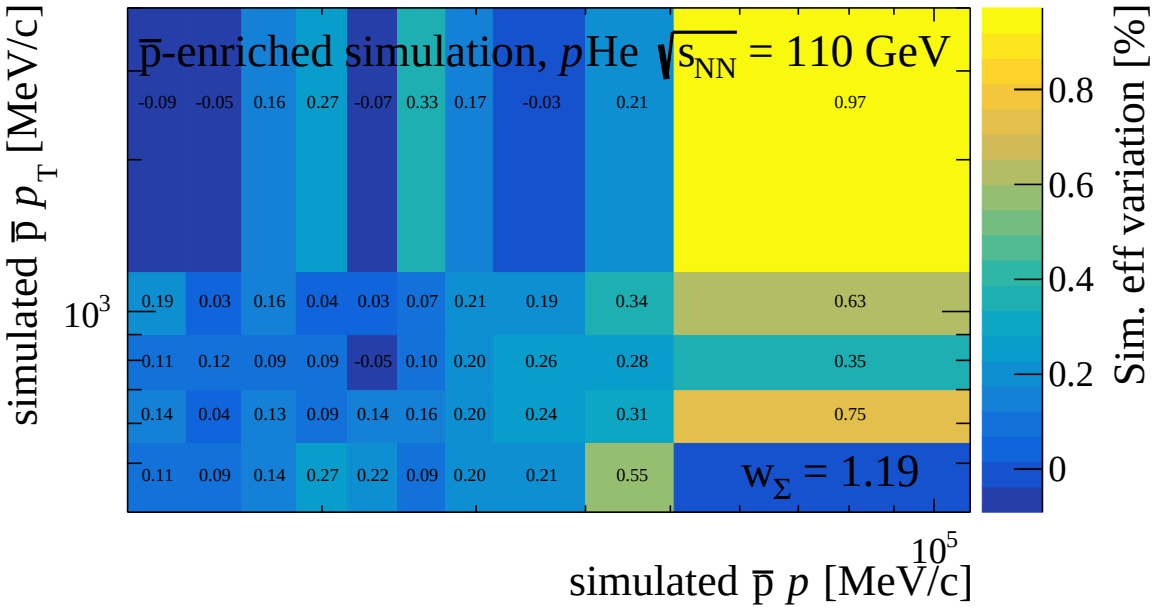


Figure II.1.35: Simulation efficiency per-cent variation for detached antiprotons in the $\overline{p}$-enriched simulated sample when enhancing the simulated $\overline{\Sigma}^-$ abundance by a factor 1.188, as measured with the fit to pHe data described in the text.

increases with the momentum as expected by the larger difference between the $\overline{\Lambda}$ and $\overline{\Sigma}^-$ simulation efficiencies. With respect to other sources of uncertainty, it can be neglected. For the latter, the variation ranges between $\sim 1.4\%$ and $\sim 4\%$ and is assigned to the systematic uncertainty.

II.1.6.3 Simulated $\overline{\Xi}^+$-to-$\overline{\Lambda}$ ratio

Similarly to the previous section, the sensitivity to the fraction of $\overline{\Lambda}$ particles originating in a $\overline{\Xi}$ decay is also studied. In this case, considering that the separation between the $\chi^2_{\rm IP}$ templates for antiprotons originating in promptly-produced $\overline{\Lambda}$ and in $\overline{\Xi}$ decays is not sufficient to be distinguished with the fit to the pHe data, a different strategy is followed. Thanks to the $\overline{\Xi}^+$ invariant mass constraint and to the efficient reconstruction of the final-state particles, a selection algorithm for $\overline{\Xi}^+ \to (\overline{\Lambda} \to \overline{p}\pi^+)\pi^+$ decays is implemented. The numbers of selected signal candidates in the pHe simulated and data samples are determined by performing a fit to the $\overline{\Xi}^+$ invariant mass distribution and compared with the selection of promptly-produced $\overline{\Lambda}$ particles developed for the exclusive approach to the analysis and described in Sec. II.2.3. The $\overline{\Xi}^+$-to-$\overline{\Lambda}_{\rm prompt}$ ratio correction factor:

$$\frac{(\overline{\Xi}^+/\overline{\Lambda}_{\rm prompt})^{\rm Data}}{(\overline{\Xi}^+/\overline{\Lambda}_{\rm prompt})^{\rm MC}} \tag{II.1.24}$$

is then measured. The simulated sample enriched with $\overline{\Lambda}_{\rm prompt}$ baryons that is used in the exclusive approach to the analysis (called $\overline{\Lambda}$-MC in the following) does not offer enough statistics for the $\overline{\Xi}^+$ decays. Another simulated sample (called $\overline{\Xi}^+$-MC) with $\sim 1M$ events is thus produced. For this, generated minimum-bias events are filtered requiring

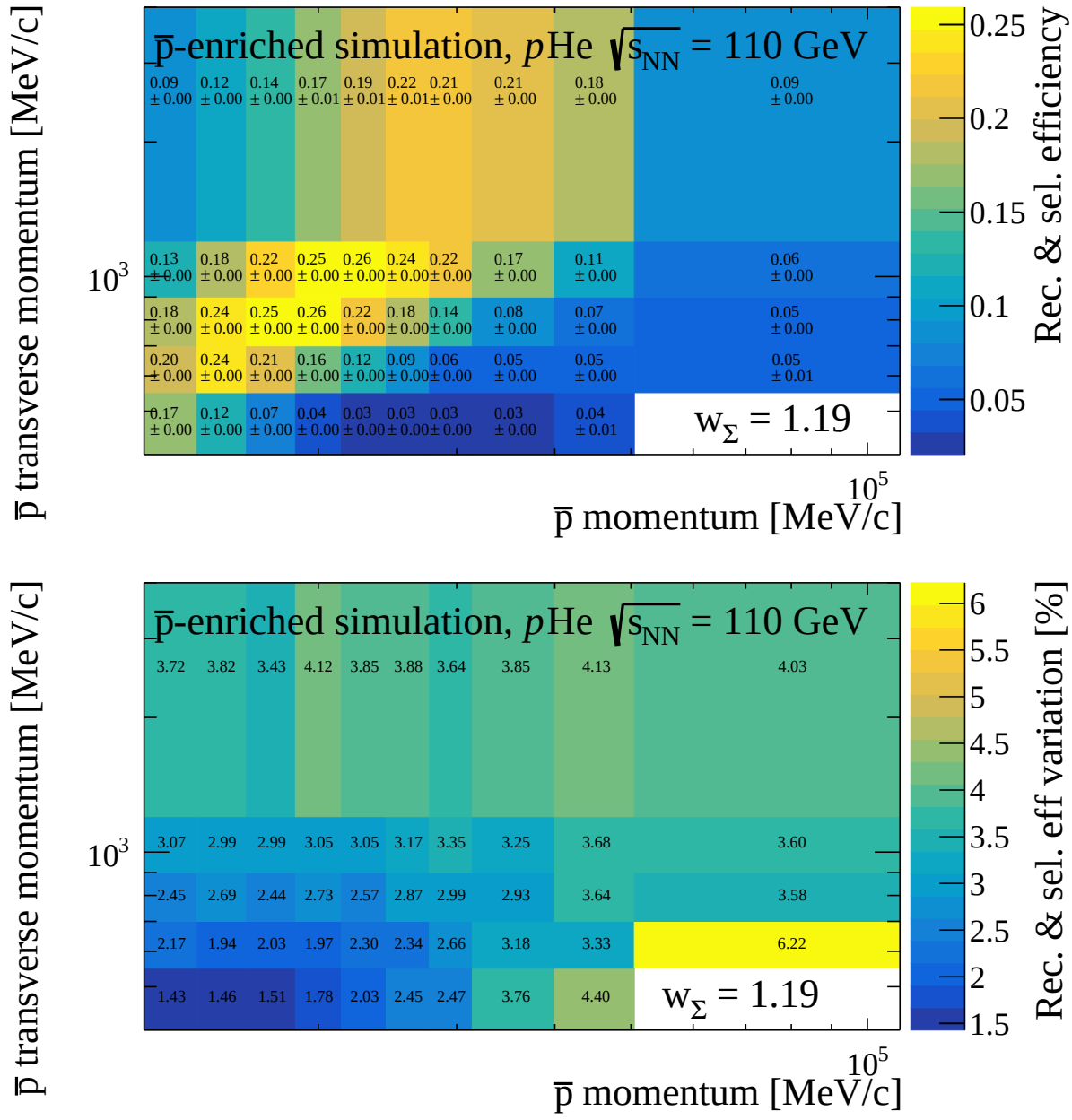


Figure II.1.36: Reconstruction and selection efficiencies for detached antiprotons in the $\overline{p}$-enriched simulated sample in bins of their simulated momentum and transverse momentum when applying the weight enhancing the $\overline{\Sigma}^{-}$ statistics (top) and percent variations with respect to the obtained values when fixing the $\overline{\Lambda}$-to-$\overline{\Sigma}^{-}$ ratio to the EPOS-LHC prediction (bottom).

Table II.1.9: Applied selection algorithm for $\overline{\Xi}^+ \to \overline{\Lambda}\pi^+$ decays in pHe data and simulation.

Prefilters			$N_{SPD} > 0$, $N_{VELO} > 0$ $N_{\text{trk}}(\eta < 0) < 5$, $\text{PV}_z \in [-700, 100]\,\text{mm}$
$\overline{\Lambda}$	$\overline{p}$, π^+	:	track fit $\chi^2/\text{ndf} < 5$, $2 < \eta < 6$
	$\overline{p}$	:	$p > 12\,\text{GeV}/c$, $p_\text{T} > 400\,\text{MeV}/c$
	π^+	:	$p > 1\,\text{GeV}/c$
	Comb.	:	$M \in [0.9, 1.5]\,\text{GeV}/c^2$, $\chi^2_{DOCA} < 10$
	$\overline{\Lambda}$	:	$\lvert M - M(\overline{\Lambda})\rvert < 30\,\text{MeV}/c^2$, DV $\chi^2 < 16$, $\lvert DV_z\rvert < 1000\,\text{mm}$
	Fisher	:	$log(IP(\overline{p})) + log(IP(\pi^+)) - log(IP(\overline{\Lambda})) > 0$
$\overline{\Xi}^+$	π^+	:	track fit $\chi^2/\text{ndf} < 5$, $p > 1.5\,\text{GeV}/c$, $p_\text{T} > 50\,\text{MeV}/c$
	Comb.	:	$1200 < M < 1500\,\text{MeV}/c^2$, $\chi^2_{DOCA} < 10$
	$\overline{\Xi}^+$	:	$\lvert M - M(\overline{\Xi}^+)\rvert < 40\,\text{MeV}/c^2$, DV $\chi^2 < 16$, $\lvert DV_z\rvert < 1000\,\text{mm}$
Offline	$\overline{p}$	:	$GhostProb < 0.15$
	π^+	:	$\text{OV}_z \in [-600, 400]\,\text{mm}$
	$\overline{\Lambda}$	:	$\text{DV}_z \in [-450, 400]\,\text{mm}$, $log(\chi^2_{\text{IP}}) > 0$, $cos(DIRA) > 0$
		:	$\left\lvert\left(\frac{\alpha+0.69}{0.18}\right)^2 + \left(\frac{p_\pi \cdot sin\theta_{\pi\overline{\Lambda}}}{100.4\,\text{MeV}/c}\right)^2 - 1\right\rvert < 0.39$
	$\overline{\Xi}^+$	:	$FD > 50\,\text{mm}$, $log(\chi^2_{\text{IP}}) < 4$, $log(cos(DIRA)) > -4 \cdot 10^{-5}$
		:	$log(IP(\overline{\Lambda})) + log(IP(\pi^+)) - log(IP(\overline{\Xi}^+)) > 0$

at least one $\overline{\Xi}^+ \to (\overline{\Lambda} \to \overline{p}\pi^+)\pi^+$ decay. In order to exclude biases in the comparison with the $\overline{\Lambda}$-MC sample, no other requirements than those described in Sec. II.2.2 are set (except for the $\overline{\Lambda}$ origin in a $\overline{\Xi}^+$ decay). The corresponding fraction of filtered events is $\varepsilon^{\overline{\Xi}^+\text{-MC}}_{\text{generator}} = (94.46 \pm 0.16) \cdot 10^{-5}$. Comparing the number of generated events N_{evts} and the filtering fraction of the prompt exclusive $\overline{\Lambda}$ production, listed in Tab. II.2.2, the ratio between the $\overline{\Xi}$ candidates in the two productions is:

$$r = \frac{N^{\overline{\Xi}^+\text{-MC}}_{\text{evts}} / \varepsilon^{\overline{\Xi}^+\text{-MC}}_{\text{generator}}}{N^{\overline{\Lambda}\text{-MC}}_{\text{evts}} / \varepsilon^{\overline{\Lambda}\text{-MC}}_{\text{generator}}} = 2.77, \tag{II.1.25}$$

assuming that the $\overline{\Xi}^+$ candidates other than those selected by the generator-level cuts are a negligible fraction.

The selection algorithm applied to enhance the $\overline{\Xi}^+$ candidates is performed in two steps, summarised in Tab. II.1.9. After the application of the preselection requirements, the trigger selection and the considered PV_z range, the signal decay chain is defined. Reconstructed tracks within the LHCb pseudorapidity acceptance and with a track fit $\chi^2/\text{ndf} < 5$ are associated to the antiproton and pion hypotheses. For the former, the same kinematic phase-space than the selection in Sec. II.1.3 is applied; for the latter, a softer momentum threshold is used. The two-track combinations between all selected tracks are then considered when their Distance Of Closest Approach (DOCA) divided by the uncertainty of its measurement (χ^2_{DOCA}) is below a threshold. Those giving an invariant mass, calculated according to Eq. I.3.5, within a symmetric interval centred on the nominal $\overline{\Lambda}$ mass value are accepted. The $\overline{\Lambda}$ candidates are also required

to decay before the end of the VELO detector and to be reconstructed with good quality. A requirement on the combination of all particles IP, discussed in Sec. II.2.3, is also applied. Selected $\overline{\Lambda}$ candidates are combined with another pion, asked to be reconstructed with good quality and to pass soft kinematic thresholds, in order to build the $\overline{\Xi}^+ \to (\overline{\Lambda} \to \overline{p}\pi^+)\pi^+$ decay. The invariant mass for the $\overline{\Lambda}$-π^+ combination is evaluated and accepted if near to the nominal $\overline{\Xi}^+$ mass and if presenting a DV within the VELO. In the second selection step, the decay is described in more details in order to enhance the statistics of signal $\overline{\Xi}^+$ decays over the background due to random association of tracks. In order to achieve a larger efficiency cancellation in the $\overline{\Xi}^+$-to-$\overline{\Lambda}$ ratio, the $\mathrm{OV}_z \in [-400, 400]$ mm range is only considered. Without this requirement, as presented in the comparison between the normalised $\overline{p}$ OV_z distributions for prompt-$\overline{\Lambda}$ (in red) and $\overline{\Xi}^+$ decays (in green) in Fig. II.1.37, the ratio would have a large bias due to non-cancelling efficency terms. The residual difference in the OV_z distribution is mitigated with a weighting procedure. The selected $\overline{\Lambda}$ candidates are included in the Armenteros-Podolanski plot [131], offering a description of all decays of a neutral particle towards an opposite-charged final-state particle pair. Applied to the $\overline{\Lambda} \to \overline{p}\pi^+$ decay, its axes are:

$$\begin{aligned} x : \alpha &= \frac{p_{\pi^+} \cdot cos(\theta_{\pi^+,\overline{\Lambda}}) - p_{\overline{p}} \cdot cos(\theta_{\overline{p},\overline{\Lambda}})}{p_{\pi^+} \cdot cos(\theta_{\pi^+,\overline{\Lambda}}) + p_{\overline{p}} \cdot cos(\theta_{\overline{p},\overline{\Lambda}})} \\ y : \pi^+ pTrans &= p_{\pi^+} \cdot sin(\theta_{\pi^+,\overline{\Lambda}}), \end{aligned} \tag{II.1.26}$$

being $p_{\pi^+(\overline{p})}$ the pion (proton) momentum and $\theta_{\pi^+(\overline{p}),\overline{\Lambda}}$ the respective angles with the $\overline{\Lambda}$ flight direction. Depending on the particle masses involved in the decay, a unique ellipse is drawn in the Armenteros-Podolanski plot and a high selection purity is reached. The $\overline{\Xi}^+$ DV is finally required to be separated from the OV by a minimum Flight Distance (FD) of 50 mm, while the particle compatibility with the primary vertex is imposed through its DIRection Angle (DIRA), defined as the angle between a line drawn from the PV to the particle DV and the sum of the 4-momenta of its decay products, and with the combination of all particles IPs.
After the selection and the z weighting, the mass peaks for the $\overline{\Lambda}$ (top) and the $\overline{\Xi}^+$ (bottom) particles are shown in Fig. II.1.38 in the simulated (left) and data (right) samples. Distributions in data exhibit a larger combinatorial background since signals are filtered in the simulation. In order to measure the $\overline{\Lambda}$ and $\overline{\Xi}^+$ particle raw yields, the two mass distributions are fitted with parametric models:

- the combination of a Voigtian and two Gaussian distributions for the signal and a second-order polynomial function for the $\overline{\Lambda}$ mass peak;
- a Voigtian for the signal and an exponential function for the background for the $\overline{\Xi}^+$ particle.

All fits achieve a good quality and the obtained results for the number of signal events in the fits, listed in Tab. II.1.10, are used to calculate:

$$\frac{(\overline{\Xi}^+/\overline{\Lambda}_{\mathrm{prompt}})^{\mathrm{Data}}}{(\overline{\Xi}^+/\overline{\Lambda}_{\mathrm{prompt}})^{\mathrm{MC}}} = 1.09 \pm 0.09, \tag{II.1.27}$$

where also the factor in Eq. II.1.25, taking into account the different statistics of the $\overline{\Xi}^+$- and the $\overline{\Lambda}$-enriched simulated samples, is considered. The result is not observed to

Table II.1.10: $\bar{\Lambda}$ and $\bar{\Xi}^+$ number of fitted signal events in the pHe simulated and data samples.

	Simulation	Data
$\bar{\Lambda}$ fit	84800 ± 300	39600 ± 300
$\bar{\Xi}^+$ fit	2750 ± 50	510 ± 40

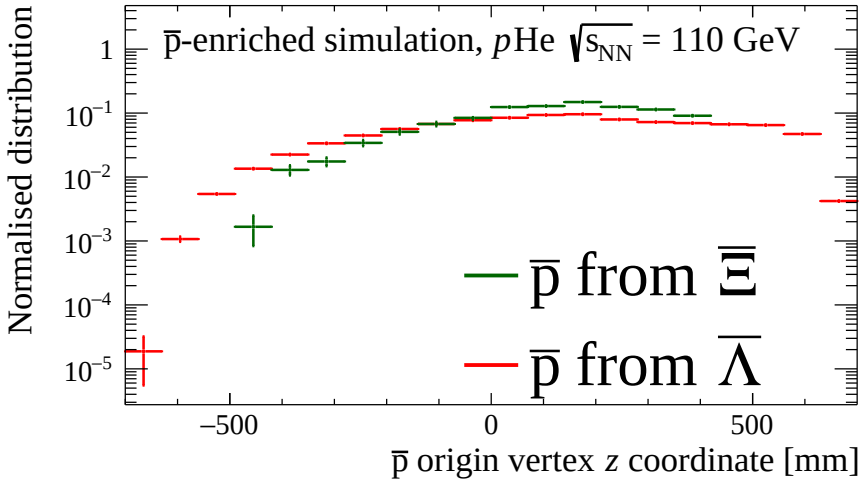


Figure II.1.37: Comparison between the normalised distributions of the origin vertex z coordinate for antiprotons coming from $\bar{\Lambda}$ and $\bar{\Xi}$ decays.

depend significantly on the antiproton kinematics and is therefore assumed as an overall scale-factor.

The effect of the 9% $\bar{\Xi}$ underestimation in the simulated sample with respect to data is then studied. The possible distortion of the template for detached antiprotons is assumed to be negligible, following the result of the previous section and considering that the $\bar{\Xi}^+$ contribution to all detached antiprotons is one order of magnitude smaller than the $\bar{\Sigma}^-$ one. Vice versa, the effect on the simulation and reconstruction and selection efficiencies is considered by introducing a weight to enhance the $\bar{\Xi}^+$ statistics such as:

$$\varepsilon_{\text{det}}^{\text{sim}} = \frac{\Sigma_{i=\bar{\Lambda}_{\text{prompt}},\,\bar{\Sigma}^-,\,\bar{\Xi}^+,\,\bar{\Xi}^0}\; w_i \cdot N_{\text{gen}\,i}^{\text{MC}} \cdot \varepsilon_i^{\text{sim}}}{\Sigma_{i=\bar{\Lambda}_{\text{prompt}},\,\bar{\Sigma}^-,\,\bar{\Xi}^+,\,\bar{\Xi}^0}\; w_i \cdot N_{\text{gen}\,i}^{\text{MC}}} \tag{II.1.28}$$

$$\varepsilon_{\text{det}}^{\text{rec \& sel}} = \frac{\Sigma_{i=\bar{\Lambda}_{\text{prompt}},\,\bar{\Sigma}^-,\,\bar{\Xi}^+,\,\bar{\Xi}^0}\; w_i \cdot N_{\text{sim}\,i}^{\text{MC}} \cdot \varepsilon_i^{\text{rec \& sel}}}{\Sigma_{i=\bar{\Lambda}_{\text{prompt}},\,\bar{\Sigma}^-,\,\bar{\Xi}^+,\,\bar{\Xi}^0}\; w_i \cdot N_{\text{sim}\,i}^{\text{MC}}}, \tag{II.1.29}$$

being now $w_{\bar{\Xi}^+,\,\bar{\Xi}^0} = 1.09 \pm 0.09$ and equal to unity elsewhere. The percent variation of the simulation and reconstruction and selection efficiencies is illustrated in Figs. II.1.39 and II.1.40, respectively. The former is found to be negligible with respect to other sources of uncertainty, which also was expected considering the results of the previous section. The average reconstruction efficiency changes by $(-0.7 \pm 0.7)\%$. The $w_{\bar{\Xi}^+,\,\bar{\Xi}^0}$ weights are applied to the nominal analysis and a relative 0.7% systematic uncertainty is assigned on R.

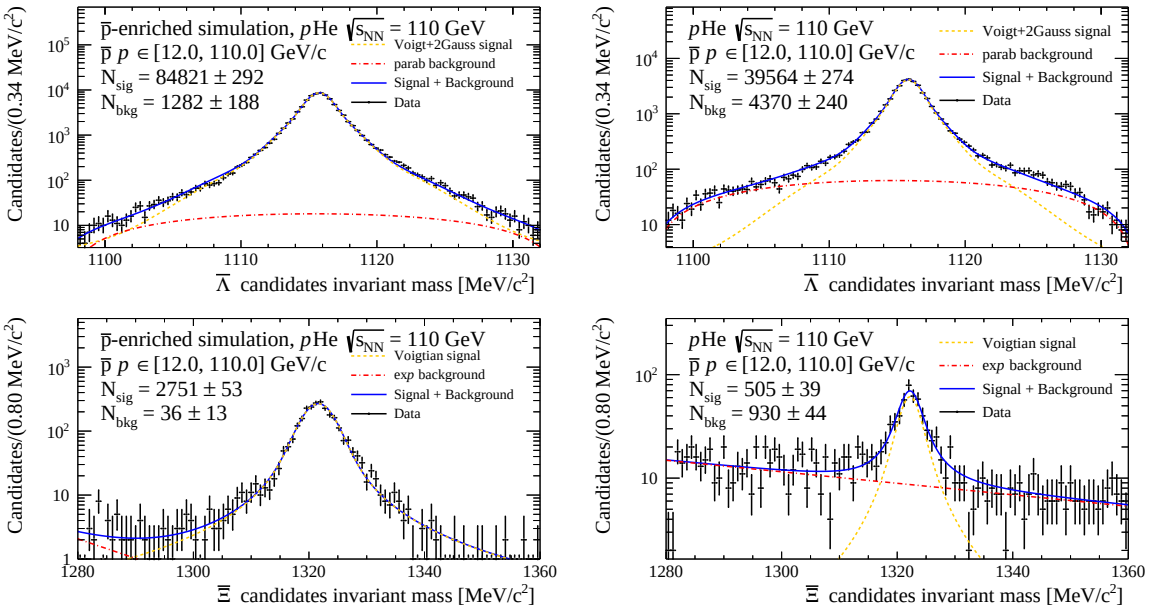


Figure II.1.38: Invariant mass distribution for the selected $\overline{\Lambda}$ (top) and $\overline{\Xi}^{+}$ (bottom) candidates in the pHe simulated (the $\overline{\Lambda}$-enriched in top left, the $\overline{\Xi}^{+}$-enriched in bottom left) and data (right) sample. The $\overline{\Lambda}$ candidates are modelled with the sum of a Voigitan and two Gaussian distributions for the signal and a second-order polynomial function for the background; $\overline{\Xi}^{+}$ with a Voigtian and an exponential function. The resulting numbers of fitted signal events are indicated.

Fig. II.1.41 shows the reconstruction and selection efficiencies for detached antiprotons after the application of the weights for the $\overline{\Sigma}^{-}$-to-$\overline{\Lambda}$ and $\overline{\Xi}$-to-$\overline{\Lambda}$ ratios.

II.1.6.4 Particle identification

The contribution to the systematic uncertainty related to the identification in the detector of prompt and detached antiprotons is studied in this section. As explained in Sec. II.1.3, a tight cut on the PID variables is set so that the MISID contamination, which would not be distinguished by the fit covering the same $\chi^2_{\rm IP}$ region than prompt antiprotons, can be fixed to the predictions of the minimum-bias simulated sample. On the opposite, a different accuracy of the simulation in reproducing the PID classifiers of prompt and detached antiprotons, not cancelling in the R ratio, could induce a bias. The PID efficiency for the $\mathrm{DLL}_{p,\pi} > 20$ and $\mathrm{DLL}_{p,K} > 10$ requirements is measured on simulation for prompt and detached antiprotons. This is defined as:

$$\varepsilon^{\rm PID}_{\rm prompt\,(det)} = \frac{N^{\rm post}_{\rm prompt\,(det)}}{N^{\rm pre}_{\rm prompt\,(det)}} \qquad \text{(II.1.30)}$$

being $N^{\rm pre\,[post]}_{\rm prompt\,(det)}$ the number of prompt (detached) candidates before [after] the application of the PID cuts. The results on the $\overline{p}$-enriched simulated sample integrating over the full kinematic phase-space are listed in Tab. II.1.11, while the efficiency dependence on the antiproton momentum (top left), transverse momentum (top right), pseudorapidity

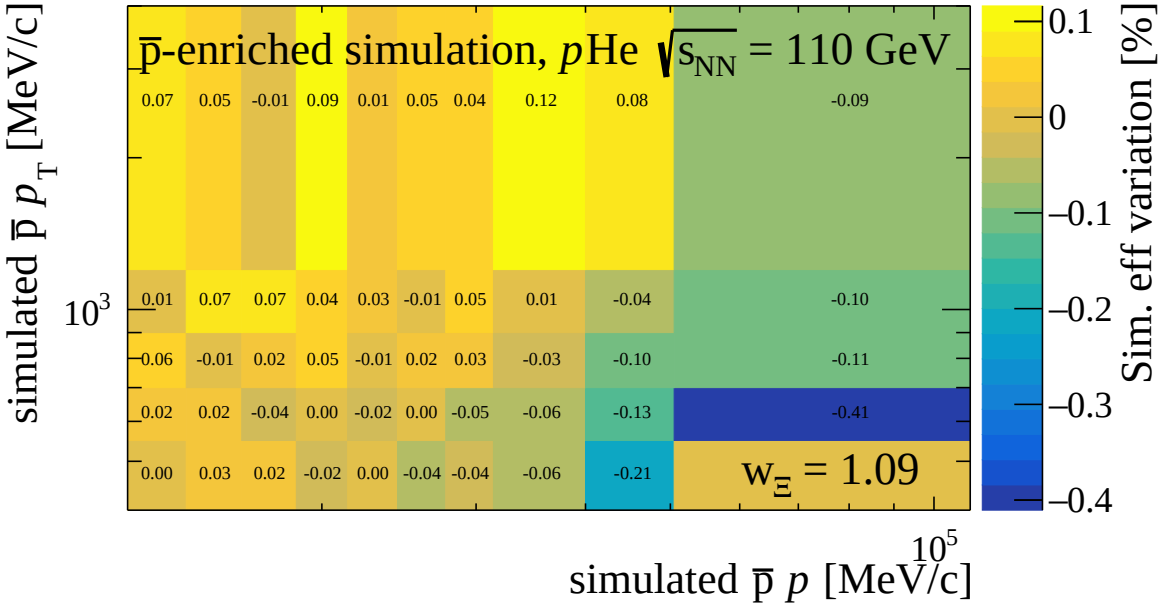


Figure II.1.39: Simulation efficiency per-cent variation for detached antiproton in the $\overline{p}$-enriched simulated sample when enhancing the simulated $\overline{\Xi}^{+}$ and $\overline{\Xi}^{0}$ abundances by a factor 1.09, as measured on the pHe data with the procedure described in the text.

Table II.1.11: PID efficiency for prompt and detached antiprotons with different applied PID selection.

	OV$_z$ range [mm]	prompt $\overline{p}$ [%]	detached $\overline{p}$ [%]	$\varepsilon^{\text{PID}}_{\text{prompt}}/\varepsilon^{\text{PID}}_{\text{det}}$
DLL$_{p,\pi} > 20$	$z \in [-700, 700]$	64.90 ± 0.03	57.74 ± 0.11	1.124 ± 0.002
DLL$_{p,K} > 10$	$z \in [-700, 100]$		60.93 ± 0.14	1.065 ± 0.002
DLL$_{p,\pi} > 17$ DLL$_{p,K} > 9.6$	$z \in [-700, 700]$	69.32 ± 0.03	62.18 ± 0.10	1.115 ± 0.002

(bottom left) and OV$_z$ (bottom right) are shown in Fig. II.1.42. A similar evolution is found as a function of the momentum, where the two peaking structures correspond to the regions where the two RICH detectors are optimised. The efficiency difference, approximately a constant 7%, is in part explained by the steeply decreasing evolution along z, worsening the performance for detached antiprotons. A second comparison restricted to OV$_z \in [-700, 100]$ mm is then performed and, as listed in Tab. II.1.11, the PID efficiency for detached antiprotons increases, with a residual difference with respect to the prompt ones $\sim 4\%$. Figure II.1.43 shows the same comparisons than Fig. II.1.42, but restricted to OV$_z \in [-700, 100]$ mm. The evolution with the particle kinematics and the event geometry is found to be similar and the residual differences are mostly due to geometric and kinematic aspects, which are expected to be precisely reproduced by the simulation. Large PID effects that could impact the systematic uncertainty are thus excluded.

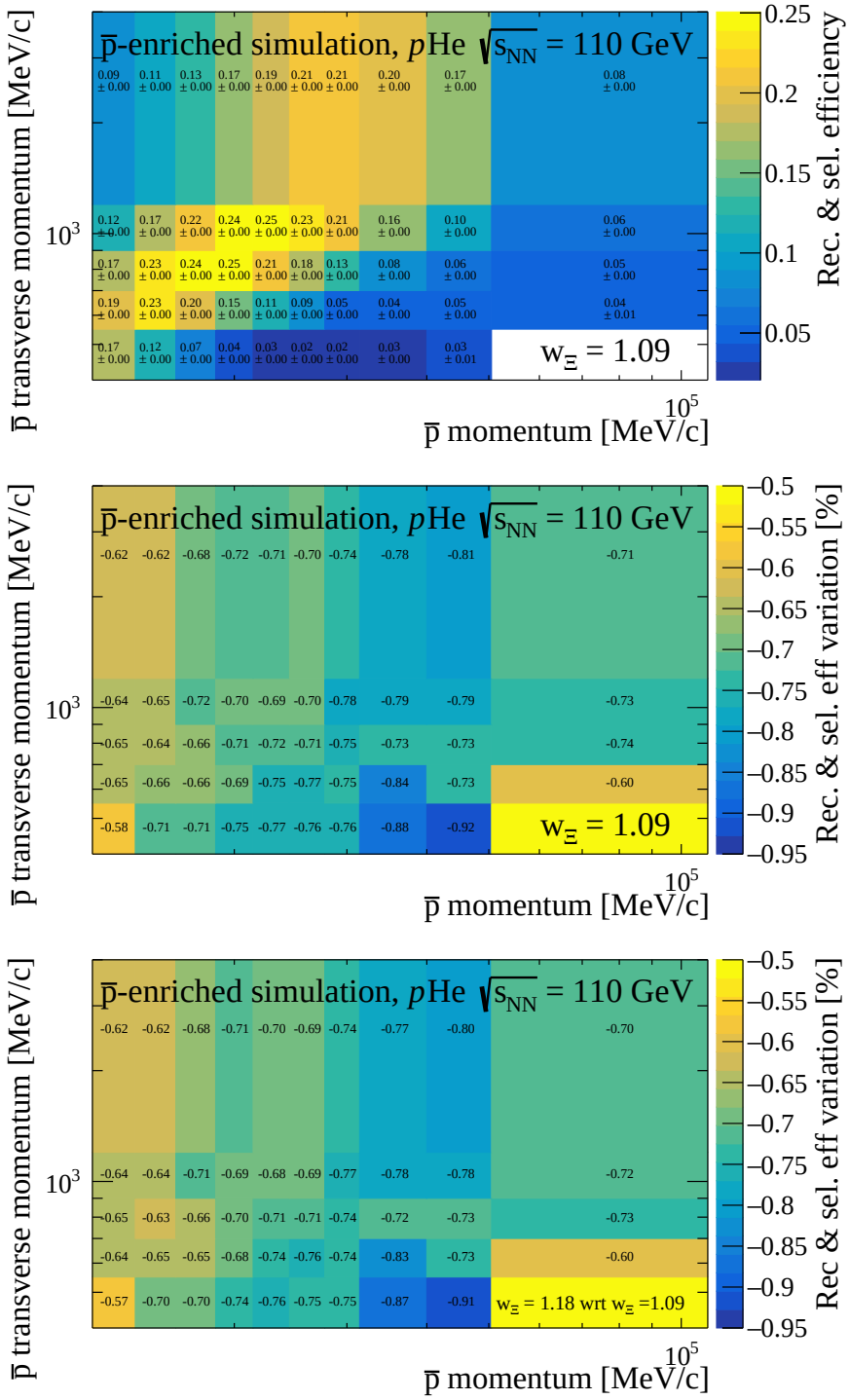


Figure II.1.40: Reconstruction and selection efficiencies for detached antiprotons in the $\overline{p}$-enriched simulation in momentum and transverse momentum bins when enhancing the $\overline{\Xi}$ statistics (top). The middle plot shows the percent variation with respect to the values with the $\overline{\Xi}^{+}$-to-$\overline{\Lambda}$ ratio fixed to the EPOS-LHC prediction, the bottom one the variation when comparing $w_{\overline{\Xi}} = 1.18$ with $w_{\overline{\Xi}} = 1.09$.

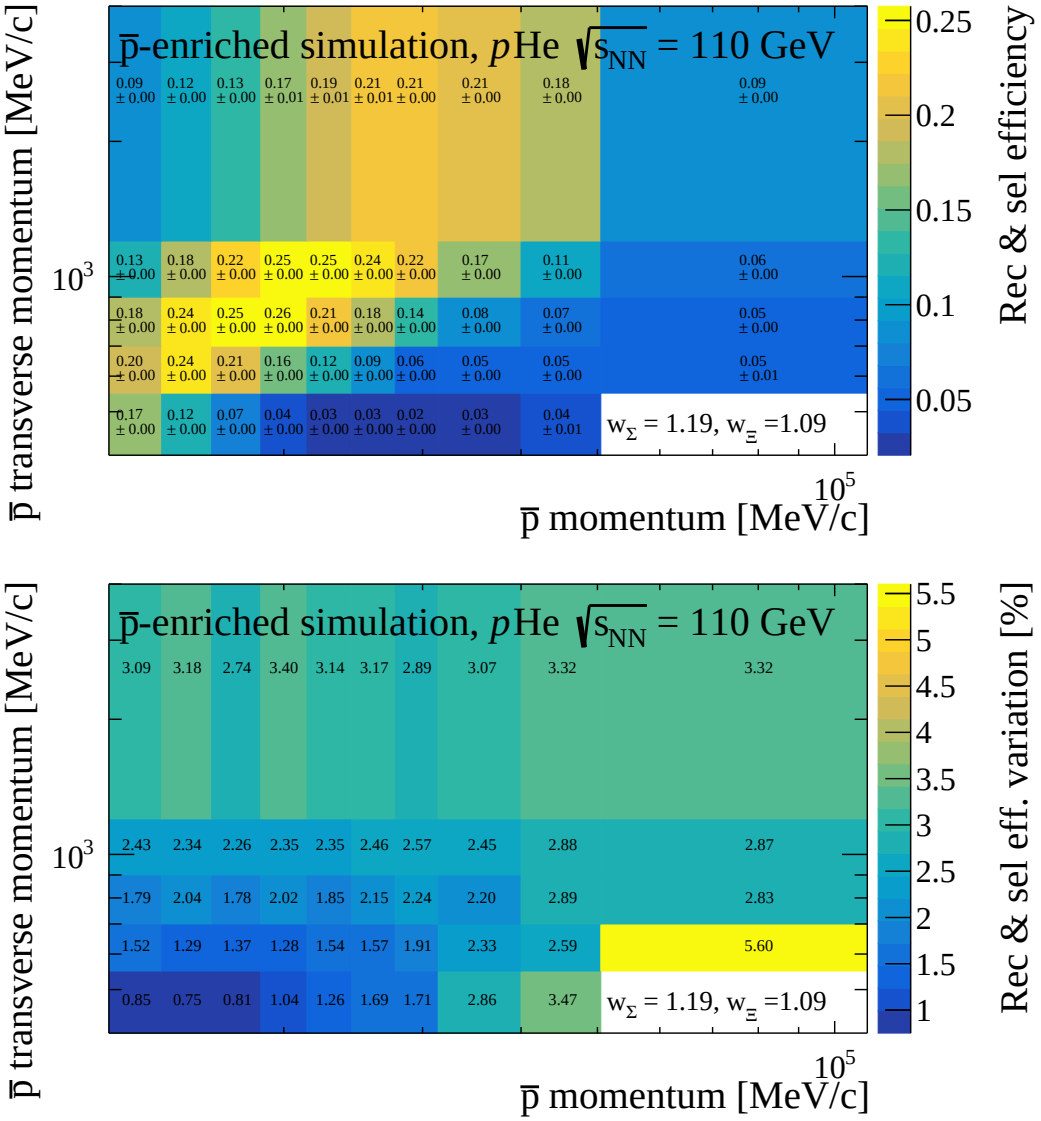


Figure II.1.41: Reconstruction and selection efficiencies for detached antiprotons after the application of all weights (top) and percent variation with respect to the result of Fig. II.1.22 in the $\overline{p}$-enriched simulated sample in bins of their momentum and transverse momentum.

A data-driven cross-check of the PID efficiency for the dominant contribution among all detached antiprotons is performed selecting in the pHe data the $\overline{\Lambda} \to \overline{p}\pi^+$ decays, as summarised in Tab. II.1.12. A similar strategy as the previous section is followed. After the preselection requirements discussed above, good-quality reconstructed particles are assumed as antiprotons and pions and combined when satisfying loose kinematic criteria. The detachment with respect to the PV is imposed through the particles χ^2_{IP} and an invariant mass interval centred on the nominal $\overline{\Lambda}$ mass value is accepted. As presented in left Fig. II.1.44, other than signal events, defined as those where the reconstructed particle hypotheses coincide with the simulated ones, a large contamination of combinatorial background events (in blue) and of $K^0_{\text{S}} \to \pi^-\pi^+$ decays where the negatively-charged pion is misidentified as an antiproton (in green) is found. To remove the former, the decay is described in the Armenteros-Podolanski plot and a tight

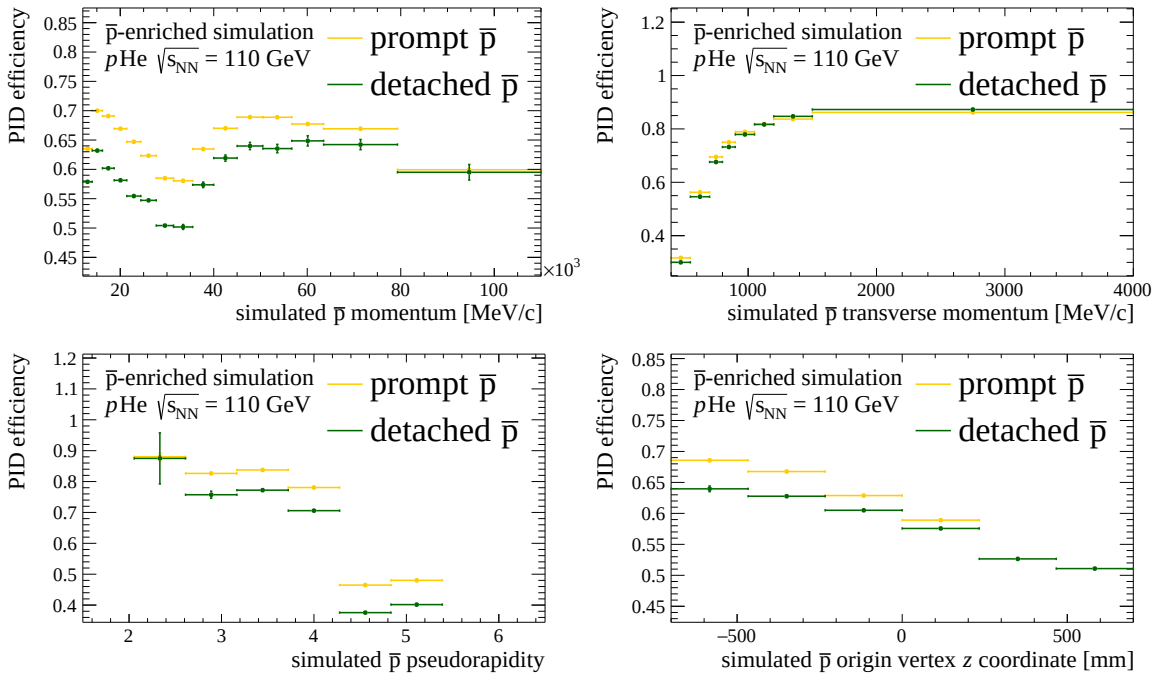


Figure II.1.42: PID efficiency for prompt (in orange) and detached (in green) antiprotons in the $\overline{p}$-enriched simulated sample considering all antiproton origin vertex z values as a function of their momentum (top left), transverse momentum (top right), pseudorapidity (bottom left) and z coordinate of the origin vertex (bottom right).

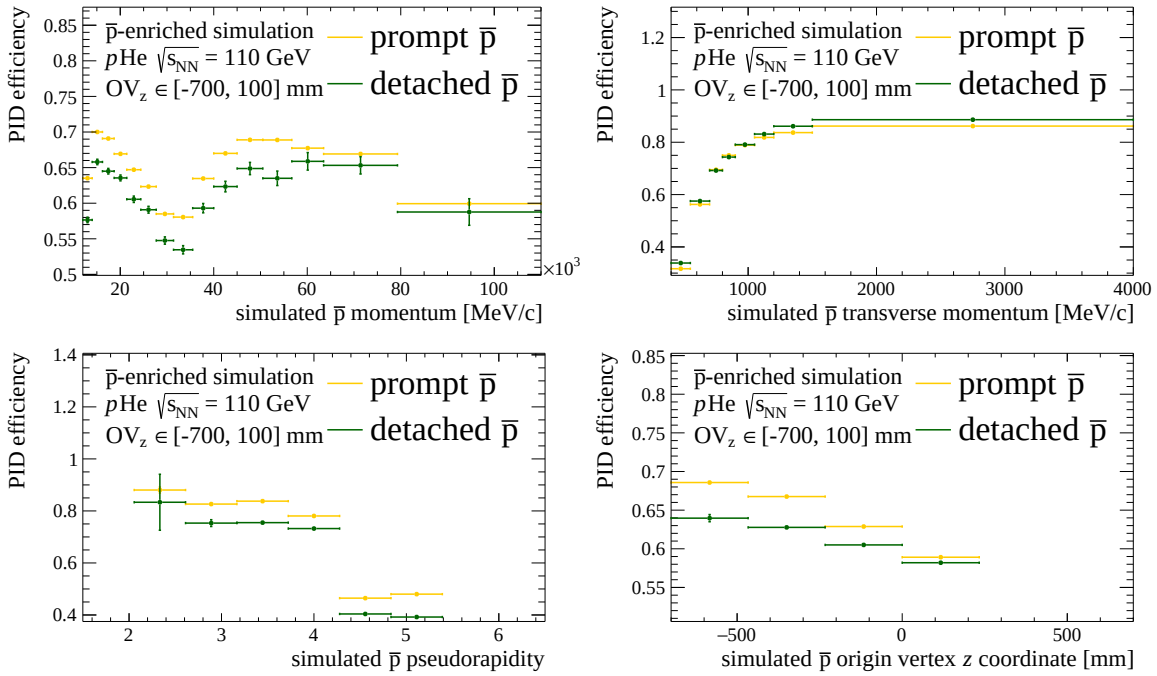


Figure II.1.43: PID efficiency for prompt (in orange) and detached (in green) antiprotons in the $\overline{p}$-enriched simulated sample only considering the antiproton origin vertex range $z \in [-700, 100]$ mm as a function of their momentum (top left), transverse momentum (top right), pseudorapidity (bottom left) and z coordinate of the origin vertex (bottom right).

Table II.1.12: Selection algorithm for the $\overline{\Lambda} \to \overline{p}\pi^+$ decays applied on pHe data to cross-check the simulated PID efficiency measurement.

	Selection cuts
Event	$N_{SPD} > 0$, $N_{VELO} > 0$, $N_{\rm trk}(\eta < 0) < 5$
$\overline{p}$	$p \in [12, 100]\,\text{GeV}/c$, $p_{\rm T} \in [0.4, 4]\,\text{GeV}/c$, $\eta < 5$ $log(\chi^2_{\rm IP}) \in [3, 6.5]$, $\text{OV}_z < 700\,\text{mm}$, $GhostProb < 0.15$
π^+	$p < 20\,\text{GeV}/c$, $p_{\rm T} < 0.5\,\text{GeV}/c$, $log(\chi^2_{\rm IP}) \in [3, 7.5]$
$\overline{\Lambda}$	$\eta < 5.5$, $M \in [1122, 1122]\,\text{MeV}/c^2$, $DIRA < 0.0027$ $WM < 470$ or $530 < WM < 560\,\text{MeV}/c^2$
Armenteros	$\left\lvert \left(\frac{\alpha+0.69}{0.18}\right)^2 + \left(\frac{p_\pi \cdot sin\theta_{\pi\bar{\Lambda}}}{100.4\,\text{MeV}/c}\right)^2 - 1 \right\rvert < 0.05$ $\pi^+\, p_{Trans} > 20\,\text{MeV}/c$, $\overline{p}\, p_{Trans} < 101\,\text{MeV}/c$

requirement on it is applied. For the latter, the Wrong Mass (WM), *i.e.* the two-tracks invariant mass under the two-pion hypothesis, is evaluated and shown in middle-left plot of Fig. II.1.44. Over a continuous distribution, a peak at the nominal $K^0_{\rm S}$ mass value is found and removed from the selection. With the application of all the requirements, the Armenteros-Podolanski plot (top), the wrong (middle) and the $\overline{\Lambda}$ invariant mass (bottom) are presented in right Fig. II.1.44 and an efficient removal of the background is achieved. The resulting purity, defined as the fraction of signal events over the total selected ones, is larger than 97.5%.

After the application of the selection algorithm, the PID efficiency is measured with Eq. II.1.30 and compared to the evaluation on the simulated candidates. The quality of the simulation (in green) in reproducing the data distributions (in red) of the quantities affecting the PID response is studied in Fig. II.1.45. An excellent representation of the antiproton kinematics is found (left), while, as a consequence of the limited simulation of low-energy backgrounds, large differences arise for the detector occupancy (right). Hence, also different PID efficiencies are expected. On the selected data sample, a $62.2 \pm 0.03\%$ PID efficiency is found, with a relative 4.5% variation with respect to the value measured on simulation. The study is also repeated as a function of the same quantities reported in the previous paragraph and Fig. II.1.46 presents the result for the $\overline{p}$ OV_z. A similar evolution for the two measurements is found, except for the first bins where only poor data statistics is available. This is a general limitation for most of the samples collected in Run2, notably the pHe one analysed here, and motivates the development of the machine-learning-based approach described in Part III.
In order to assign a contribution to the systematic uncertainty, the PID cuts applied in the analysis are made looser up to make the efficiency measured on simulation for detached antiprotons to coincide with the data one. With the same set of cuts, listed in II.1.11, the efficiency for prompt antiprotons is also measured. The $\varepsilon^{\rm PID}_{\rm prompt}/\varepsilon^{\rm PID}_{\rm det}$ ratio is found to change by a relative 0.9% with respect to the cuts considered for the analysis. This shows that the difference in PID response is not critically dependent on the chosen working

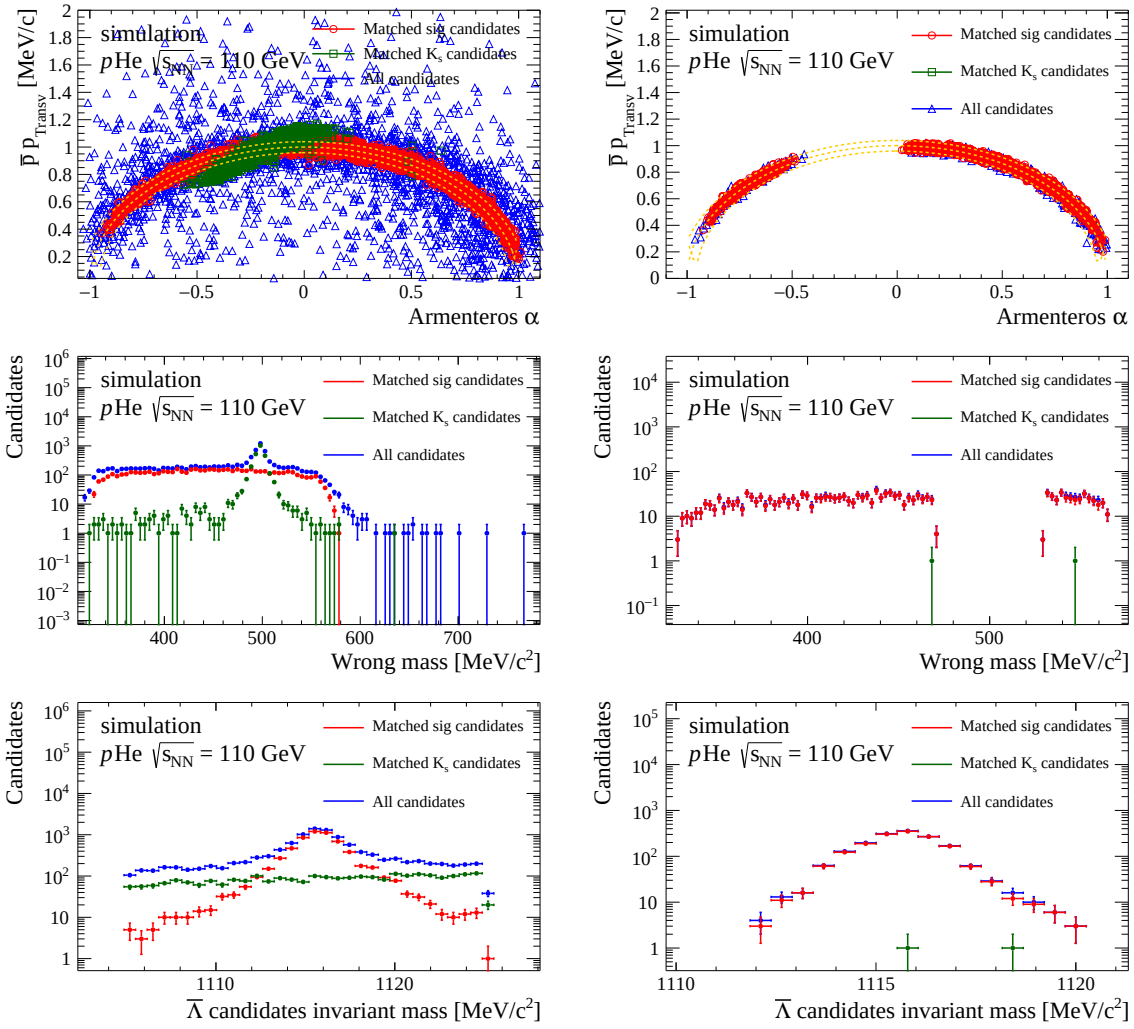


Figure II.1.44: Distributions in the pHe simulated sample of the Armenteros-Podolanski variables (top), the wrong (middle) and the $\bar{\Lambda}$ invariant mass (bottom) before (left) and after (right) the application of the selection algorithm intended to enhance the signal events (in red) over the contamination from random association of tracks (in blue) and from $K^0_S \to \pi^+\pi^-$ decays (in green) where the negatively-charged pion is misidentified as an antiproton.

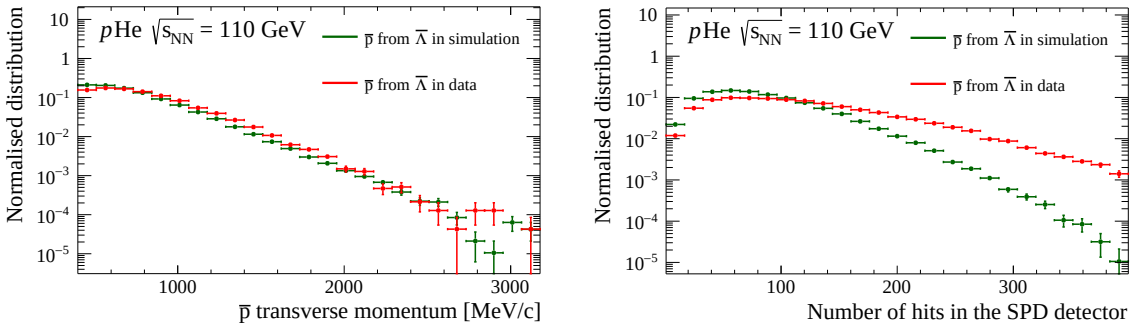


Figure II.1.45: Normalised distributions of the transverse momentum (left) and of the number of hits in the SPD detector (right) for antiprotons originating in a $\bar{\Lambda}$ decay in pHe simulation (in green) and data (in red).

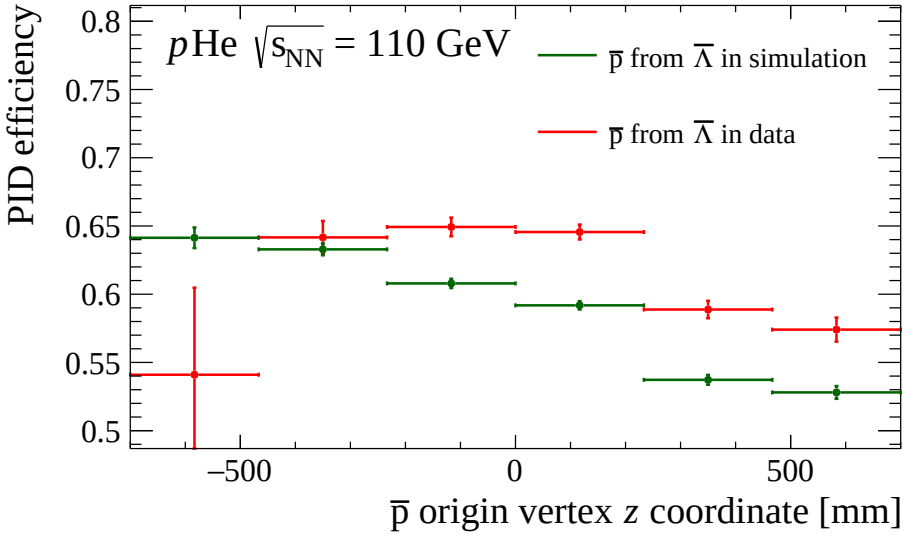


Figure II.1.46: PID efficiency as a function of the z coordinate of the antiproton origin vertex when measured for antiprotons produced in $\bar{\Lambda}$ decays in the pHe simulation (in green) and in data (in red).

point and the 0.9% variation is taken as an estimation of the related systematic uncertainty.

An additional validation of the ability of the simulation to predict the difference in PID efficiency between prompt and detached antiprotons will be provided in Sec. III.1.4.2, where the data-driven modelling of the PID response I developed during my *Ph.D.* programme is presented.

II.1.6.5 Tracking

The different efficiency of the prompt and detached antiproton reconstruction is then studied. This is defined as:

$$\varepsilon_{\text{trk}} = \frac{N_{\text{rec.ed}}}{N_{\text{rec.ble}}}, \tag{II.1.31}$$

being:

- $N_{\text{rec.ed}}$ the number of particles that are reconstructed in all tracking detectors and with at least 70% of the energy deposits;
- $N_{\text{rec.ble}}$ the number of particles leaving at least one radial and one ϕ signal in three VELO modules, one in the x and one between the u or v stations of both the TT and IT/OT detectors.

The efficiency is expected to mostly depend on the number of hits constituting the track, which is a function of the antiproton kinematics and of its OV coordinates. The results for prompt (in orange) and detached (in green) antiprotons, obtained with the $\overline{p}$-enriched simulated sample, are listed in Tab. II.1.13 integrating over the full kinematic phase-space and are presented in Fig. II.1.47 as a function of their simulated momentum (top left), transverse momentum (top right), pseudorapidity (bottom left) and OV_z (bottom

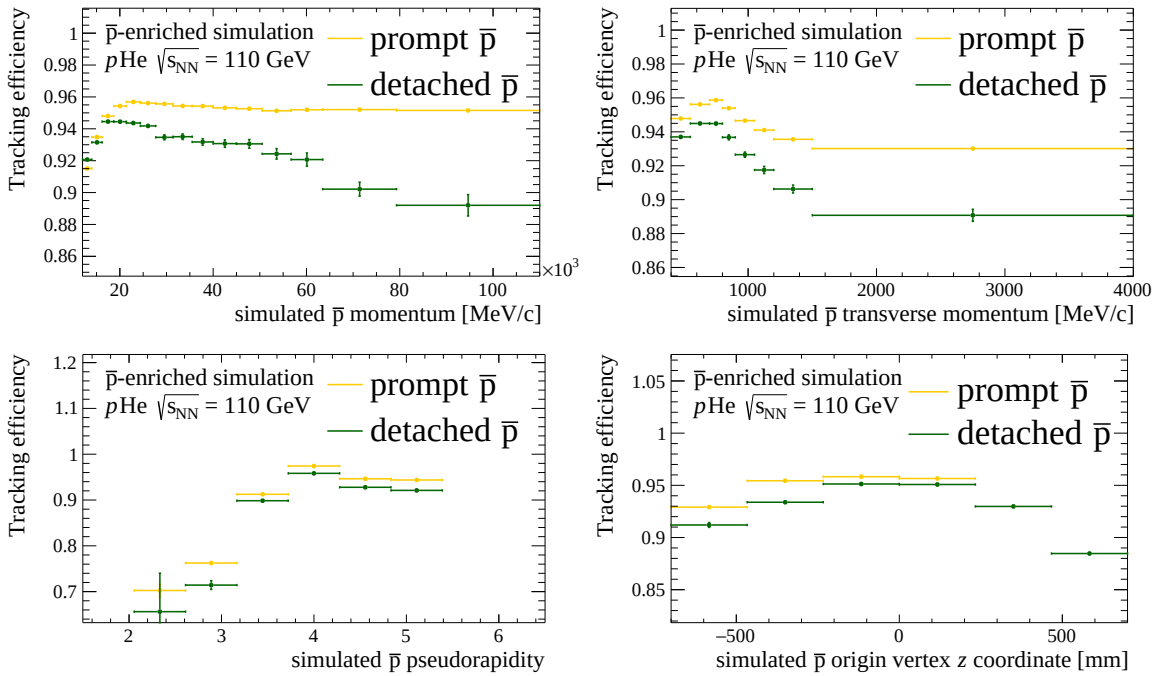


Figure II.1.47: Tracking efficiency for prompt (yellow) and detached (green) antiprotons in the $\overline{p}$-enriched simulated sample as a function of their momentum (top left), transverse momentum (top right), pseudorapidity (bottom left) and z coordinate of the origin vertex (bottom right).

Table II.1.13: Tracking efficiency for prompt and detached antiprotons in the $\overline{p}$-enriched simulated sample.

OV_z range	prompt $\overline{p}$ [%]	detached $\overline{p}$ [%]
$z \in [-700, 700]$ mm	94.852 ± 0.013	93.47 ± 0.05
$z \in [-700, 100]$ mm	94.852 ± 0.013	94.33 ± 0.06

right). A lower efficiency by $\sim 1.5\%$ is found for detached antiprotons. The difference, as in the previous section, mostly arise in the $\mathrm{OV}_z > 100$ mm region, only populated by detached antiprotons and affected by a poor efficiency. A different tracking efficiency between prompt and detached antiprotons can also be seen for $\mathrm{OV}_z < -200$ mm. This can be explained by Fig. II.1.48, showing the tracking efficiency for prompt (top) and detached (bottom) antiprotons as a function of their pseudorapidity and OV_z. For $\mathrm{OV}_z < -200$ mm, the two antiproton categories are reconstructed with the same efficiency at large pseudorapidity values, while differences arise for $\eta < 4$. Since the detached antiprotons originate in anti-hyperon decays, their OV is on average more distant from the beam axis than prompt antiprotons, and their pseudorapidity distribution averages to lower values. As a consequence, detached antiprotons are reconstructed with a lower number of energy deposits.
In order to verify that the measured efficiency difference comes from detached antiprotons with $\mathrm{OV}_z > 100$ mm, the study is repeated only considering $\mathrm{OV}_z \in [-700, 100]$ mm, as in the previous section. Provided the same OV_z region for the two categories, the residual

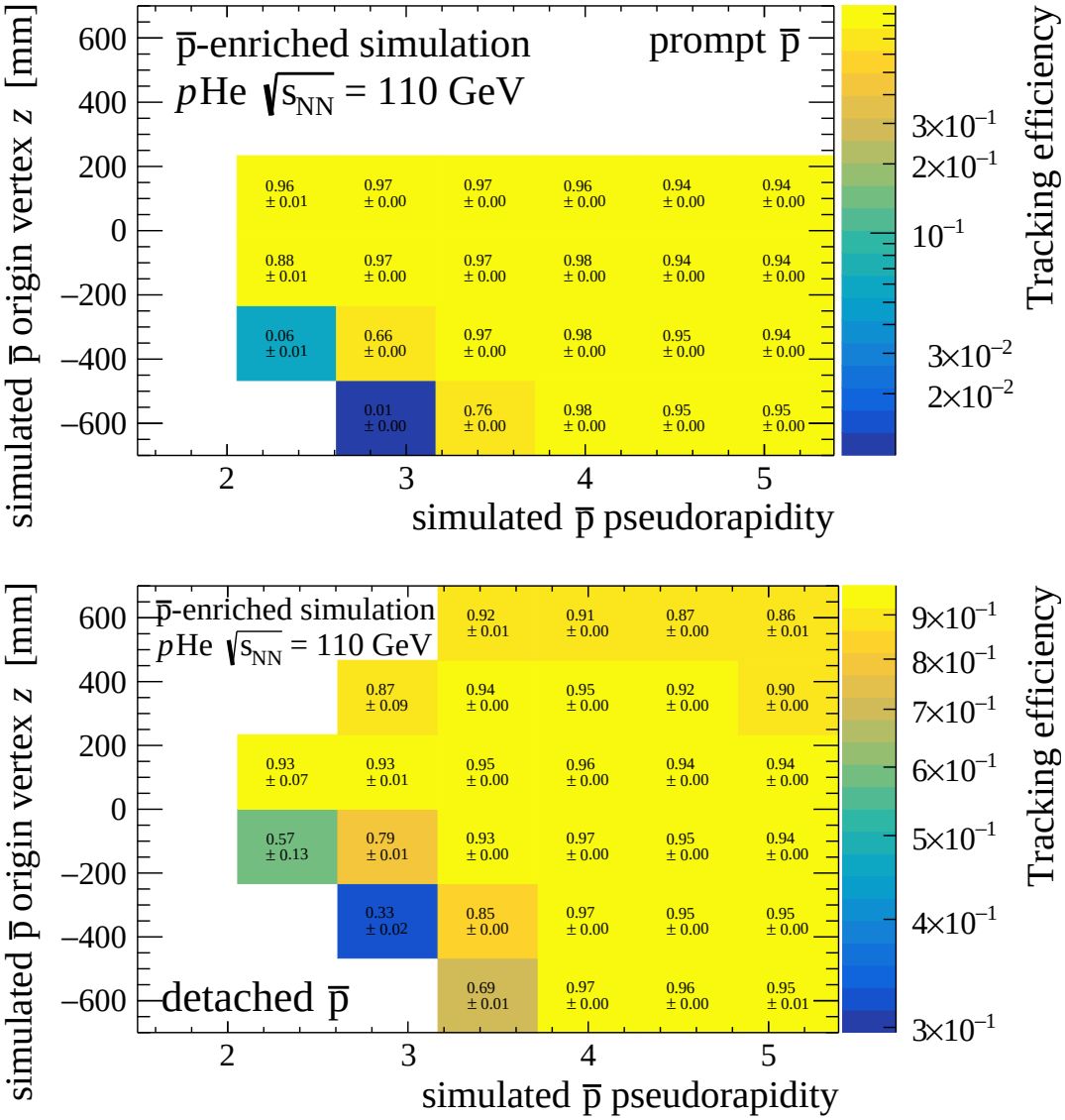


Figure II.1.48: Tracking efficiency for prompt (top) and detached (bottom) antiprotons in the $\overline{p}$-enriched simulated sample as a function of their pseudorapidity and z coordinate of their origin vertex.

efficiency difference between prompt and detached antiprotons is $\sim 0.5\%$ on the full kinematic phase-space, as indicated in Tab. II.1.13, and is mostly due to geometric effects at the most upstream OV_z values, as presented in Fig. II.1.49. Since the particle kinematics and the event geometry are known to be reproduced by the simulation with high precision, the efficiencies are expected to be accurate as well. Data-driven confirmations for this are documented in Ref. [61] and in Sec. II.2.6.1 for the prompt and the detached component, respectively. The systematic uncertainty due to the tracking is hence neglected.

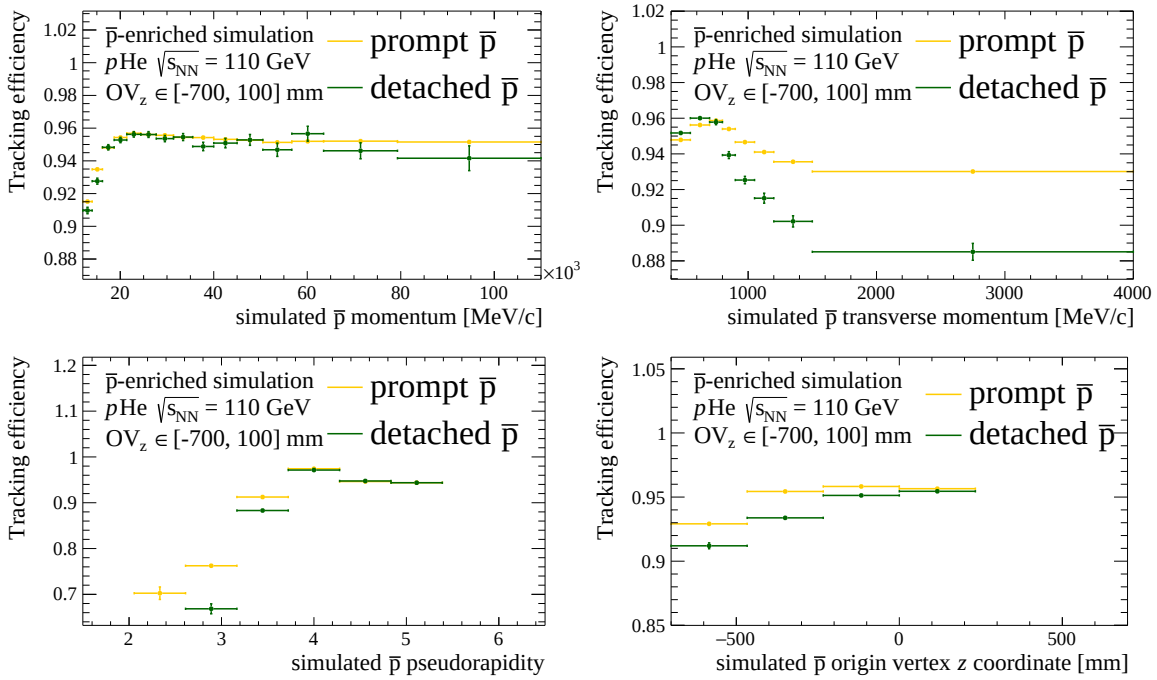


Figure II.1.49: Tracking efficiency for prompt (yellow) and detached (green) antiprotons in the $\overline{p}$-enriched simulated sample as a function of their momentum (top left), transverse momentum (top right), pseudorapidity (bottom left) and z coordinate of the origin vertex (bottom right) only considering $\mathrm{OV}_z \in [-700, 100]$ mm.

Table II.1.14: Summary of the systematic uncertainties.

Statistical uncertainty (fit)	(most bins) $< 2.5\%$
MC statistics (efficiencies)	(most bins) $< 1.8\%$
Parametrization	(most bins) $< 2\%$
Prompt $\overline{p}$ template	4.4 %
$\overline{\Sigma}^-$-to-$\overline{\Lambda}$ simulated ratio	1.4- 4 %
$\overline{\Xi}^+$-to-$\overline{\Lambda}$ simulated ratio	0.7%
PID	0.9%
Tracking	negligible

II.1.7 Results for the inclusive approach

The detached-to-prompt antiproton ratio

$$R = \frac{\sigma(p\mathrm{He} \to (\overline{\Lambda} \to \overline{p}\pi^+)\, X) + \sigma(p\mathrm{He} \to (\overline{\Sigma}^- \to \overline{p}\pi^0)\, X)}{\sigma(p\mathrm{He} \to \overline{p}_{\mathrm{prompt}}\, X)} = \frac{N^{\mathrm{Data}}_{\mathrm{det}}/\varepsilon_{\mathrm{det}}}{N^{\mathrm{Data}}_{\mathrm{prompt}}/\varepsilon_{\mathrm{prompt}}} \qquad \text{(II.1.32)}$$

is measured on a sample of pHe collisions collected with the LHCb experiment with a $\sqrt{s_{\mathrm{NN}}}$ =110 GeV nucleon-nucleon centre-of-mass energy. The detached and prompt antiproton abundances $N^{\mathrm{Data}}_{\mathrm{det,\,prompt}}$ are determined with a template fit to the antiproton candidates χ^2_{IP} distribution in the pHe data; the simulation and reconstruction and selection efficiencies $\varepsilon_{\mathrm{det,\,prompt}}$ are measured with the simulated sample. Several contributions to the systematic uncertainties, summarised in Tab. II.1.14, are scrutinised employing the simulation or data-driven checks and summed in quadrature to the fit uncertainties.
The obtained results in the considered binning scheme for the antiproton momentum and transverse momentum are listed in Tabs. II.1.15 and II.1.16 and presented in Fig. II.1.50. In the figure top plot the uncertainty is only statistical, in the bottom is the squared sum of the statistical and systematic contributions. Figure II.1.51 presents the resulting per-cent relative uncertainty of the measurement. The statistical contribution ranges from 1.9 to 6.2%, but is below 2.5% in most bins; the total uncertainty is in most bins lower than 6.5%. The contribution arising from the systematic uncertainties is dominant. The results are then separately projected onto the momentum and transverse momentum axes and compared in Fig. II.1.52 with the predictions of the theoretical models introduced in Sec. II.0.1, re-evaluated to exclude the low-transverse momentum and high momentum bins that are removed from the analysis because of the too large MISID contamination. For the projections, the contributions to the uncertainties coming from the fit and the efficiencies are considered as not correlated, the systematics as fully correlated among bins and are summed in quadrature and linearly, respectively. In the figures, inner error bars show the fit uncertainty only, the outer ones the total, obtained summing in quadrature the statistic and systematic contributions. The measured ratio R is found to be slightly decreasing with the momentum and the transverse momentum and follows the same evolution predicted by most theoretical models. On absolute, its value is measured as significantly larger than theoretical models, especially with respect to EPOS-LHC that is considered for the generation of the simulated samples employed in this analysis.

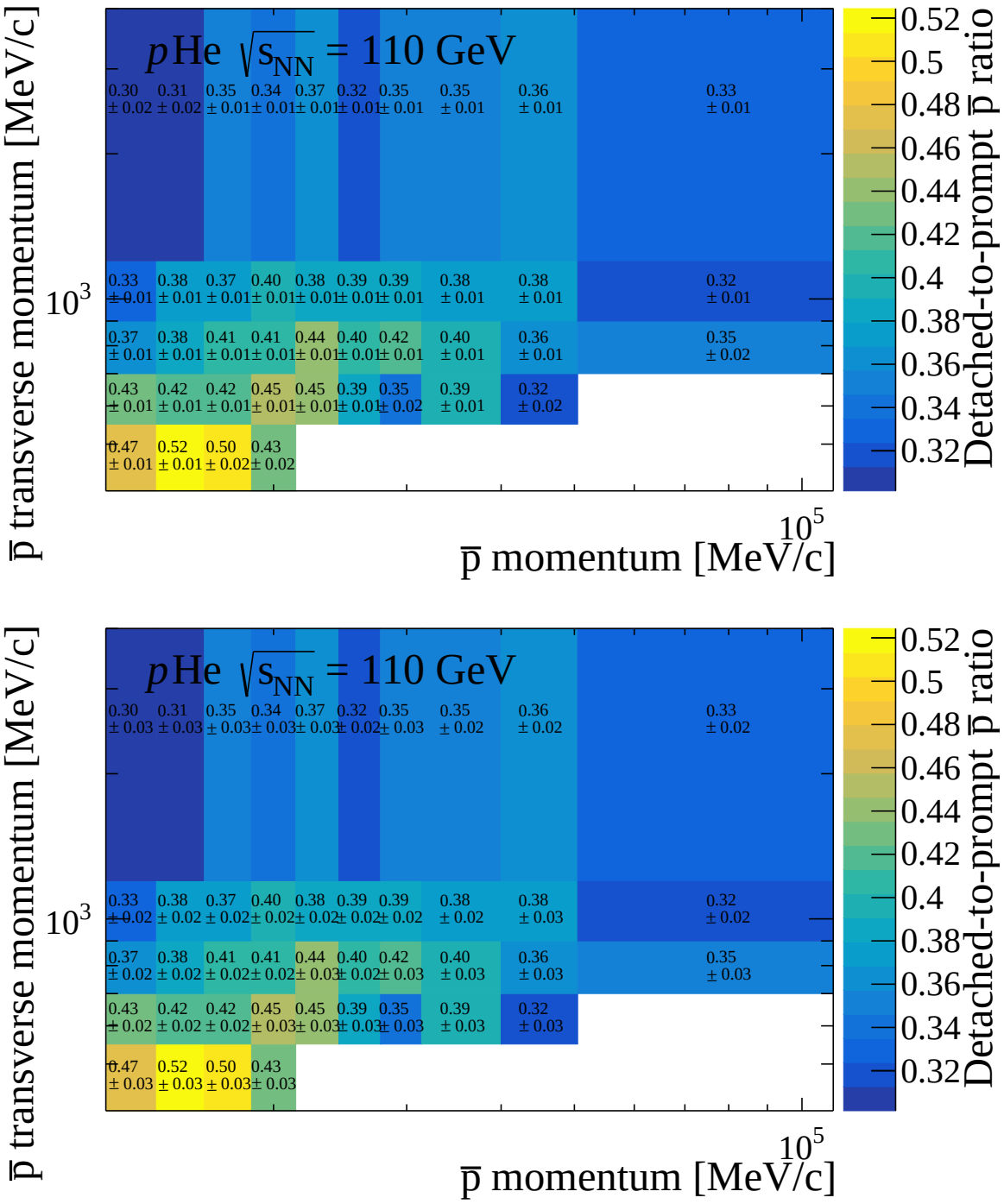


Figure II.1.50: Detached-to-prompt $\overline{p}$ ratio measured on the LHCb pHe 2016 data as a function of its momentum and transverse momentum with displayed statistical only (top) and total (bottom) uncertainties.

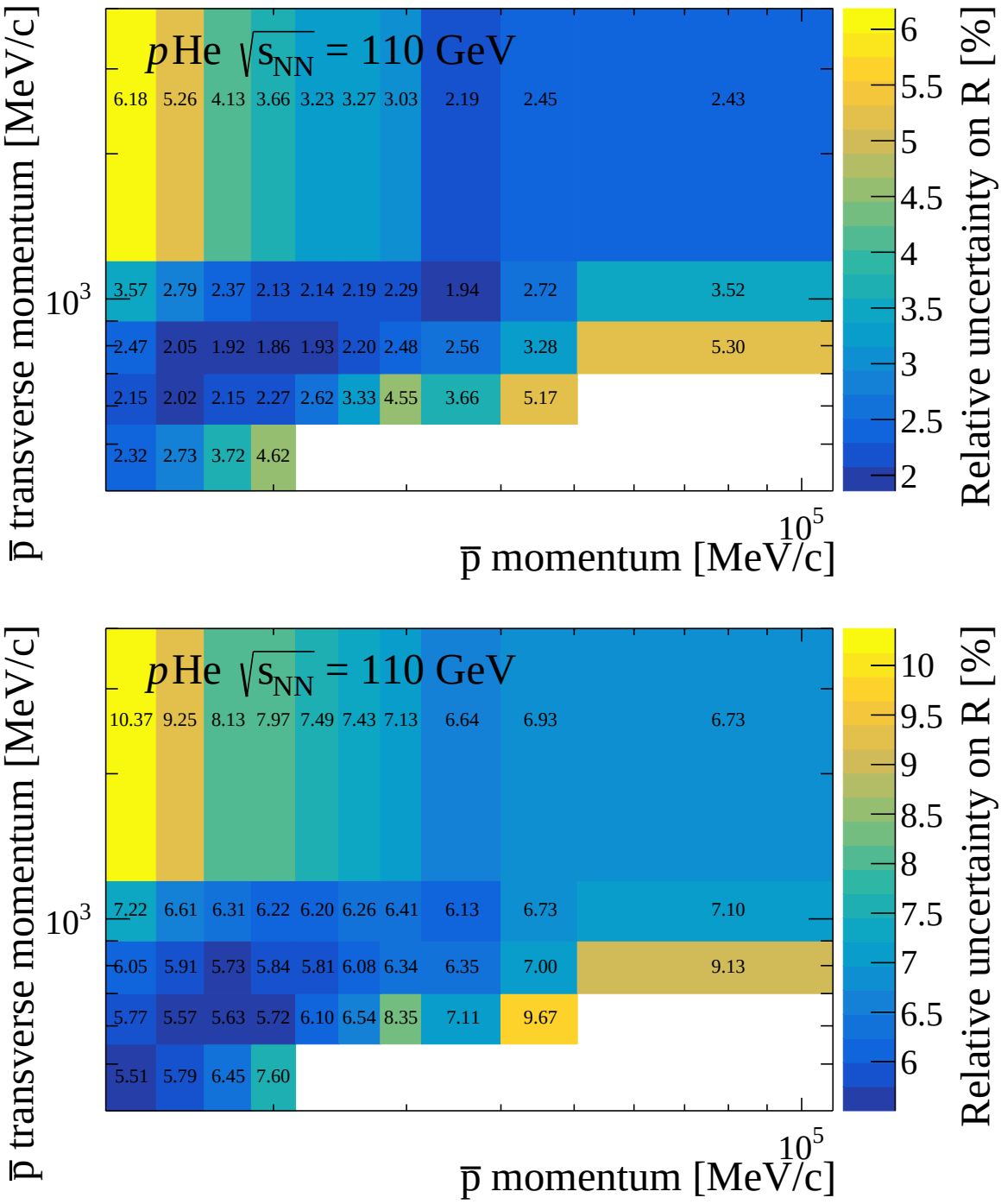


Figure II.1.51: Relative statistical (top) and total (bottom) uncertainty for the detached-to-prompt $\overline{p}$ ratio measured on the LHCb pHe 2016 data as a function of its momentum and transverse momentum.

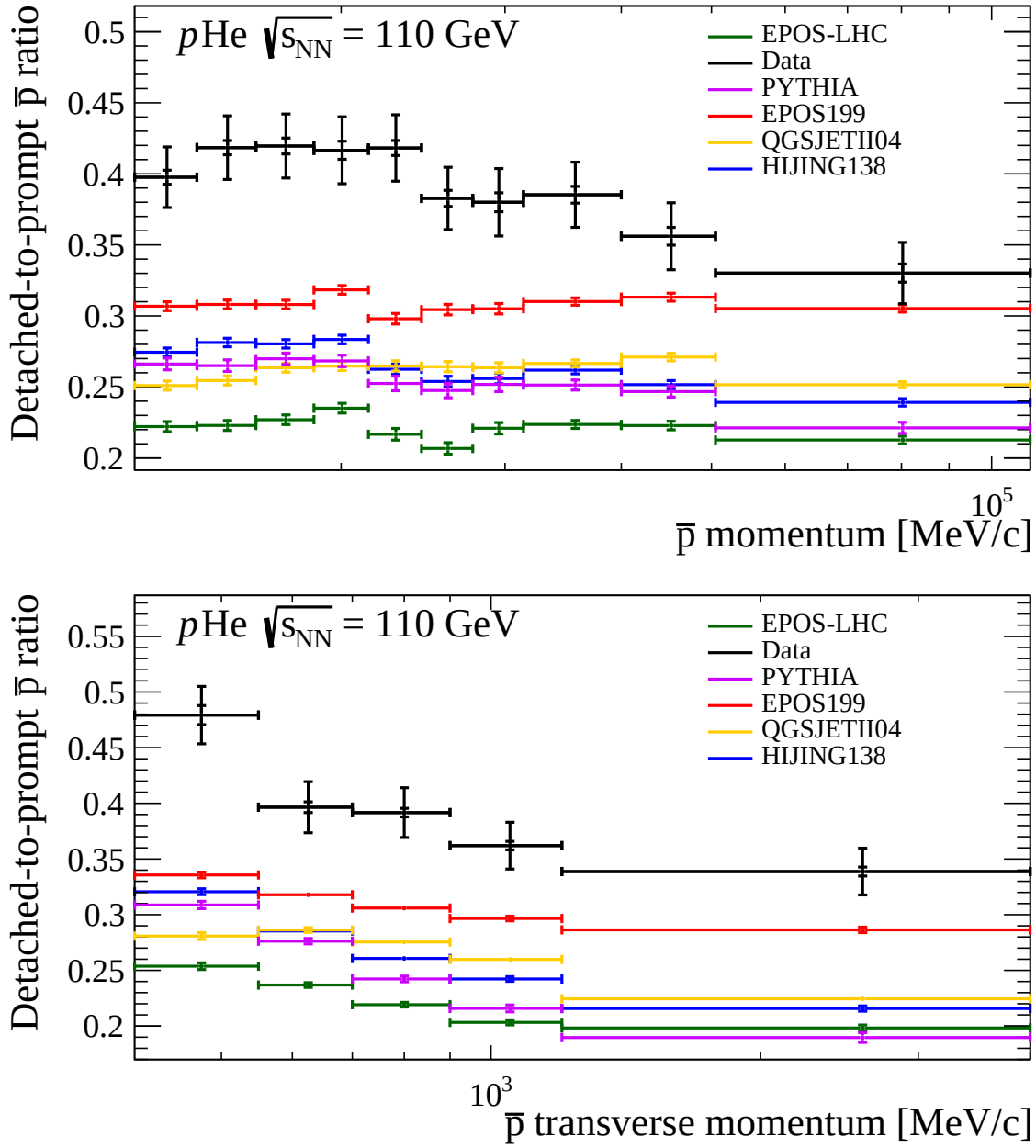


Figure II.1.52: Comparison of the measured detached-to-prompt $\bar{p}$ ratio on the LHCb pHe 2016 data as a function of the momentum (top) and the transverse momentum (bottom) with the predictions of the most common generators used for CR physics. For the measurement, the inner error bars present the fit and efficiency contibutions to the uncertainty only, the outer ones the total uncertainty. While the kinematic evolution of the ratio is consistent with the prediction, an evident underestimation is observed.

Table II.1.15: Numerical results for the detached-to-prompt antiproton ratio in intervals of the antiproton momentum and transverse momentum.

p range [GeV/c]	p_{T} range [GeV/c]	R	δR_{stat}	δR_{tot}	$\frac{\delta R_{stat}}{R}$ [%]	$\frac{\delta R_{tot}}{R}$ [%]
12.0 - 14.0	0.4 - 0.55	0.469	0.011	0.026	2.3	5.5
12.0 - 14.0	0.55 - 0.7	0.429	0.009	0.025	2.1	5.8
12.0 - 14.0	0.7 - 0.9	0.368	0.009	0.022	2.5	6.0
12.0 - 14.0	0.9 - 1.2	0.334	0.012	0.024	3.6	7.2
12.0 - 14.0	1.2 - 4.0	0.301	0.019	0.031	6.2	10.4
14.0 - 16.2	0.4 - 0.55	0.524	0.014	0.030	2.7	5.8
14.0 - 16.2	0.55 - 0.7	0.420	0.008	0.023	2.0	5.6
14.0 - 16.2	0.7 - 0.9	0.384	0.008	0.023	2.1	5.9
14.0 - 16.2	0.9 - 1.2	0.377	0.011	0.025	2.8	6.6
14.0 - 16.2	1.2 - 4.0	0.308	0.016	0.028	5.3	9.3
16.2 - 18.7	0.4 - 0.55	0.502	0.019	0.032	3.7	6.4
16.2 - 18.7	0.55 - 0.7	0.421	0.009	0.024	2.1	5.6
16.2 - 18.7	0.7 - 0.9	0.406	0.008	0.023	1.9	5.7
16.2 - 18.7	0.9 - 1.2	0.370	0.009	0.023	2.4	6.3
16.2 - 18.7	1.2 - 4.0	0.348	0.014	0.028	4.1	8.1
18.7 - 21.4	0.4 - 0.55	0.435	0.020	0.033	4.6	7.6
18.7 - 21.4	0.55 - 0.7	0.448	0.010	0.026	2.3	5.7
18.7 - 21.4	0.7 - 0.9	0.408	0.008	0.024	1.9	5.8
18.7 - 21.4	0.9 - 1.2	0.402	0.009	0.025	2.1	6.2
18.7 - 21.4	1.2 - 4.0	0.340	0.012	0.027	3.7	8.0
21.4 - 24.4	0.55 - 0.7	0.445	0.012	0.027	2.6	6.1

Table II.1.16: Numerical results for the detached-to-prompt antiproton ratio in intervals of the antiproton momentum and transverse momentum.

p range [GeV/c]	p_{T} range [GeV/c]	R	δR_{stat}	δR_{tot}	$\frac{\delta R_{stat}}{R}$ [%]	$\frac{\delta R_{tot}}{R}$ [%]
21.4 - 24.4	0.7 - 0.9	0.440	0.009	0.026	1.9	5.8
21.4 - 24.4	0.9 - 1.2	0.380	0.008	0.024	2.1	6.2
21.4 - 24.4	1.2 - 4.0	0.365	0.012	0.027	3.2	7.5
24.4 - 27.7	0.55 - 0.7	0.389	0.013	0.025	3.3	6.5
24.4 - 27.7	0.7 - 0.9	0.402	0.009	0.024	2.2	6.1
24.4 - 27.7	0.9 - 1.2	0.387	0.008	0.024	2.2	6.3
24.4 - 27.7	1.2 - 4.0	0.318	0.010	0.024	3.3	7.4
27.7 - 31.4	0.55 - 0.7	0.346	0.016	0.029	4.5	8.3
27.7 - 31.4	0.7 - 0.9	0.424	0.010	0.027	2.5	6.3
27.7 - 31.4	0.9 - 1.2	0.389	0.009	0.025	2.3	6.4
27.7 - 31.4	1.2 - 4.0	0.354	0.011	0.025	3.0	7.1
31.4 - 40.0	0.55 - 0.7	0.395	0.014	0.028	3.7	7.1
31.4 - 40.0	0.7 - 0.9	0.399	0.010	0.025	2.6	6.3
31.4 - 40.0	0.9 - 1.2	0.379	0.007	0.023	1.9	6.1
31.4 - 40.0	1.2 - 4.0	0.347	0.008	0.023	2.2	6.6
40.0 - 50.5	0.55 - 0.7	0.323	0.017	0.031	5.2	9.7
40.0 - 50.5	0.7 - 0.9	0.360	0.012	0.025	3.3	7.0
40.0 - 50.5	0.9 - 1.2	0.376	0.010	0.025	2.7	6.7
40.0 - 50.5	1.2 - 4.0	0.360	0.009	0.025	2.5	6.9
50.5 - 110.0	0.7 - 0.9	0.354	0.019	0.032	5.3	9.1
50.5 - 110.0	0.9 - 1.2	0.317	0.011	0.023	3.5	7.1
50.5 - 110.0	1.2 - 4.0	0.332	0.008	0.022	2.4	6.7

II.2 Exclusive approach

II.2.1 Analysis strategy

The inclusive approach described in the previous chapter measures the relative abundance of antiprotons produced in all anti-hyperon decays with respect to the prompt component. A complementary approach to the analysis, exclusively addressing the antiproton production in $\overline{\Lambda}_{\text{prompt}} \to \overline{p}\pi^+$ decays, is described in this chapter. These represent the dominant contribution among all detached antiprotons and can be selected in data with high purity exploiting the efficient reconstruction of the two final-state charged particles and the narrow $\overline{\Lambda}$ invariant mass.
The antiproton production cross-section in $\overline{\Lambda}_{\text{prompt}} \to \overline{p}\pi^+$ decays is measured as:

$$\sigma(p\text{He} \to (\overline{\Lambda}_{\text{prompt}} \to \overline{p}\pi^+)X,\ \sqrt{s_{\text{NN}}} = 110\,\text{GeV}) = \frac{N^{\text{Data}}}{\varepsilon \cdot \mathcal{L}} \tag{II.2.1}$$

being:

- N^{Data} the number of reconstructed and selected signal events in the collected pHe data sample, measured with a fit to the $\overline{\Lambda}$ particle invariant mass distribution;
- ε the experimental efficiency, measured on the simulated sample.
- $\mathcal{L}$ the luminosity of the analysed pHe sample, measured as described in Sec. I.4.2.2.

The ratio between the obtained cross-section results and those for the prompt antiproton production from Ref. [61] provides an independent measurement of the dominant detached-to-prompt antiproton contribution and can cross-check the conclusions of the inclusive approach.

The chapter is organised as follows. In Sec. II.2.2, the simulated samples produced for this analysis are described. These are employed to define a selection algorithm, described in Sec. II.2.3, which enhances the signal statistics over the background. The fit to the invariant mass of the selected $\overline{\Lambda}$ candidates used to measure N^{Data} is described in Sec. II.2.4, all terms contributing to the experimental efficiency in Sec. II.2.5. The contributions to the systematic uncertainty are discussed in Ref. II.2.6 and Sec. II.2.7 finally presents the results for this approach.

Table II.2.1: Selection requirements for the generation of the events in the two samples employed in the exclusive approach to the analysis.

Sample	Selection requirements		
$\Lambda/\overline{\Lambda}$ -enriched	$\Lambda/\overline{\Lambda}$	:	$p_{\mathrm{T}} > 300$ MeV/c, $\eta \in [1.5, 5.9]$
	$p/\overline{p}$, π^-/π^+:		$\eta \in [1.5, 5.9]$
$\overline{\Lambda}$-enriched	$\overline{\Lambda}$	:	$p_{\mathrm{T}} > 300$ MeV/c, $\eta \in [2, 5.5]$
	$\overline{\Lambda}$	:	$ct < 1\,\mathrm{m} \cdot M/[18(E_{CM} + p_{z,CM})]$
	$\overline{p}$, π^+	:	$\eta \in [2, 5.5]$

II.2.2 Simulated samples

Three simulated samples, produced following the same procedure as in Sec. II.1.2, are used for this analysis:

- a minimum-bias one, common to the prompt antiproton production measurement [61] and to the inclusive approach presented in the previous chapter. Although all generated events are accepted, the statistics of the $\overline{\Lambda}$ decays is insufficient;
- a sample (called $\Lambda/\overline{\Lambda}$ -enriched in the following), originally produced to study the tracking and particle identification efficiencies in fixed-target data. As summarised in Tab. II.2.1, this sample is enriched with $\Lambda \to p\pi^-$ and $\overline{\Lambda} \to \overline{p}\pi^+$ candidates. Events are accepted only when all particles are produced in the LHCb pseudorapidity acceptance and the Λ or $\overline{\Lambda}$ particle present a minimum 300 MeV/c transverse momentum;
- a sample (called $\overline{\Lambda}$-enriched in the following), specifically produced for this analysis, where antiprotons reconstructed combining the information from all tracking detectors are measured. As summarised in Tab. II.2.1, only $\overline{\Lambda} \to \overline{p}\pi^+$ decays are accepted in this case, stricter pseudorapidity thresholds are set and a constraint on the $\overline{\Lambda}$ particle time of flight is added. Decays occurring downstream the VELO detector, which can not be reconstructed in this analysis, are removed.

As listed in Tab. II.2.2, approximately 5 millions events (N_{evts}) are accepted by the respective applied selections. Considering the fraction of events surviving the generator cuts ($\varepsilon_{\mathrm{generator}}$), the corresponding ratio for the number of $\overline{\Lambda}$ particles ($N_{\overline{\Lambda}}$) between the two samples is:

$$\frac{N_{\overline{\Lambda}}(\overline{\Lambda})}{N_{\overline{\Lambda}}(\Lambda/\overline{\Lambda})} = \frac{N_{\mathrm{evts}}(\overline{\Lambda})/\varepsilon_{\mathrm{generator}}(\overline{\Lambda})}{N_{\mathrm{evts}}(\Lambda/\overline{\Lambda})/\varepsilon_{\mathrm{generator}}(\Lambda/\overline{\Lambda})} = 6.09 \pm 0.06. \tag{II.2.2}$$

To measure the generation and simulation efficiencies, as described in Sec. II.2.5.1 and II.2.5.2, respectively, three samples corresponding to those above but with a disabled simulation of the particles propagation through the detector are also produced.

Table II.2.2: Number of accepted events and fraction of the events surviving the selection requirements listed in Tab. II.2.1 for the two simulated samples employed in the exclusive approach to the analysis.

Simulated sample	N_{evts}	$\varepsilon_{\text{generator}}$
$\Lambda/\overline{\Lambda}$ -enriched	5130327	0.0808(8)
$\overline{\Lambda}$-enriched	5037822	0.01304(5)

II.2.3 $\overline{\Lambda}_{\text{prompt}} \to \overline{p}\pi^+$ decays selection

A selection algorithm to enhance the statistics of $\overline{\Lambda}_{\text{prompt}} \to \overline{p}\pi^+$ decays over random combination of tracks is applied to pHe data and simulation in two steps, summarised in Tab. II.2.3 and II.2.4, respectively.
After the preselection, requiring as in the previous chapter the trigger configuration discussed in Sec. I.4.2.2, a maximum of 5 reconstructed tracks with negative pseudorapidity and a reconstructed PV with $z \in [-700, 100]$ mm, the signal decay is defined. High-quality reconstructed particles satisfying loose kinematic requirements are associated to the antiproton and pion hypotheses and then combined to form a $\overline{\Lambda}$ when their χ^2_{DOCA} is below 10 and when the $\overline{\Lambda}$ DV is well reconstructed. The resulting candidate is only accepted when the reconstructed invariant mass shows a maximum distance from the nominal $\overline{\Lambda}$ value of 40 MeV/c^2, an interval which is chosen to be larger with respect to the expected mass resolution in order to also allow background events to be modelled with a parametric function. Two strategies are compared to exploit in the selection the final-state particles detachment with respect to the PV, consequent to the $\overline{\Lambda}$ flight before the decay. These consider:

- the antiproton and pion candidates χ^2_{IP}, as performed in Sec. II.1.6.4;
- a linear combination of the logarithms of all particles IPs, firstly studied in Ref. [132] and defined as:

$$\mathcal{F}_{\text{IP}} = log(IP(\overline{p})) + log(IP(\pi^+)) - log(IP(\overline{\Lambda})). \tag{II.2.3}$$

  As the three IP variables are approximately linearly correlated, the $\mathcal{F}_{\text{IP}}$ discriminant has close-to-optimal performance.

The distributions of the π^+ χ^2_{IP} (left) and of $\mathcal{F}_{\text{IP}}$ (right) are compared in Fig. II.2.1 for all candidates selected with the requirements in Tab. II.2.3 except the detachment one (in blue) and for the signal ones only (in red), identified requiring that all reconstructed particle identities match the simulated ones. Presenting a larger discriminating power, the $\mathcal{F}_{\text{IP}}$ cut is chosen. After the application of all selection requirements, the $\overline{\Lambda}$ candidates invariant mass distributions for all the considered samples are shown in Fig. II.2.2. As expected, a sizeable contamination from background events is observed. This can also be seen in the left Fig. II.2.3, describing the decay in the Armenteros-Podolanski plot defined as in Eq. II.1.26. Other than an elliptic pattern from signal events, a background smooth distribution is observed.
A more detailed description of the signal decay is performed to reduce the background

Table II.2.3: Preselection of $\overline{\Lambda}_{\text{prompt}} \to \overline{p}\pi^+$ decays.

Prefilters	$N_{SPD} > 0$, $N_{VELO} > 0$ $N_{\text{trk}}(\eta < 0) < 5$, $\text{PV}_z \in [-700, 100]$ mm
$\overline{p}$, π^+ $\overline{p}$ π^+	track $\chi^2/\text{ndf} < 5$, $GhostProb < 0.4$ $p > 8\,\text{GeV}/c$, $p_{\text{T}} > 0.3\,\text{GeV}/c$ $p > 2\,\text{GeV}/c$
Combination	$M \in [1, 1.2]\,\text{GeV}/c^2$, $\chi^2_{\text{DOCA}} < 10$
$\overline{\Lambda}_{\text{prompt}}$	$\lvert M - M(\overline{\Lambda})\rvert < 40\,\text{MeV}/c^2$, DV $\chi^2 < 16$
$\mathcal{F}_{\text{IP}}$	$log(IP(\overline{p})) + log(IP(\pi^+)) - log(IP(\overline{\Lambda})) > 0$

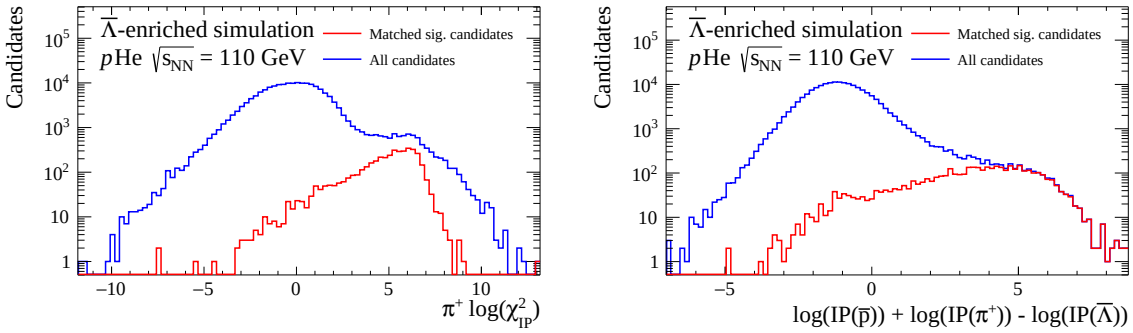


Figure II.2.1: Comparison between the distributions of the π^+ candidates $log(\chi^2_{\text{IP}})$ (left) and $\mathcal{F}_{\text{IP}}$ (right) for all events selected in the $\overline{\Lambda}$-enriched simulated sample with all cuts in Tab. II.2.3 except the detachment one (blue) and for the signal ones only (red), defined requiring all reconstructed and simulated particle hypotheses to coincide. The latter variable offers a higher discrimination power and is thus chosen for the selection algorithm.

contamination. This is required to be as symmetric as possible to the one applied in the prompt antiproton production measurement, so that most of the uncertainty is removed when performing the $\overline{\Lambda}_{\text{prompt}}$-to-prompt antiproton ratio. The applied requirements, listed in Tab. II.2.4, are grouped as:

- Quality and geometry cuts, to remove random associations between a pion and an antiproton candidate. The detachment cuts are tightened considering the particles IPs, χ^2_{IP}, $\mathcal{F}_{\text{IP}}$ and $\mathcal{F}_{\chi^2_{\text{IP}}}$, being the last defined with the same formula as $\mathcal{F}_{\text{IP}}$ but considering the particles χ^2_{IP}. The $\overline{\Lambda}$ candidates are required to decay before the end of the VELO detector and to point to the PV through their DIRA;

- Fiducial region cuts, to restrict the kinematic and geometric phase-spaces to match those considered in the prompt analysis. The momentum and transverse momentum thresholds for the antiproton candidate are tightened and the reconstructed PV is required to be compatible with the beam geometry other than to lie in the range $z \in [-700, 100]$ mm;

- a veto to suppress the contamination from $K^0_{\text{S}} \to \pi^+\pi^-$ decays where the negatively-charged pion is incorrectly identified as an antiproton. Similarly to Sec. II.1.6.4,

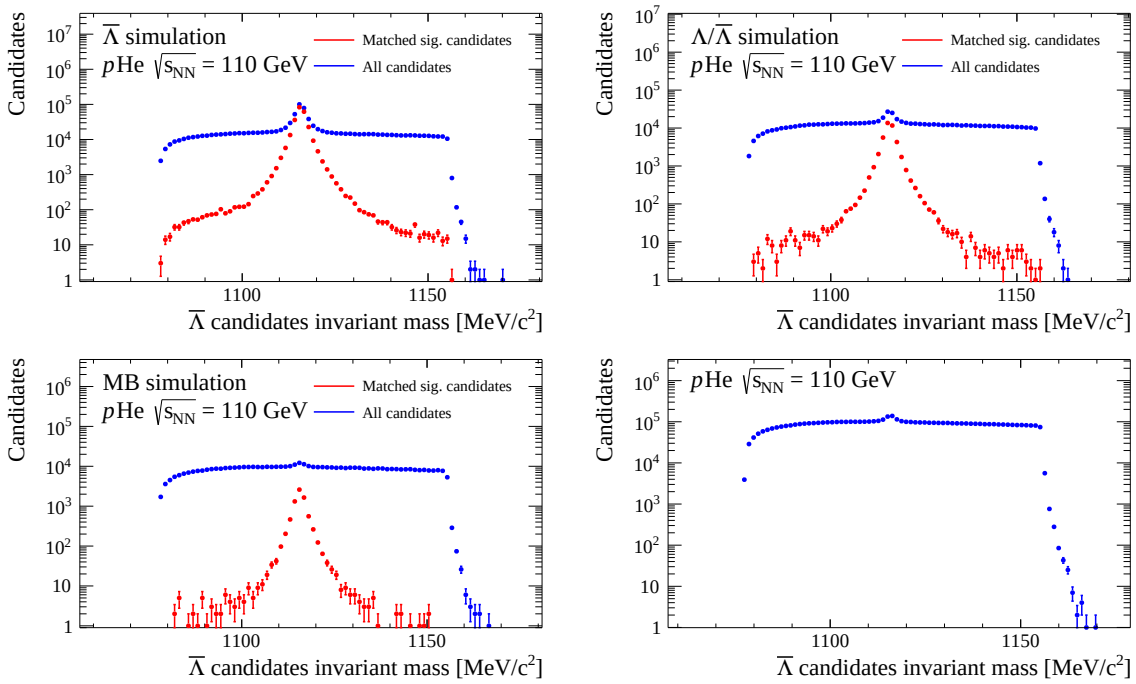


Figure II.2.2: Invariant mass distribution for the $\overline{\Lambda}$ candidates in the $\overline{\Lambda}$-enriched (top left), the $\Lambda/\overline{\Lambda}$ -enriched (top right), the minimum-bias (bottom left) simulations and in the data sample (bottom right) after the application of the selection algorithm listed in Tab. II.2.3. In all histograms, blue markers indicate all events passing the selection criteria, red the signal ones only identified by requiring the simulated identities of all the particles to match $\overline{\Lambda} \to \overline{p}\pi^+$.

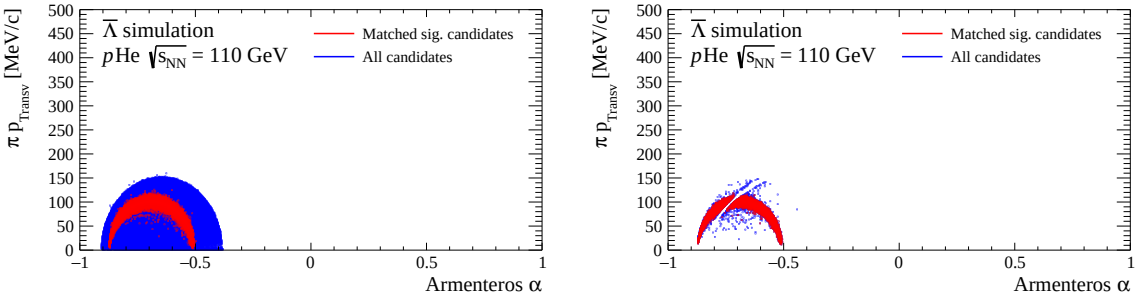


Figure II.2.3: Armenteros-Podolanski plot for the $\overline{\Lambda} \to \overline{p}\pi^+$ decay in the $\overline{\Lambda}$-enriched simulated sample before (left) and after (right) the application of the offline selection requirements but the Armenteros-Podolanski one. Blue markers indicate all candidates, red the signal ones only.

this is done by re-evaluating the invariant mass replacing the antiproton with the pion mass hypothesis, obtaining the Wrong Mass (WM). As presented in Fig. II.2.4 for the $\overline{\Lambda}$-enriched simulated sample, an enhanced peak centred in the K^0_S nominal mass value is found over a continuous distribution. This is roughly modelled with a Gaussian with mean μ and width σ and the values in $\mu \pm 2\sigma$ are removed.

After the application of all the selection requirements, the $\overline{\Lambda}$ candidates invariant mass distributions in all the considered samples and the Armenteros-Podolanski plot for the $\overline{\Lambda}$-enriched simulation are presented in Fig. II.2.5 and in right Fig. II.2.3, respectively. The signal significance is increased with respect to Fig. II.2.2, but a contamination of

Table II.2.4: Selection algorithm applied offline for $\overline{\Lambda}_{\text{prompt}} \to \overline{p}\pi^+$ decays.

Quality cuts	$\overline{\Lambda}$:	$\lvert \text{DV}_z \rvert < 700\,\text{mm}$, $log(cos(DIRA)) > -2 \cdot 10^{-5}$, IP$< 5\,\text{mm}$
	PV :	$\sqrt{[(x\text{-}(0.81\,\text{mm-}z \cdot 490))^2 + (y\text{-}(\text{-}0.22\,\text{mm-}z \cdot 45))^2]} < 0.4\,\text{mm}$
Decay geometry	$\overline{p}$:	$log(\chi^2_{\text{IP}}) > 1$
	π^+ :	$log(\chi^2_{\text{IP}}) > 2$
	$\mathcal{F}$:	$\mathcal{F}_{\text{IP}} > 1.5$, $\mathcal{F}_{\chi^2_{\text{IP}}} > 4$
Fiducial region	$\overline{p}$:	$p \in [12, 110]\,\text{GeV}/c$, $p_{\text{T}} \in [0.4, 4]\,\text{GeV}/c$, $2 < \eta < 5$
	π^+ :	$2 < \eta < 5.5$
	$\overline{\Lambda}$:	$2 < \eta < 5.5$ & $p_{\text{T}} > 300\,\text{MeV}/c$
K^0_{S} veto		$WM < 490$ or $511 < WM < 580\,\text{MeV}/c^2$
Armenteros		$\left\lvert \left(\frac{\alpha+0.69}{0.18}\right)^2 + \left(\frac{p_\pi \cdot sin\theta_{\pi\overline{\Lambda}}}{100.4\,\text{MeV}/c}\right)^2 - 1 \right\rvert < 0.39$

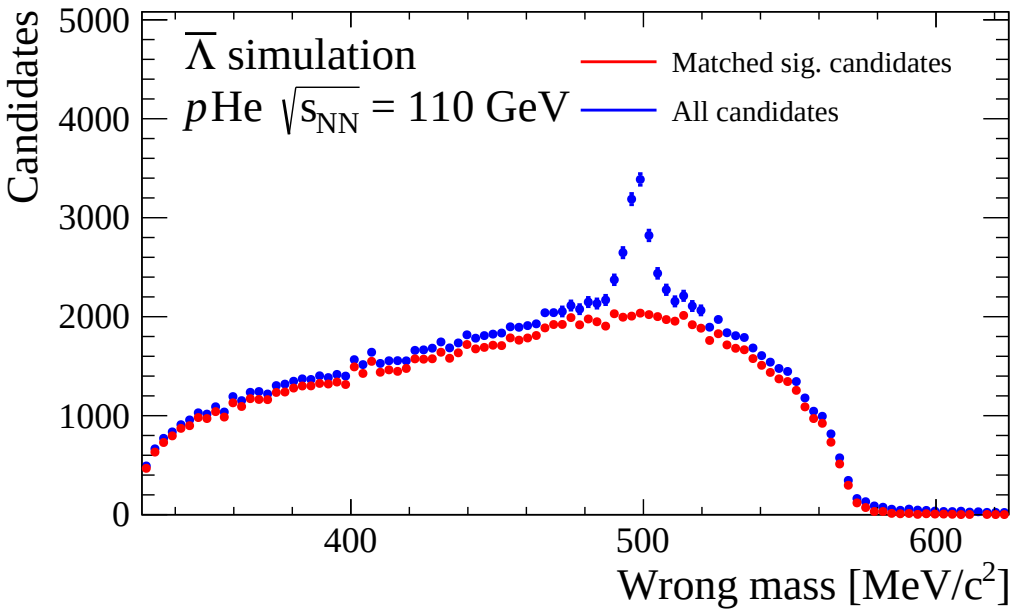


Figure II.2.4: Distribution before the K^0_{S} veto cut application for the wrong mass in the $\overline{\Lambda}$-enriched simulated sample. Over a continuous distribution of signal candidates, an enhanced peak centred in the nominal K^0_{S} value can be seen.

candidates classified as background can still be observed. In most of these events, called unmatched in the following, one of the two final-state tracks is not uniquely matched to a simulated particle and the simulated $\overline{\Lambda}$ identity is null. Their distribution for the $\overline{\Lambda}$ invariant mass is illustrated in Fig. II.2.6 and a mixture of signal and background events can be seen. In order to distinguish the two contributions and optimise a further selection in the Armenteros-Podolanski plot, the binned distribution is modelled with the sum of a Breit-Wigner for the signal (in orange) and a first-order polynomial function (in red) for the background. The values obtained from the fit for the Breit-Wigner mean μ and width

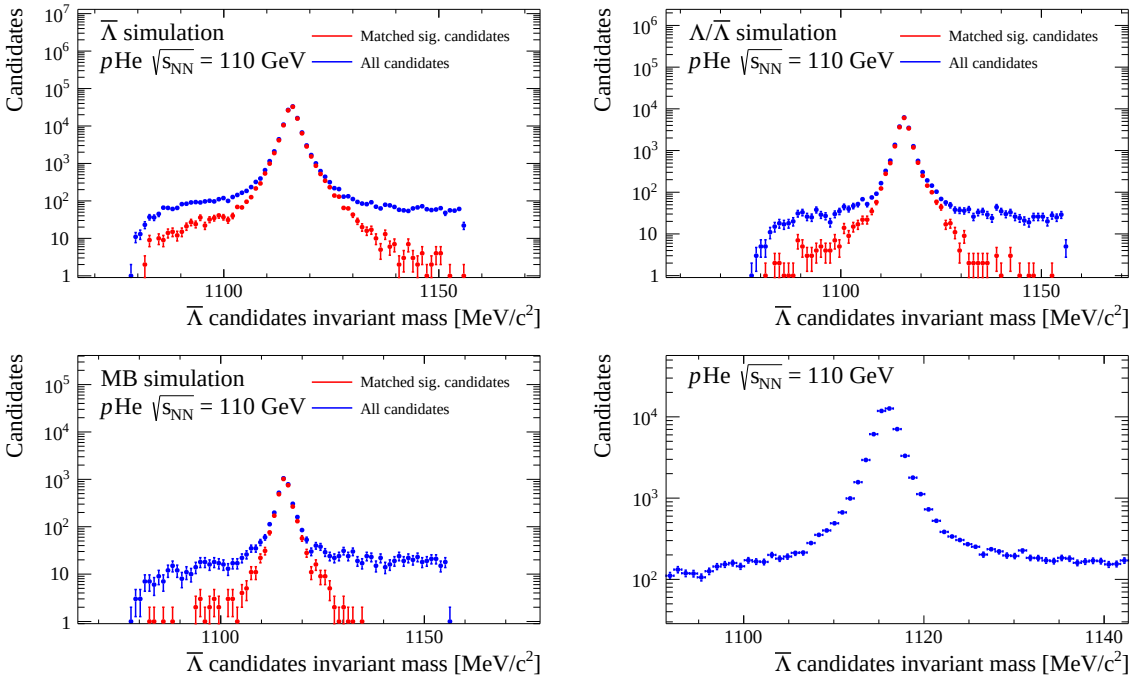


Figure II.2.5: Invariant mass distribution for the $\overline{\Lambda}$ candidates in the $\overline{\Lambda}$-enriched (top left), in the $\Lambda/\overline{\Lambda}$ -enriched (top right), in the minimum-bias (bottom left) simulations and in the data sample (bottom right) after the full selections except the Armenteros-Podolanski one. In all histograms, blue markers indicate all selected events, red the signal ones only identified by requiring the true identity of all the particles to correspond to the $\overline{\Lambda} \to \overline{p}\pi^+$ ones. By comparing with the distributions of Fig. II.2.2, a large background reduction is achieved.

σ are used to define a $\mu \pm 2\sigma$ region, where, according to the fitted model, 87% of the unmatched signal and $\sim 23\%$ of the background candidates lie. Unmatched candidates in this region are added to the matched ones in order to optimise the signal-to-noise ratio of a further selection in the Armenteros-Podolanski plot. The contamination from unmatched background events, when referred to the total simulated signal statistics in the $\overline{\Lambda}$-enriched sample, is $\sim 0.5\%$ and is not expected to introduce a significant uncertainty. A non-linear transformation is applied to the Armenteros-Podolanski plot axes to map the theoretical ellipse for the signal distribution into a circumference with a radius equal to one. A selection on the distance r_α from such a circumference is set as:

$$\left| \left(\frac{\alpha + 0.69}{0.18} \right)^2 + \left(\frac{p_\pi \cdot sin\theta_{\pi\bar{\Lambda}}}{100.4\,\mathrm{MeV}/c} \right)^2 - 1 \right| < r_\alpha. \qquad \text{(II.2.4)}$$

A scan of 50 r_α values between 0 and 1 is performed and that indicated in Tab. II.2.4, optimising the signal-to-noise ratio, is chosen. Figure II.2.7 shows the comparison between the invariant mass distributions (left) and the Armenteros-Podolanski plots (right) in the $\overline{\Lambda}$-enriched simulated sample before (top) and after (bottom) the Armenteros-Podolanski selection. A large suppression of the combinatorial background is achieved, but its residual distribution as a function of the $\overline{\Lambda}$ invariant mass is observed to be distorted. When evaluating the signal fraction in the collected data, this effect will be carefully studied as a potential source of systematic uncertainty. The cut efficiency, measured on the $\overline{\Lambda}$-enriched

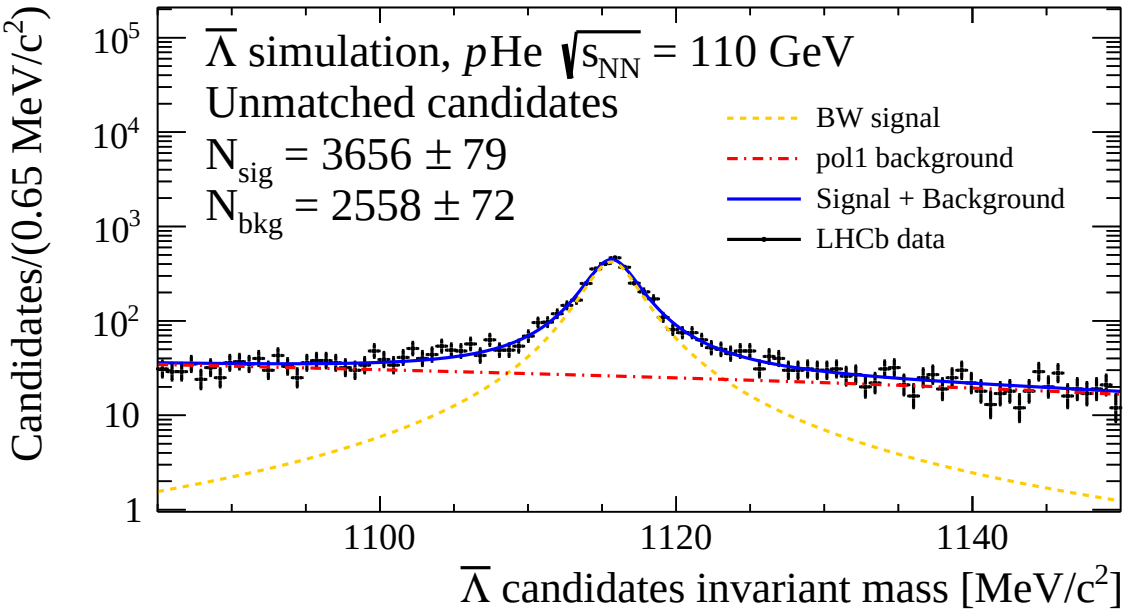


Figure II.2.6: Invariant mass distribution for the $\overline{\Lambda}$ candidates in the $\overline{\Lambda}$-enriched simulated sample where the matching between the simulated proton or pion particle and the reconstructed track is failed. To distinguish the signal and background contributions, the distribution is modelled with the sum of a Breit-Wigner for the signal (in orange) and a first order polynomial function (in red) for the background. The resulting number of signal and background events is indicated.

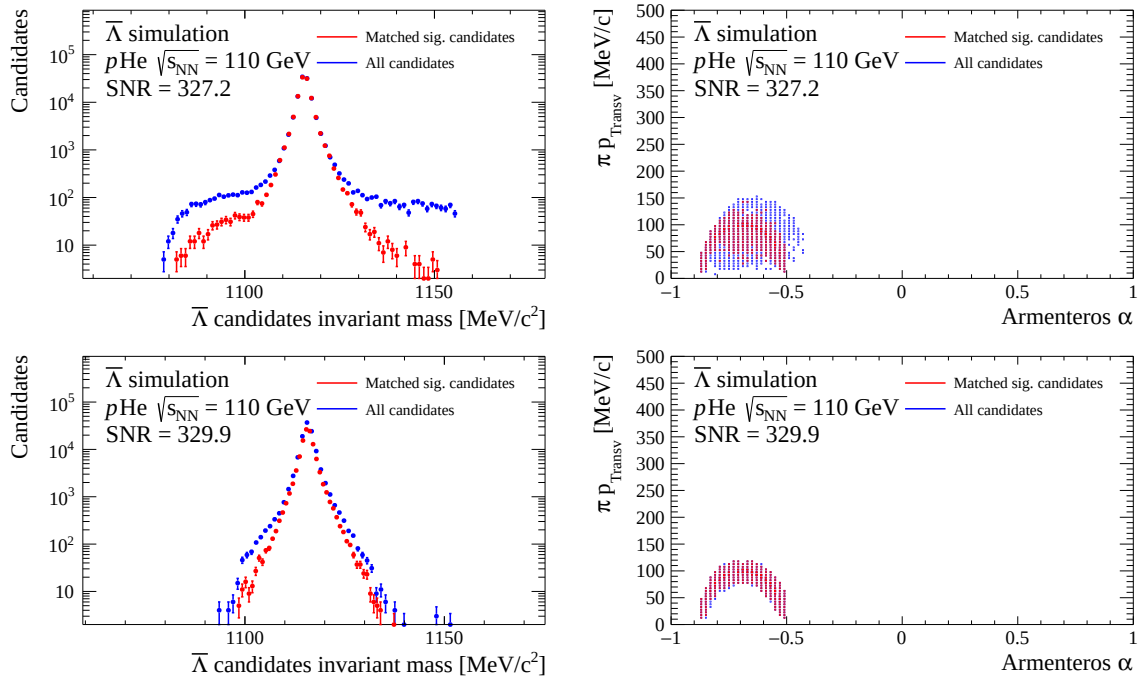


Figure II.2.7: Comparison between the invariant mass of the $\overline{\Lambda}$ candidates (left) and the Armenteros-Podolanski plots (right) before (top) and after (bottom) the Armenteros-Podolanski selection in the $\overline{\Lambda}$-enriched simulated sample. In all plots, blue markers indicate all the events, red the sum of the matched signal and the unmatched in the $\overline{\Lambda}$ mass peak ones.

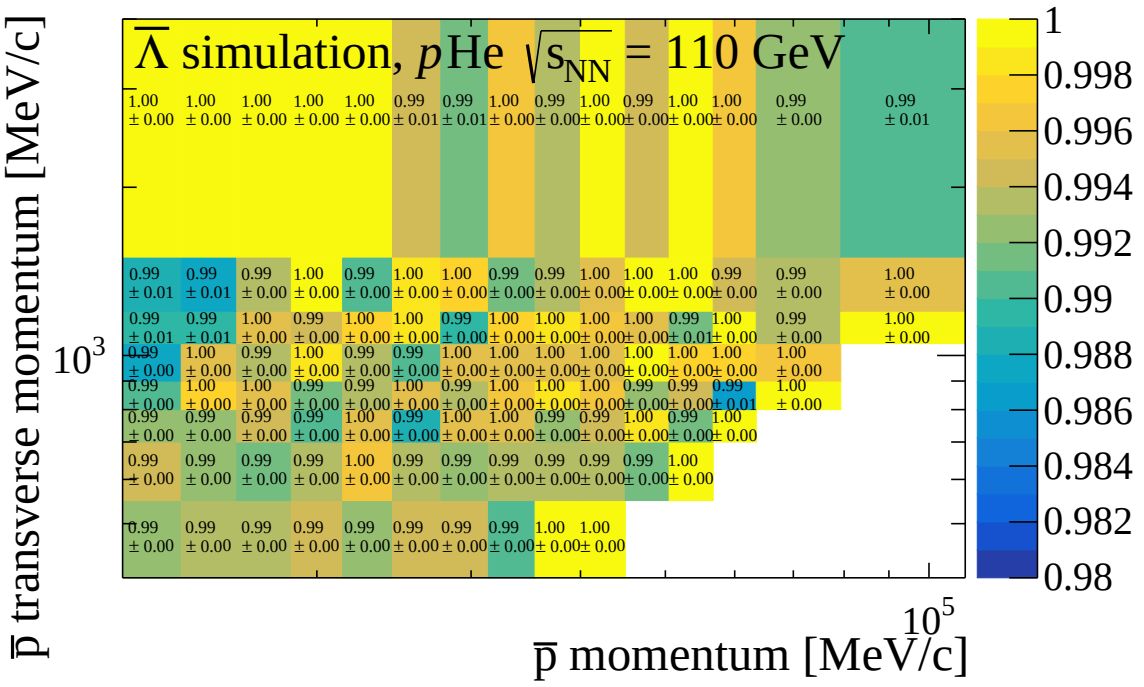


Figure II.2.8: Armenteros-Podolanski cut application efficiency in bins of the antiproton momentum and transverse momentum in the $\overline{\Lambda}$-enriched simulated sample. No evident bias on the kinematic distribution is set.

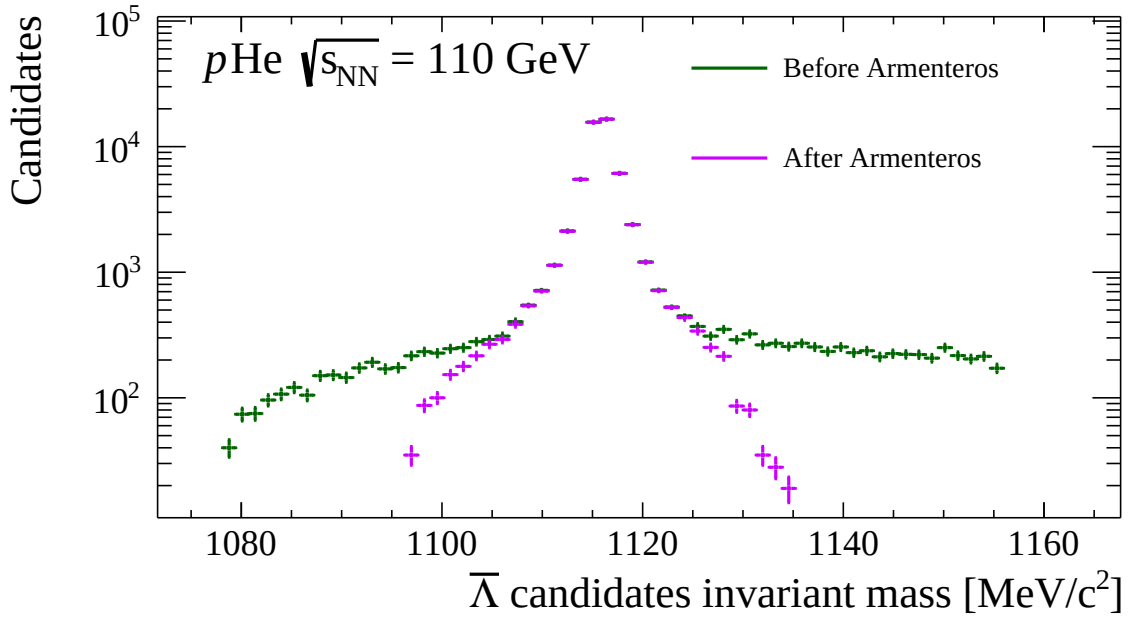


Figure II.2.9: Invariant mass distribution for the $\overline{\Lambda}$ candidates in the data sample before (green) and after (violet) the Armenteros-Podolanski selection. As demonstrated on the simulated samples, it is fully efficient on the signal but causes a distortion of the background mass shape.

simulated sample as the fraction of signal events passing the cut, is:

$$\varepsilon_{\text{Arm}} = (99.37 \pm 0.02)\% \tag{II.2.5}$$

As illustrated in Fig. II.2.8, this is also measured in the antiproton momentum and transverse momentum intervals that will be considered for the analysis in order to exclude biases in the kinematic distribution due to this selection. No relevant effect is observed. The comparison on the pHe data of the $\overline{\Lambda}$ invariant mass before (in green) and after (in violet) the Armenteros-Podolanski cut is shown in Fig. II.2.9 and, also in this case, it is found to be fully efficient on the signal peak and to produce an evident distortion of the background.

After the application of the selection algorithm, the normalised distributions for some variables describing the signal decay in the simulated and data samples are compared in order to exclude biases. This is done before and after the application of the Armenteros-Podolanski selection, in order to separately verify that the distortion effect is the same. Figure II.2.10 presents the results for the antiproton momentum (top left), transverse momentum (top right), pseudorapidity (middle left), PV_z (middle right) and number of reconstructed tracks (bottom left) and of hits in the SPD detector (bottom right). A precise description of data is achieved for the antiproton kinematics and for the event geometry, while differences arise with respect to the detector occupancy as a consequence of the assumptions on the detector material in the simulated sample. Since the discrepancies are reduced when comparing the number of tracks rather than of the hits in the SPD detector, these are mostly due to low-energy processes. The consequences on the reconstruction efficiency are thus expected to be negligible. Figure II.2.11 shows the same comparisons considering the $\overline{\Lambda}$ candidates invariant mass before (left) and after (right) the Armenteros-Podolanski selection. Following the different generator cuts applied to produce the samples, a different background fraction is observed with respect to the signal peak in the left plot. While the minimum-bias sample well reproduces both data signal and background, the others have the latter suppressed. This is in both enriched samples independent on the invariant mass, being the shapes well simulated. The same conclusions are valid after the Armenteros-Podolanski selection, but are less evident as a consequence of the background shape distortion.

II.2.4 Signal extraction

In order to count the number of reconstructed and selected signal events, the $\overline{\Lambda}$ candidates invariant mass distribution is modelled with a parametric function. This is performed on both data and simulation, considering that, as discussed in previous section, $\sim 6\%$ of the simulated candidates have a failed matching between a simulated particle and a reconstructed track. The background distortion caused by the selection in the Armenteros-Podolanski plot is also separately studied by modelling the distributions before and after the cut application. With the measurement of the efficiency performed in Eq. II.2.5, a relation:

$$N^{\text{Data}}_{\text{After Arm}} = N^{\text{Data}}_{\text{Before Arm}} \cdot \varepsilon_{\text{Arm}}, \tag{II.2.6}$$

being $N^{\text{Data}}_{\text{Before (After) Arm}}$ the number of fitted signal events before (after) the Armenteros-Podolanski selection, is also expected to be found on data.

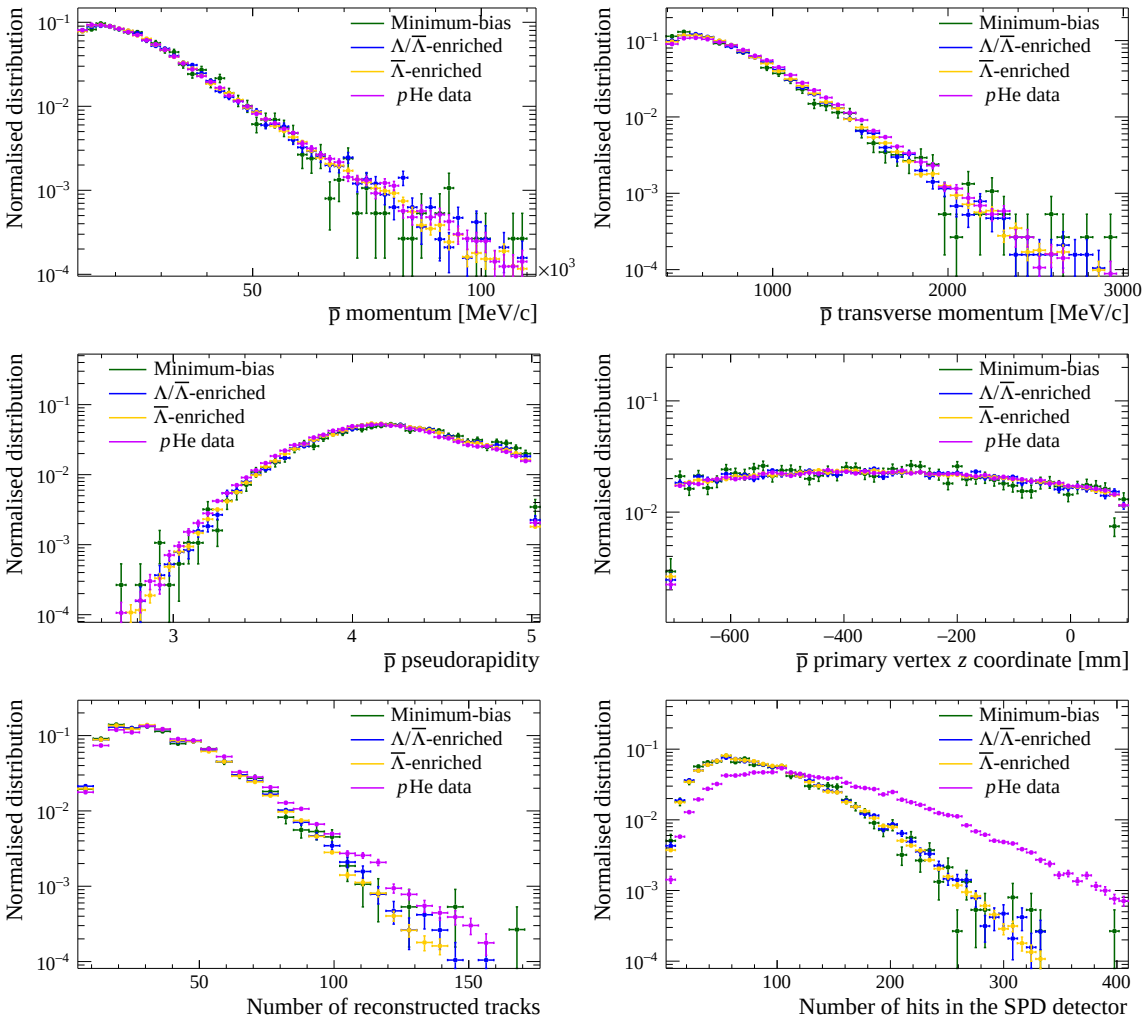


Figure II.2.10: Comparison among the data and simulated normalised distributions of the antiproton momentum (top left), transverse momentum (top right), pseudorapidity (middle left), z coordinate of the primary vertex (middle right) and number of reconstructed tracks (bottom left) and of hits in the SPD detector (bottom right). A precise description of data for the event kinematics and geometry is achieved, but not for the detector occupancy because of the limited simulation of low-energy processes.

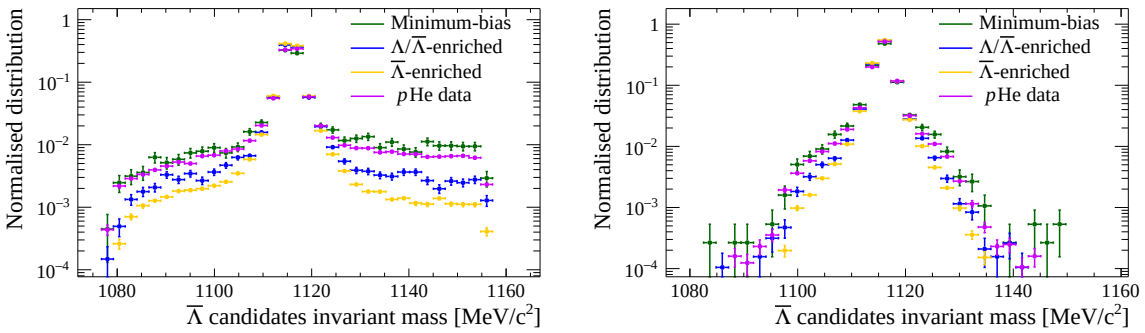


Figure II.2.11: Comparison among the data and simulated normalised distributions for the $\overline{\Lambda}$ candidates invariant mass before (left) and after (right) the Armenteros-Podolanski selection.

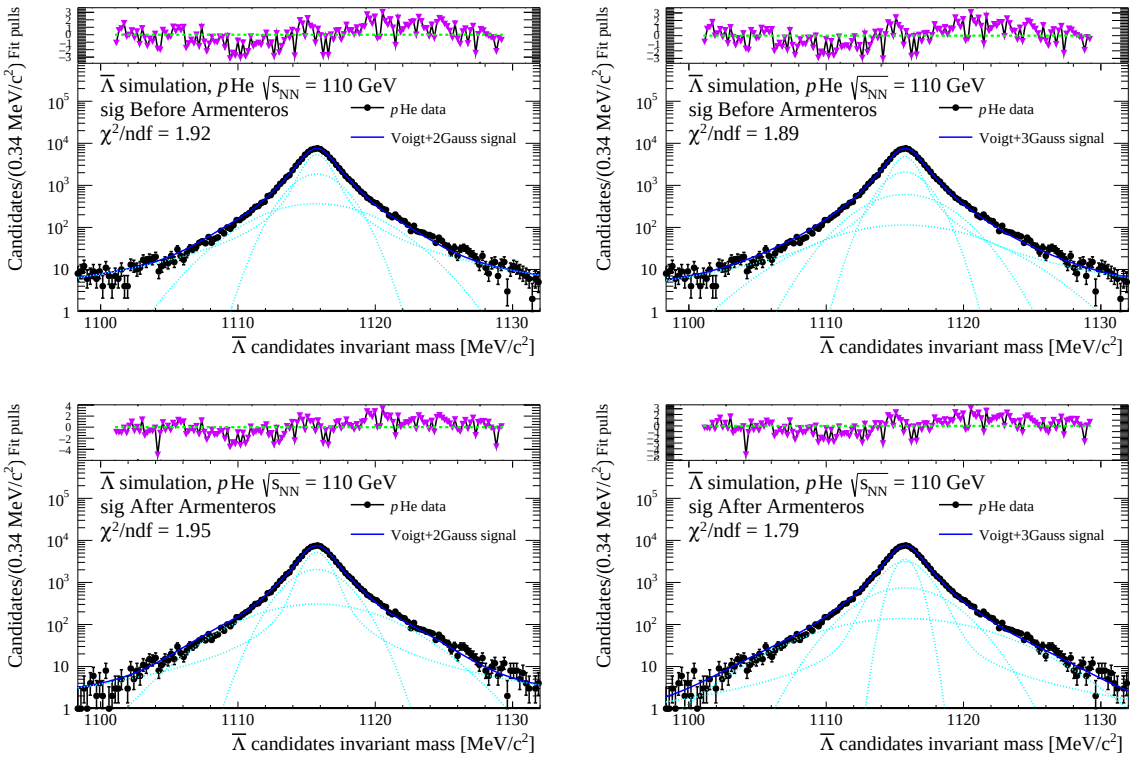


Figure II.2.12: Parametrization of the invariant mass distribution of the matched signal $\overline{\Lambda}$ particles in the $\overline{\Lambda}$-enriched simulated sample with the sum of a Voigtian and two (left) or three (right) Gaussian distributions before (top) and after (bottom) the Armenteros-Podolanski selection. Both fits show a fluctuating distribution of the pulls around zero, as presented in the top part and a similar fit χ^2/ndf, as indicated in all plots. The model with a Voigtian and two Gaussians, offering a lower number of free parameters, is chosen.

The simulation is employed to test different parametric functions for the modelling of signal and background events. To exclude possible contaminations of the mass shapes, unmatched candidates are neglected for this study. Regarding the signal, physics-inspired models are considered, such as a Breit-Wigner distribution and its convolution with a Gaussian (Voigtian). Both are found not to correctly describe the data. The model is thus complicated by adding to the Voigtian two (left) or three (right) Gaussian distributions centred on the same mean value, as presented in Fig. II.2.12 before (top) and after (bottom) the Armenteros-Podolanski selection. Given that the pull distribution and the fit χ^2/ndf are similar between the two models, that with the lower number of parameters is chosen. The other, including an additional Gaussian, will be used to measure the systematic uncertainty due to the choice of a parametrization in Sec. II.2.6.2.

Concerning the background, as illustrated in Fig. II.2.13, the available statistics in the selected sample makes a comparison of more parametric models only meaningful before the Armenteros-Podolanski selection. A model with a simple second-order polynomial and a physics-motivated model, an Argus function

$$\text{Argus}(m, m_0, c, p) = \mathcal{N} \cdot m \cdot \left[1 - \left(\frac{m}{m_0}\right)^2\right]^p \cdot \exp\left[c \cdot \left(1 - \left(\frac{m}{m_0}\right)^2\right)\right], \qquad \text{(II.2.7)}$$

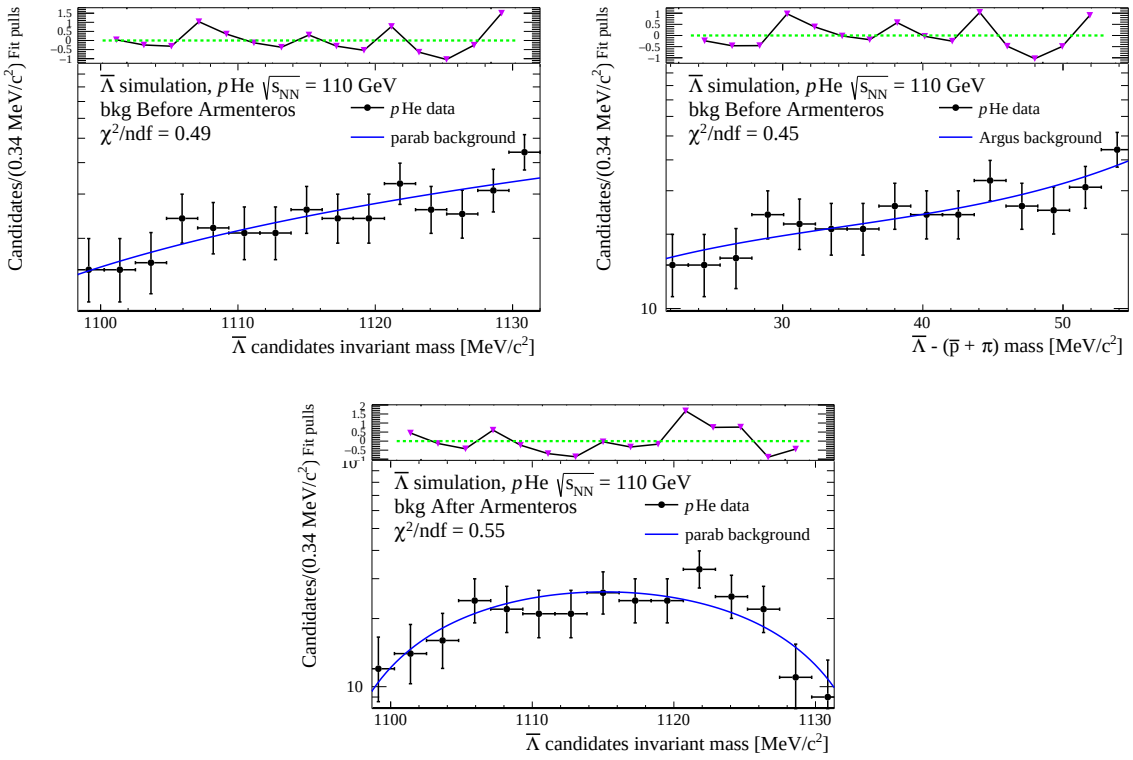


Figure II.2.13: Parametrization of the background candidates invariant mass distribution in the $\overline{\Lambda}$-enriched simulated sample with a polynomial (left) and an Argus function (right) before (top) and with a polynomial function only (bottom) after the Armenteros-Podolanski cut application.

with m the difference between the $\overline{\Lambda}$ invariant mass and the $\overline{p}$ and π^+ nominal values and all the other symbols free parameters of the fit, are employed. The results are presented in the top plots of Fig. II.2.13 and both achieve a similar quality. In order to employ the same model before and after the Armenteros-Podolanski selection and isolate its distortion effect, the polynomial function model, also offering a good description of simulated background candidates after the Armenteros-Podolanski selection (bottom Fig. II.2.13), is chosen.

The model with one Voigtian and two Gaussian distributions for the signal and a second-order polynomial function for the background are combined in a unique function and applied to the $\Lambda/\overline{\Lambda}$ and the $\overline{\Lambda}$-enriched simulated samples, as shown in top and bottom plot of Fig. II.2.14, respectively. With these fits, independently on the matching between the reconstructed and simulated quantities, the number of signal events is measured. Assuming that in both samples the signal reconstructed and selected events other than those satisfying the respective generation requirements are a negligible fraction, the ratio between the fitted signal abundances in the two samples is expected to reproduce that between the equivalent minimum-bias statistics in Eq. II.2.2. With the results displayed in Fig. II.2.14:

$$\frac{N_{\overline{\Lambda}}(\overline{\Lambda}-\mathrm{MC})}{N_{\overline{\Lambda}}(\Lambda/\overline{\Lambda}-\mathrm{MC})} = \frac{109141 \pm 1047}{18264 \pm 284} = 5.98 \pm 0.11. \qquad \text{(II.2.8)}$$

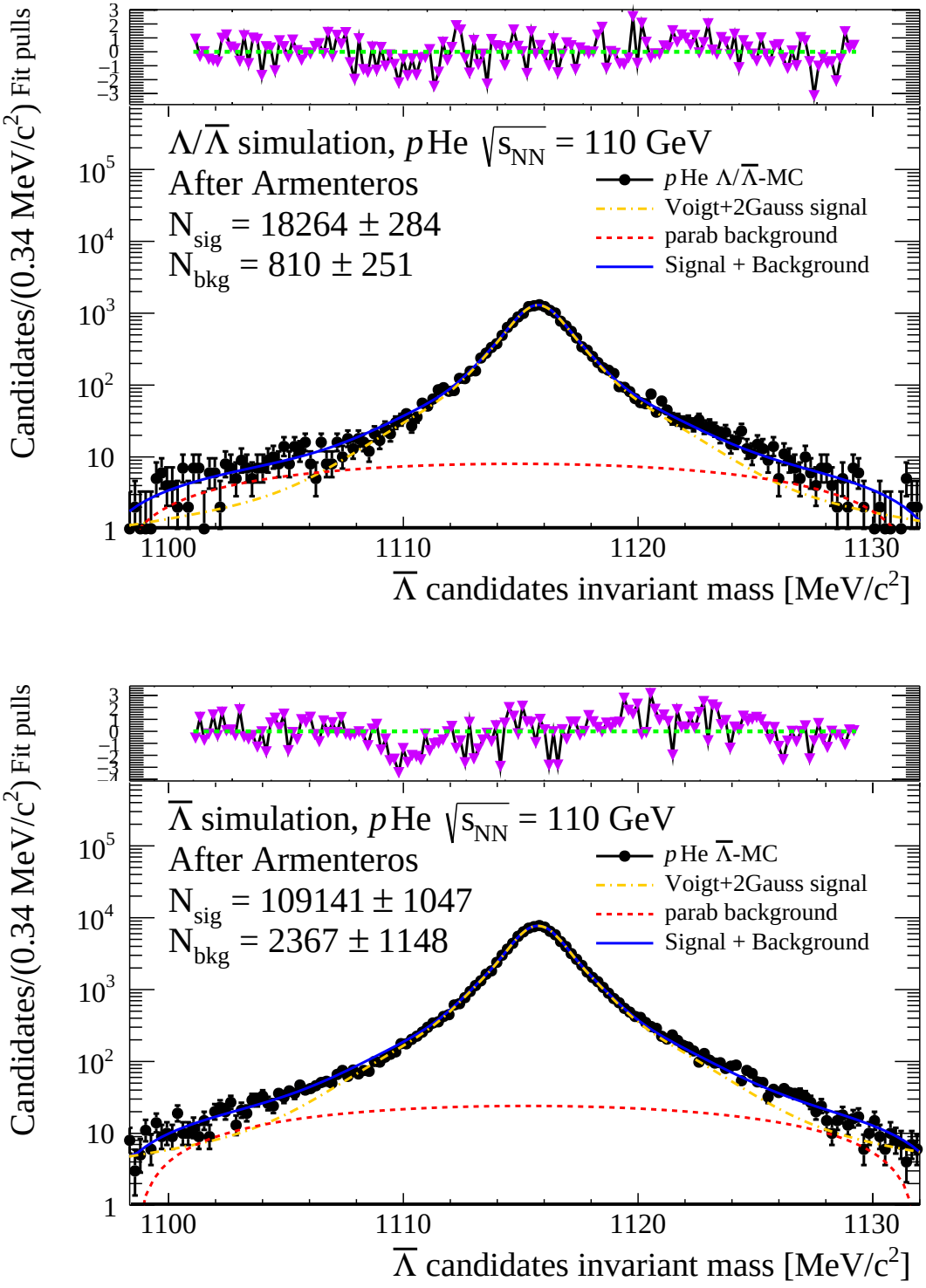


Figure II.2.14: Fit to the distribution of the $\bar{\Lambda}$ candidates invariant mass for the $\Lambda/\bar{\Lambda}$ (top) and the $\bar{\Lambda}$-enriched (bottom) simulated sample after the Armenteros-Podolanski selection with a Voigtian and two Gaussian functions for the signal and a second-order polynomial function for the background. The fit pulls (in purple) and the resulting number of signal and background events are shown.

The result consistency with the expectations proves the absence of biases in the total experimental efficiencies between the two samples.
The fit results for the other parameters in the $\overline{\Lambda}$-enriched simulated sample are used as an initialization of the fit to the data sample. A maximum $\pm 10\%$ variation of all parameters is allowed in order to account for possible imprecisions in the simulation while conserving the tested shapes. Figure II.2.15 shows the fit results on the pHe data sample before (top) and after (bottom) the Armenteros-Podolanski selection. By considering the obtained number of signal events in the two cases and applying Eq. II.2.6:

$$\varepsilon_{\text{Arm}} = \frac{50720 \pm 352}{51249 \pm 329} = (98.97 \pm 0.9)\%, \tag{II.2.9}$$

in agreement with the value calculated on the simulation and indicated in Eq. II.2.5. The choice of the background functional form is also validated by repeating the fit on the distribution before the Armenteros-Podolanski cut application with the physics-inspired Argus distribution. In this case, giving the low available background statistics in the simulated samples and the consequent large uncertainty affecting the fitted parameters, an independent initialization is applied. The result is presented in Fig. II.2.16 and a consistent number of signal events with respect to the polynomial model is obtained. The simplifying assumption for the choice of the background description is thus proved to negligibly contribute to the uncertainty.

In order to determine the cross-section dependence on the antiproton kinematics, a bidimensional binning scheme for its momentum and transverse momentum is considered with limits (in GeV/c):

p: [12, 14, 16.2, 18.7, 21.4, 24.4, 27.7, 31.4, 35.5, 40, 45, 50.5, 56.7, 63.5, 79.3, 110]
p_{T}: [0.4, 0.55, 0.7, 0.8, 0.9, 1.05, 1.5, 2.0, 4].

The parametrization of the $\overline{\Lambda}$ invariant mass distribution is repeated in each bin. All the model parameters are initialised with the results obtained on the sample integrated over the full kinematic phase-space and then left free to vary. As a consequence of the lower number of events in the kinematic bins and hence the larger statistical uncertainties, the fit regularity is monitored by requiring that:

- the Gaussian and Voigtian distributions widths are larger than $0.1\,\text{MeV}/c^2$, this removing spikes in the shapes where the fit can converge;
- the distance between the intersections of the second-order polynomial function with the x-axis is larger than $5\,\text{MeV}/c^2$;
- the fitted number of signal and background events are affected by a maximum 60% and 300% uncertainty, respectively.

In case one of these conditions is not met, the signal fit model is simplified by firstly removing the Breit-Wigner component in the Voigtian and then iteratively one Gaussian function. At last, a fit with one single Gaussian, which is a sufficiently smooth profile, is adopted. For the simulated samples, in case also this fit fails, which only happens in bins with extremely poor statistics, the number of signal events is determined as the candidates with the simulated identities matching the signal decay. Figure II.2.17

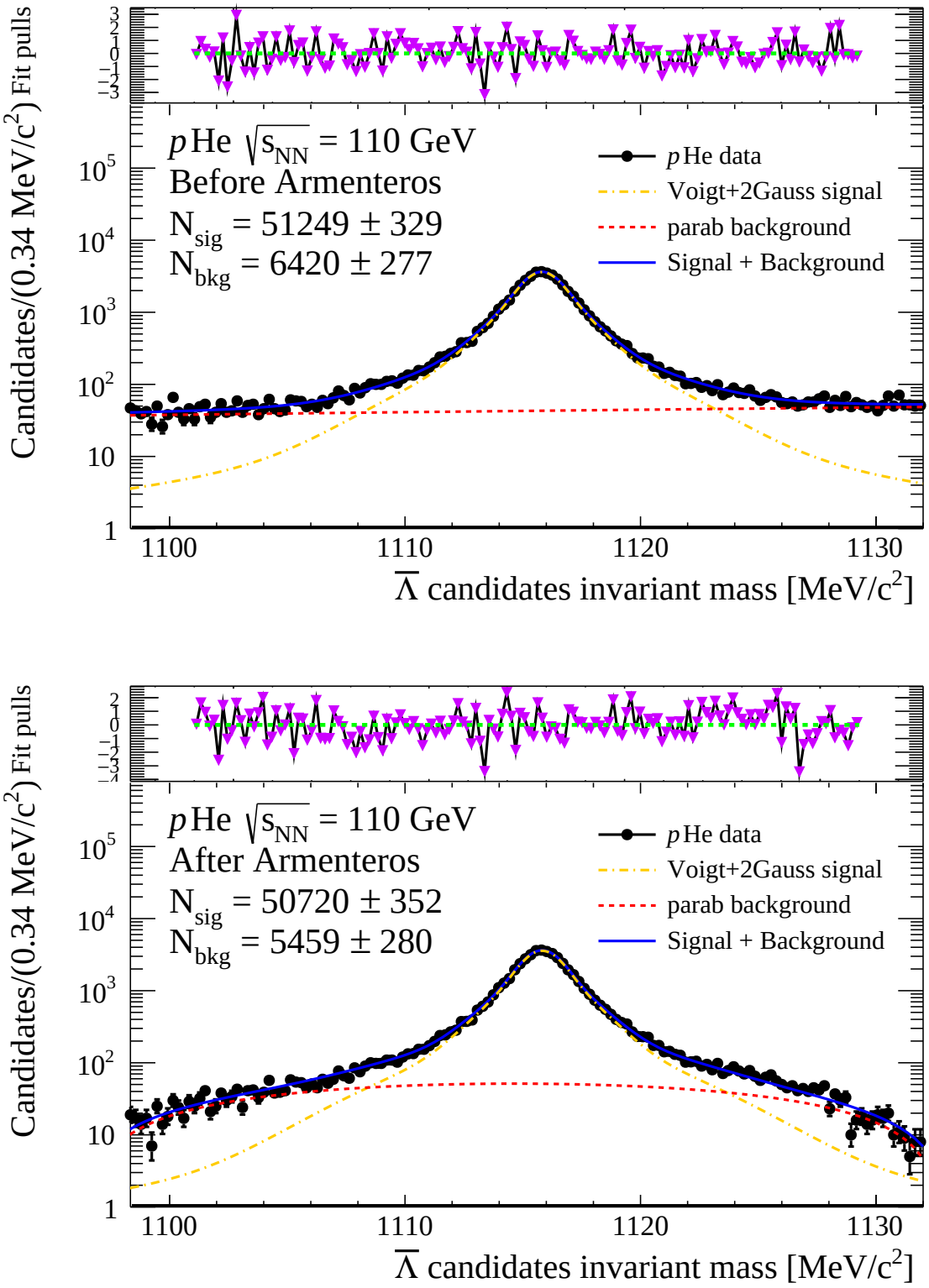


Figure II.2.15: Fit to the distribution of the $\overline{\Lambda}$ candidates invariant mass in data before (top) and after (bottom) the Armenteros-Podolanski selection with a Voigtian and two Gaussian functions for the signal and a second-order polynomial function for the background. The fit pulls (in purple) and the resulting number of signal and background events are shown.

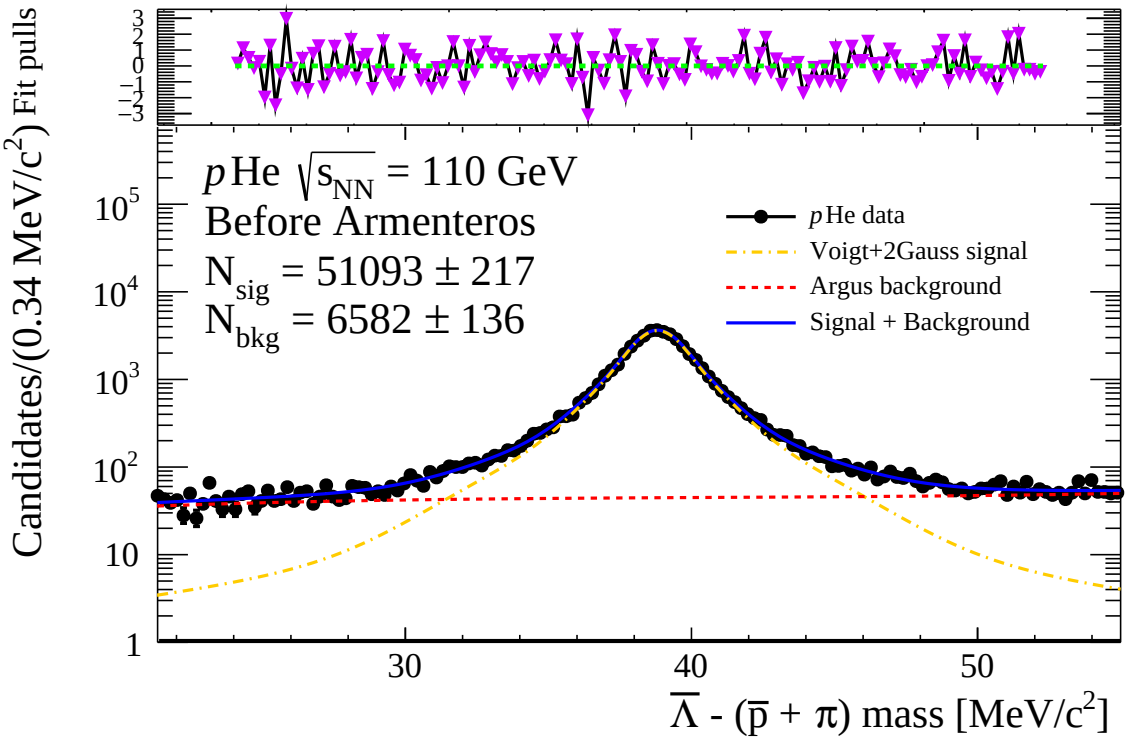


Figure II.2.16: Fit to the difference between the $\bar{\Lambda}$ invariant mass and the proton and pion nominal masses in the pHe data sample before the Armenteros-Podolanski selection application. The employed model is the sum of a Voigtian and two Gaussian distributions for the signal and an Argus function for the background. The number of signal events resulting from the fit is consistent with that obtained with a polynomial background, indicating the background simplified model to negligibly contribute to the uncertainty.

shows some examples of the fits in randomly-chosen kinematic bins. The results for the fitted number of signal events in the considered kinematic phase-space are presented in Fig. II.2.18, II.2.19 and II.2.20 for the $\Lambda/\bar{\Lambda}$ - and $\bar{\Lambda}$-enriched simulations and for data, respectively. For the projections of the bidimensional results onto the momentum (left) or transverse momentum (right) axes, the fitted number of events is compared with all selected candidates and, for the simulated samples, with the matched signal ones. In all cases, the fit results evolution with the antiproton kinematic is similar and, for the simulated samples, the candidates with failed particle-track matching can be seen as the difference between the red and green curves.

II.2.5 Experimental efficiency

All terms contributing to the total experimental efficiency ε in Eq. II.2.1 are measured as a function of the antiproton momentum and transverse momentum employing the simulated samples. The efficiency is defined as the ratio between the number of reconstructed and

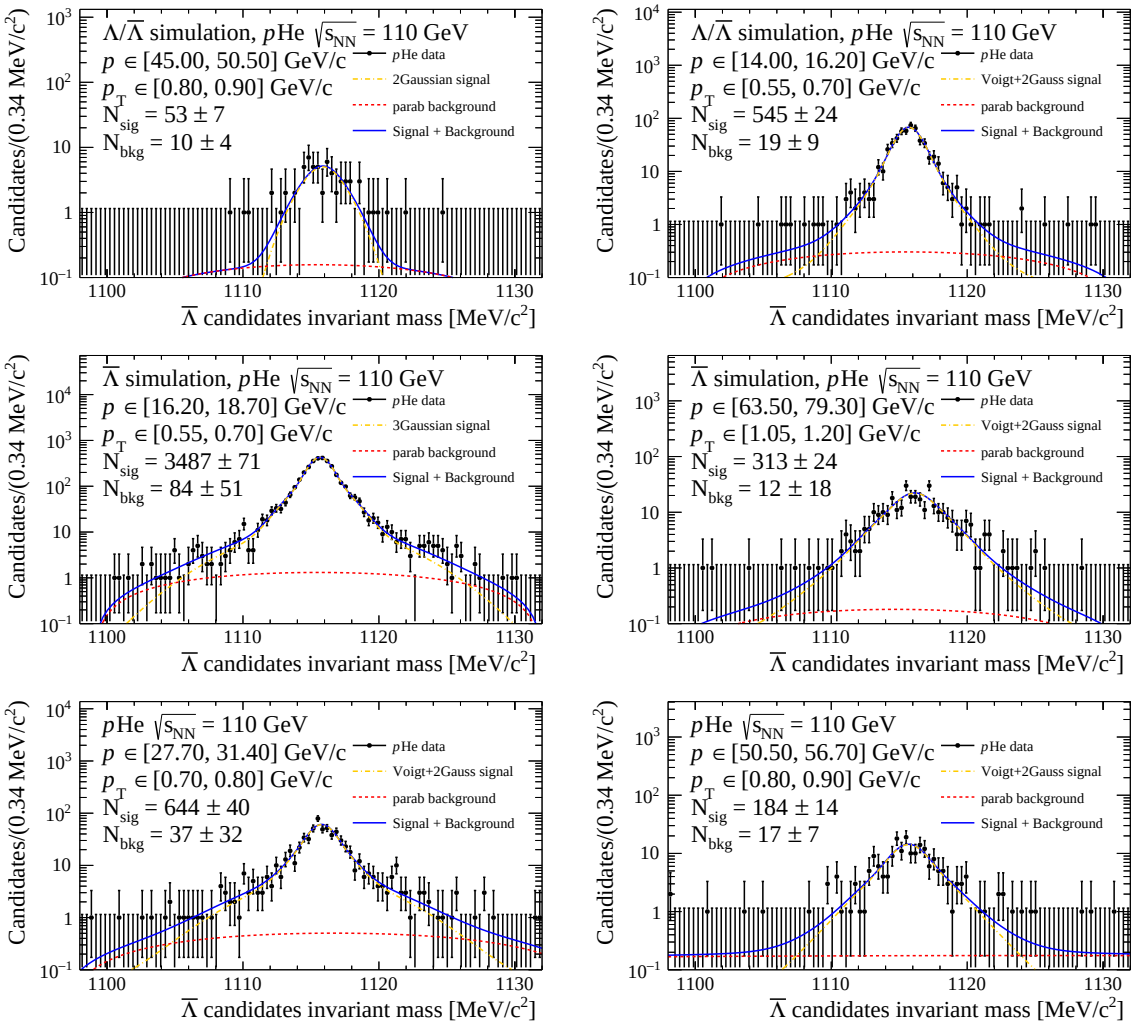


Figure II.2.17: Examples of the fits to the $\overline{\Lambda}$ candidates invariant mass distribution in the indicated intervals of the antiproton candidate momentum and transverse momentum in the $\Lambda/\overline{\Lambda}$ (top), in the $\overline{\Lambda}$-enriched (middle) simulated samples and in data (bottom). The default considered model is the sum of a Voigtian and two Gaussian for the signal and a second-order polynomial function for the background. In case, because of the poor statistics, the fit procedure fails according to the conditions detailed in the text, the model is simplified by removing firstly the Breit-Wigner component in the Voigtian and then one Gaussian per-iteration.

selected $\overline{\Lambda}_{\text{prompt}} \to \overline{p}\pi^+$ candidates and those produced in the pHe collisions

$$\varepsilon = \frac{N_{\text{rec \& sel}}}{N_{\text{produced in } p\text{He}}} \tag{II.2.10}$$

within the considered antiproton phase-space:

$$\begin{gathered} p \in [12, 110]\ \text{GeV}/c, \quad p_{\text{T}} \in [0.4, 4.0]\ \text{GeV}/c, \quad \eta < 5 \\ \text{PV}_z \in [-700, 100]\,\text{mm}. \end{gathered} \tag{II.2.11}$$

All quantities are considered after the event reconstruction for $N^{\text{rec \& sel}}$ and in the simulation for $N^{\text{produced in } p\text{He}}$, so that the efficiency measurement also takes into account

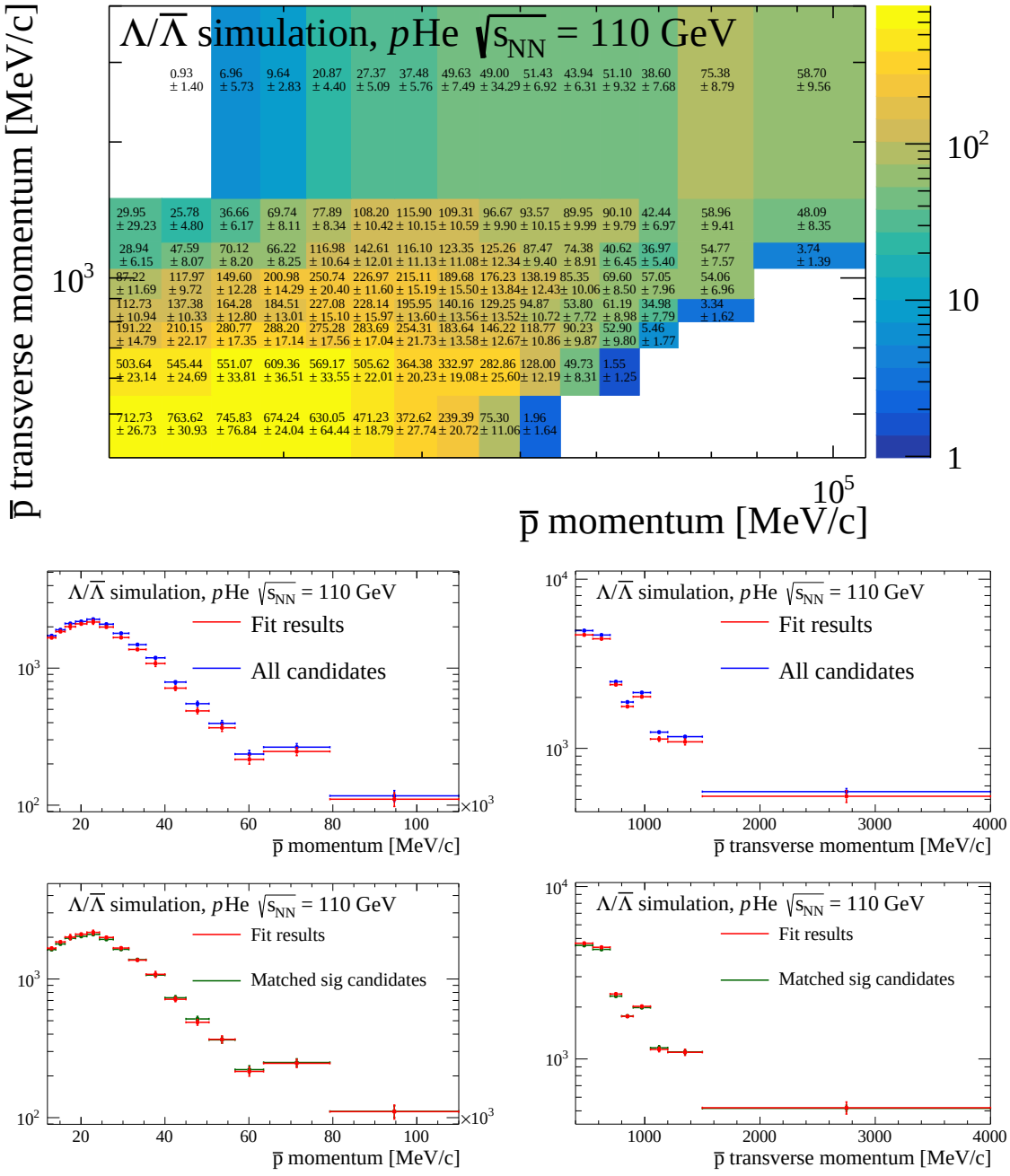


Figure II.2.18: Fitted number of signal events in the considered bins for the antiproton momentum and transverse momentum (top) and projections onto the momentum (left) and transverse momentum (right) axes for the $\Lambda/\bar{\Lambda}$ -enriched simulated sample. For the one-dimensional plots, the fit results (in red) are compared with all selected candidates (in blue) and with signal matched ones (in green).

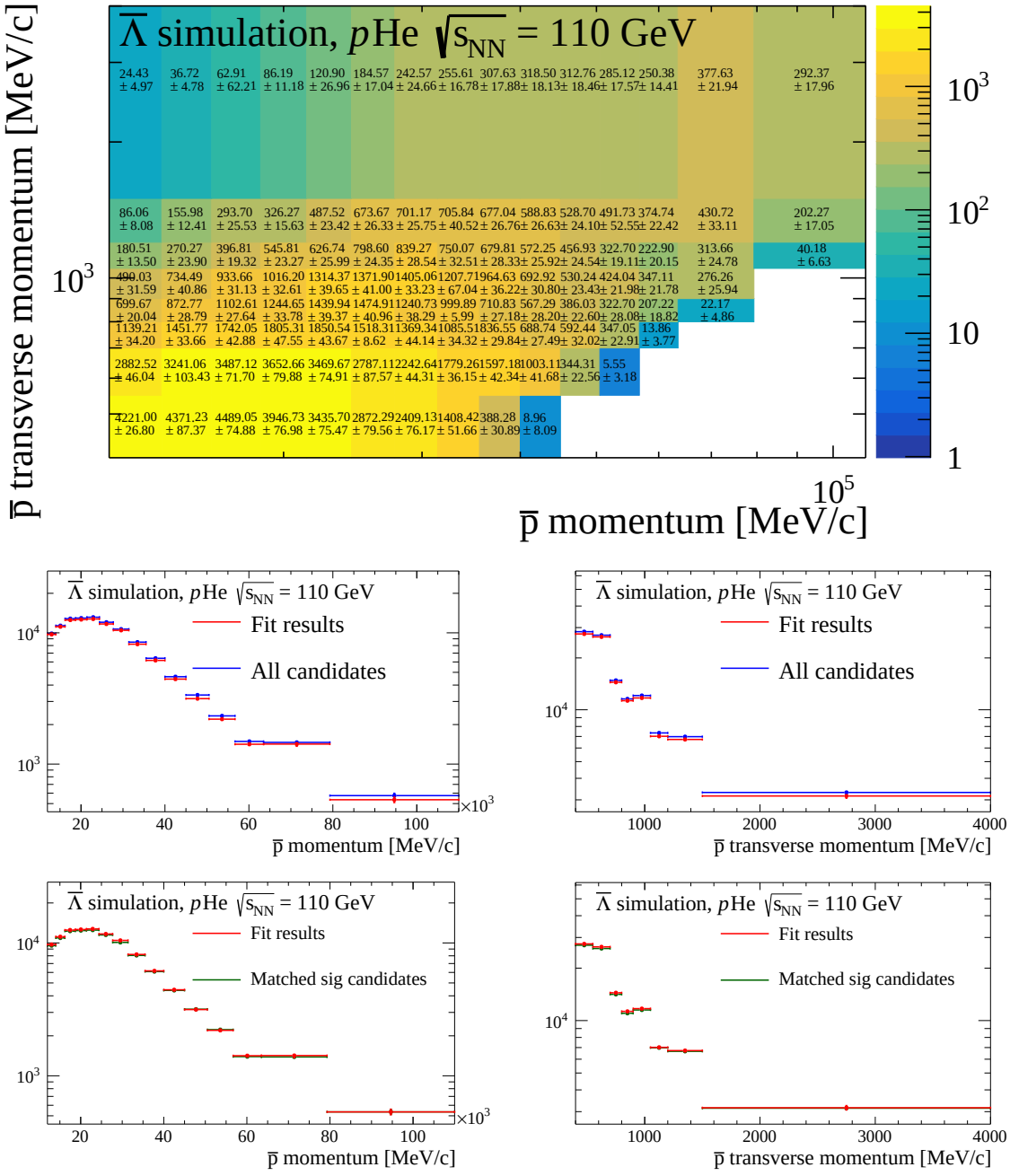


Figure II.2.19: Fitted number of signal events in the considered bins for the antiproton momentum and transverse momentum (top) and projections onto the momentum (left) and transverse momentum (right) axes for the $\overline{\Lambda}$-enriched simulated sample. For the one-dimensional plots, the fit results (in red) are compared with all selected candidates (in blue) and with signal matched ones (in green).

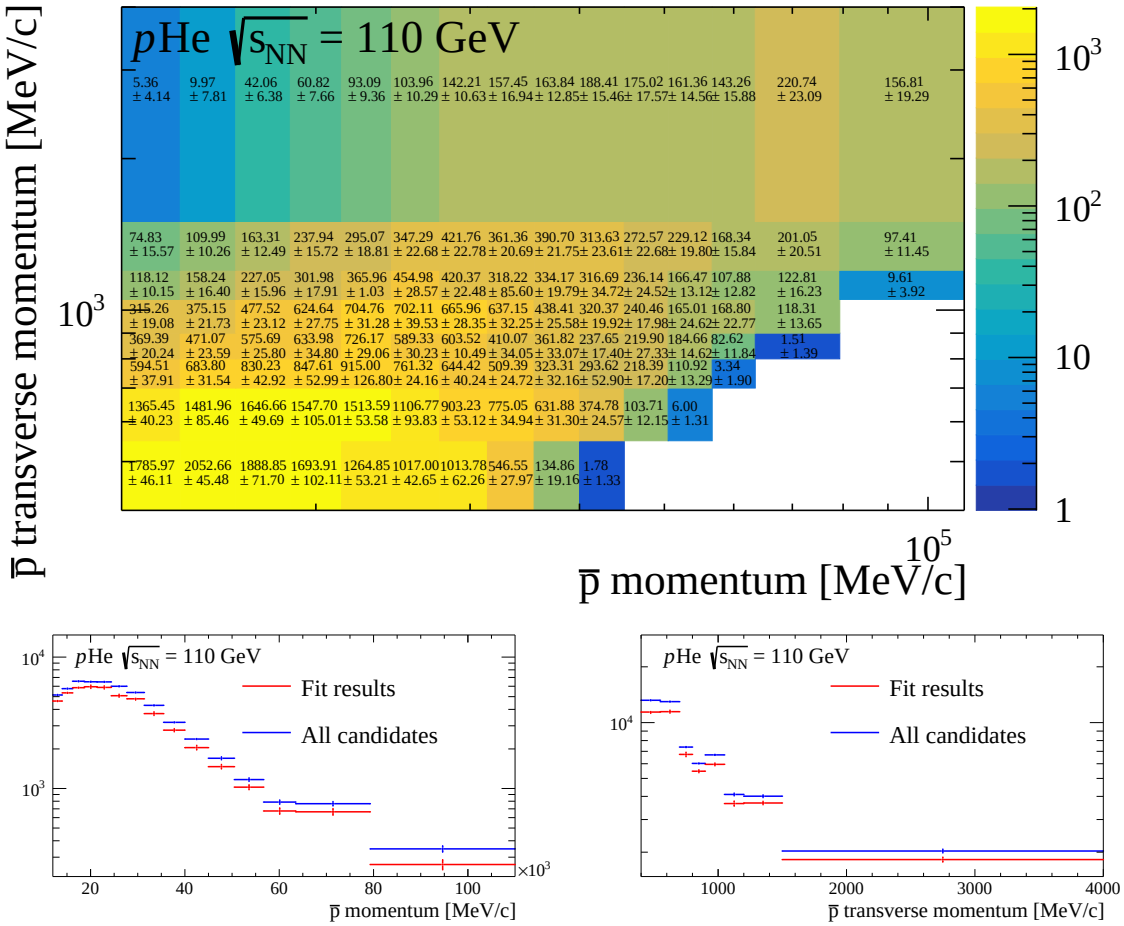


Figure II.2.20: Fitted number of signal events in the considered bins for the antiproton momentum and transverse momentum (top) and projections onto the momentum (left) and transverse momentum (right) axes for the data sample. For the one-dimensional plots, the fit results (in red) are compared with all selected candidates (in blue).

the resolution effects and the consequent candidate migrations between neighbour kinematic bins. Other than the phase-space requirements, the efficiency is only measured for promptly-produced $\overline{\Lambda}$ particles.

In the following, three terms of the total efficiency are separately studied as the generation ε_{gen}, simulation ε_{sim} and reconstruction and selection $\varepsilon_{\text{rec \& sel}}$ contributions in Sec. II.2.5.1, II.2.5.2 and II.2.5.3, respectively. These are defined as:

$$\varepsilon = \varepsilon_{\text{gen}} \cdot \varepsilon_{\text{sim}} \cdot \varepsilon_{\text{rec \& sel}} = \frac{N_{\text{gen}}}{N_{\text{produced in } p\text{He}}} \cdot \frac{N_{\text{sim}}}{N_{\text{gen}}} \cdot \frac{N_{\text{rec \& sel}}}{N_{\text{sim}}}, \tag{II.2.12}$$

being:

- $N^{\text{rec \& sel}}$ the number of reconstructed and selected candidates in the simulation, obtained with the same fit procedure counting the number of signal events in data;
- N^{sim} the number of simulated candidates;

- N^{gen} the number of generated $\overline{\Lambda}_{\text{prompt}} \to \overline{p}\pi^+$ decays in EPOS-LHC, including those not producing a corresponding candidate in the simulation because of an interaction with the detector material or because the parent-particle leaves the detector volume before the decay;
- $N^{\text{produced in } p\text{He}}$ the number of decays produced in the pHe collisions, including those that are removed from the selection applied to filter minimum-bias events.

II.2.5.1 Generation efficiency

The contribution to the efficiency due to the generator level cuts that are applied, as described in Sec. II.2.2, to produce the $\Lambda/\overline{\Lambda}$ and $\overline{\Lambda}$ simulated samples is studied in this section. Notably, all events where the pion, presenting a lower momentum than the antiprotons, exits the detector acceptance region are removed from the generated sample. In order to measure this effect, the samples with the disabled propagation of the generated particles across the detector are used. The same filtering configuration as the $\Lambda/\overline{\Lambda}$ and the $\overline{\Lambda}$-enriched simulated samples is considered, while, to speed-up the minimum-bias production, only events presenting a $\overline{\Lambda} \to \overline{p}\pi^+$ decay in the event are accepted, but not requirements on its characteristics are set. The corresponding fraction of events passing the generator-level selection is $\varepsilon^{\text{MB-MC}}_{\text{generator}} = (5.2207 \pm 0.0011)\%$, to be compared with those for the enriched samples indicated in Tab. II.2.2.

The two enriched samples are separately required to be statistically equivalent with the minimum-bias one, *i.e.* to satisfy:

$$\frac{N^{\Lambda/\overline{\Lambda}\,(\overline{\Lambda})}_{\text{evts}}}{\varepsilon^{\Lambda/\overline{\Lambda}\,(\overline{\Lambda})}_{\text{generator}}} = \frac{N^{\text{MB-MC}}_{\text{evts}}}{\varepsilon^{\text{MB-MC}}_{\text{generator}}}. \tag{II.2.13}$$

The generation efficiency is then measured as the ratio of the number of candidates in the two samples within the simulated phase-space of Eq. II.2.11:

$$\varepsilon_{\text{gen}}(p, p_{\text{T}})^{\Lambda/\overline{\Lambda}\,(\overline{\Lambda})} = \frac{N^{\Lambda/\overline{\Lambda}\,(\overline{\Lambda})}_{\text{gen}}}{N^{\text{MB-MC}}_{\text{gen}}}. \tag{II.2.14}$$

The results for the $\Lambda/\overline{\Lambda}$ and the $\overline{\Lambda}$-enriched simulated samples are presented in bins of the antiproton kinematics in Fig. II.2.21 and directly compared as a function of the momentum (top), the transverse momentum (middle) and the pseudorapidity (bottom) in Fig. II.2.22. On average, a lower efficiency is observed for the $\overline{\Lambda}$-enriched sample, which is expected considering the tighter applied cuts. The different evolution with the momentum originates from the selection on the maximum $\overline{\Lambda}$ flight distance, being the two quantities largely correlated.

II.2.5.2 Simulation efficiency

Secondly, the contribution to the efficiency of the generated particles propagation across the detector material is studied. As also described in the previous chapter, $\overline{\Lambda}$ candidates are removed from the generated sample, when, before decaying:

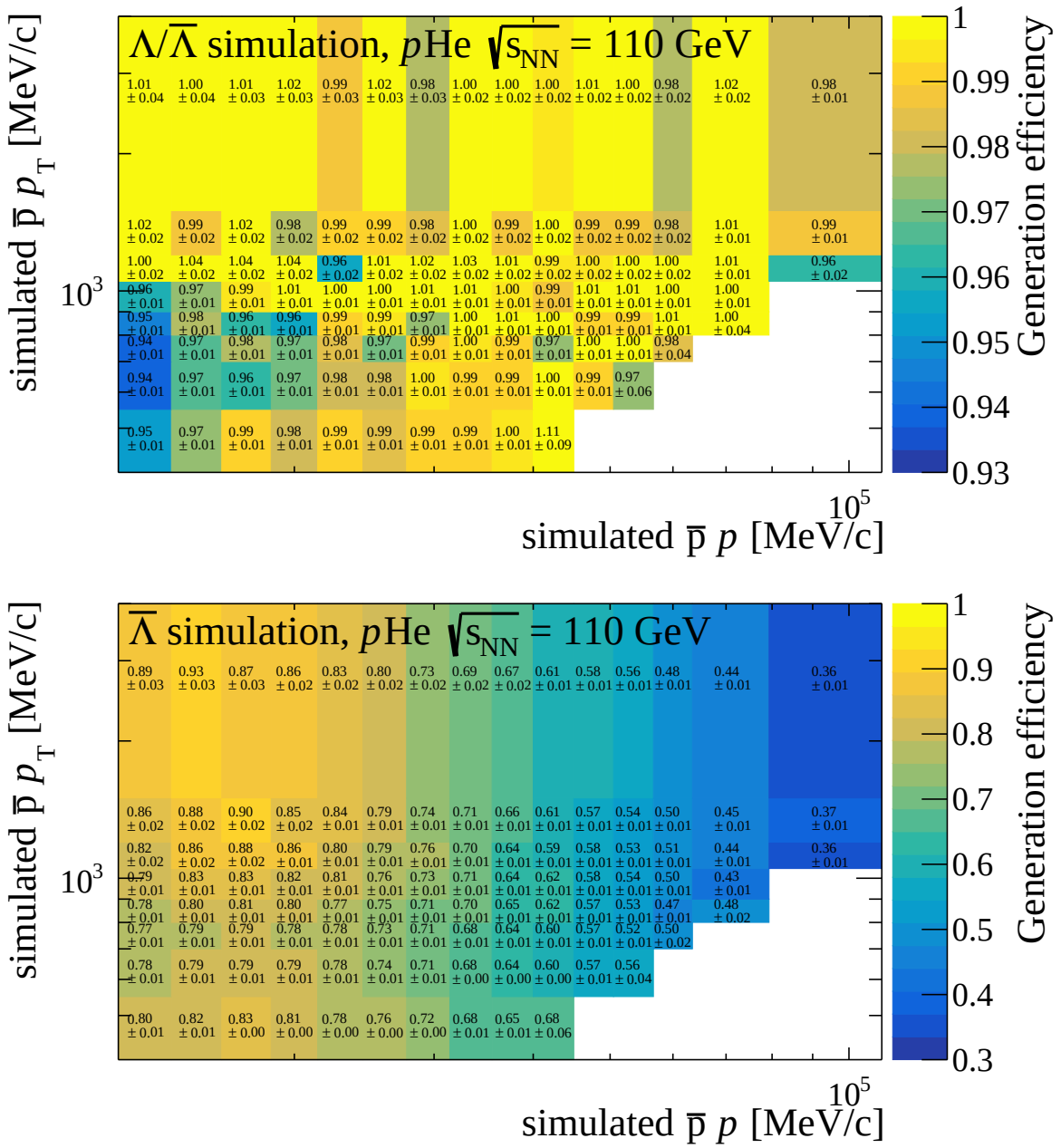


Figure II.2.21: Generation efficiency for the $\Lambda/\bar{\Lambda}$ - (top) and $\bar{\Lambda}$-enriched (bottom) simulated samples in the considered antiproton momentum-transverse momentum bins.

- a disruptive interaction with the detector material occurs;
- leave the detector volume.

The former effect is significant for both simulated samples, while, as a consequence of the threshold on the $\bar{\Lambda}$ maximum flight distance applied at the generator level in the $\bar{\Lambda}$-enriched simulated sample, the latter only affects the $\Lambda/\bar{\Lambda}$ -enriched one.
The simulation efficiencies are measured as:

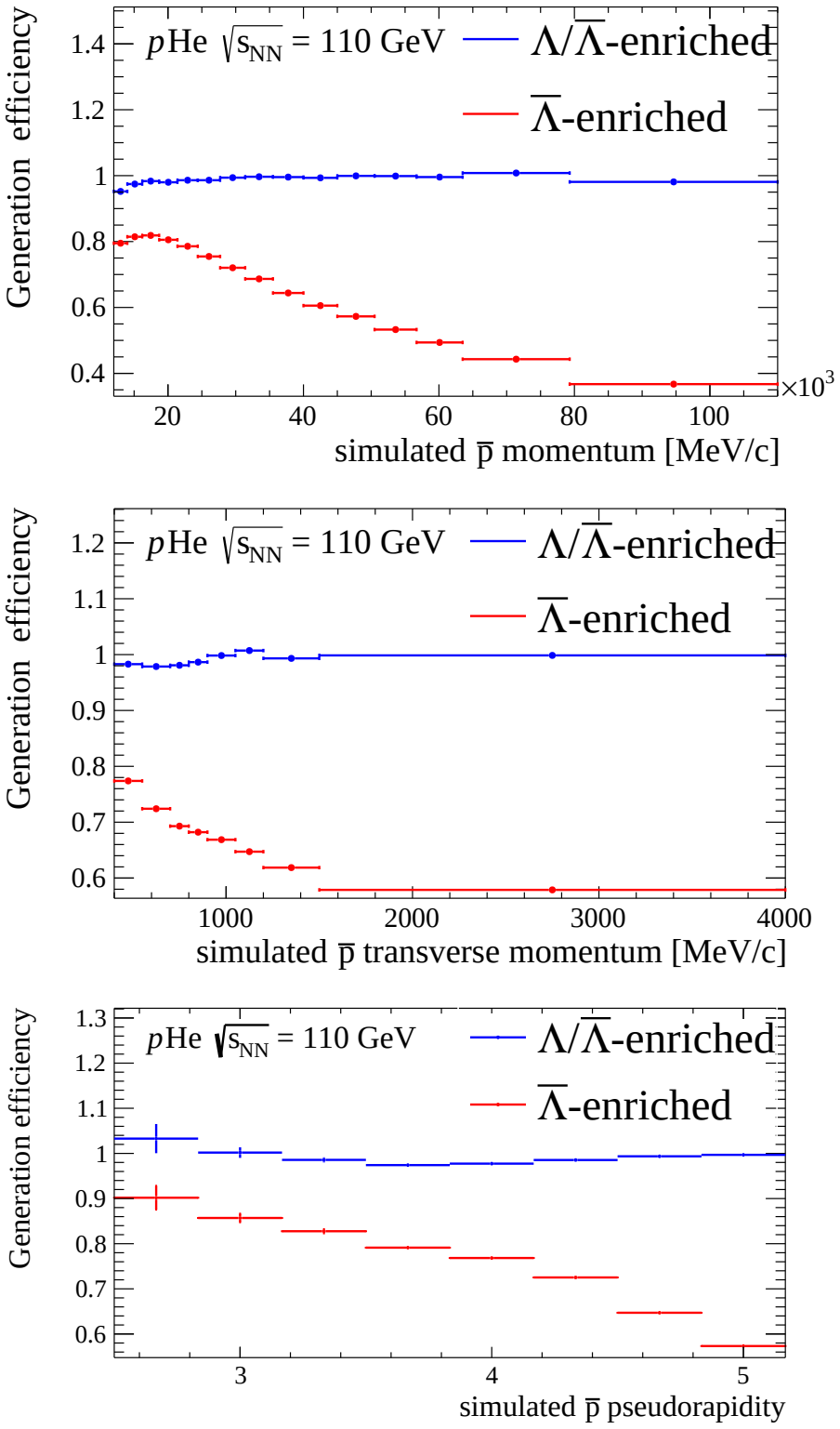


Figure II.2.22: Comparison between the measured generation efficiency for the $\Lambda/\overline{\Lambda}$ - (in blue) and the $\overline{\Lambda}$-enriched (in red) simulated samples as a function of the antiproton simulated momentum (top), transverse momentum (middle) and pseudorapidity (bottom).

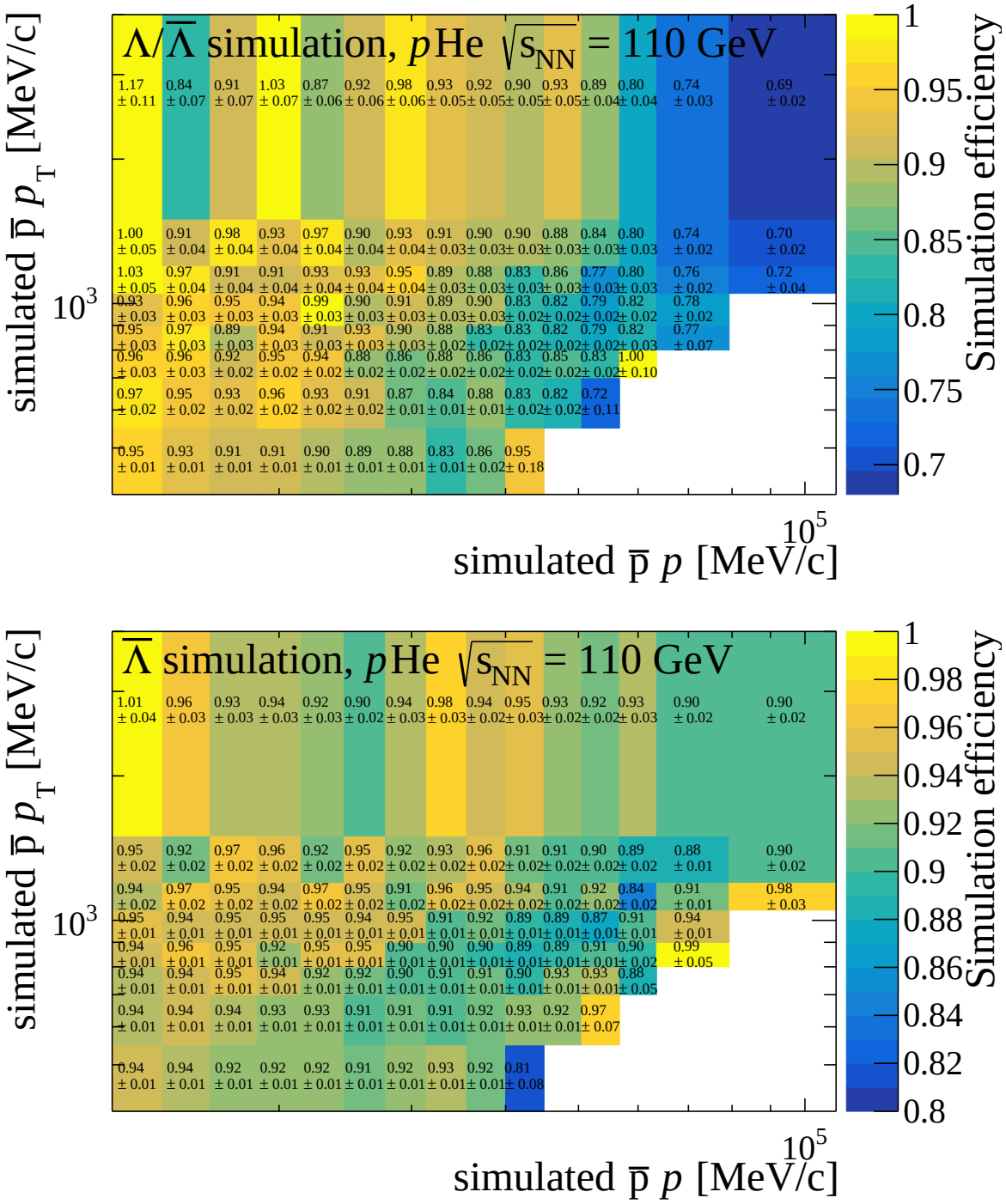


Figure II.2.23: Simulation efficiency for the $\Lambda/\bar{\Lambda}$ - (top) and $\bar{\Lambda}$-enriched (bottom) simulated samples in the considered antiproton momentum-transverse momentum bins.

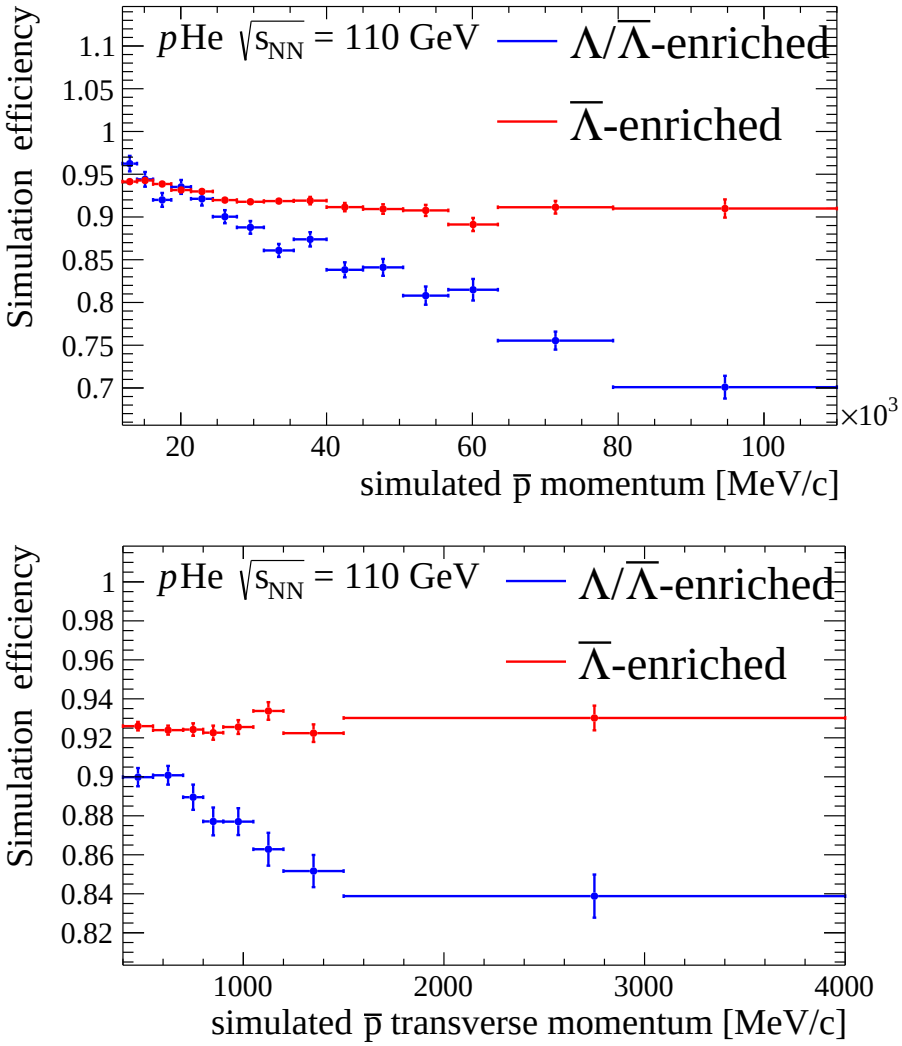


Figure II.2.24: Comparison between the measured simulation efficiency for the $\Lambda/\bar{\Lambda}$ - (in blue) and the $\bar{\Lambda}$-enriched (in red) simulated samples as a function of the antiproton simulated momentum (top) and transverse momentum (bottom).

$$\varepsilon_{\text{sim}}^{\Lambda/\bar{\Lambda}(\bar{\Lambda})}(p, p_{\text{T}}) = \frac{N_{\text{sim}}^{\Lambda/\bar{\Lambda}(\bar{\Lambda})}}{N_{\text{gen}}^{\Lambda/\bar{\Lambda}(\bar{\Lambda})}}. \tag{II.2.15}$$

As in the previous section, the generated-only and the simulated samples are required to be statistically equivalent and their candidates to belong to the fiducial phase-space defined in Eq. II.2.11. For the simulated samples, the requirement of a prompt $\bar{\Lambda}$ particle also removes most of the candidates originating in an interaction with the detector material that, being absent in the generated sample, would bias the ε_{sim} measurement. The obtained results are presented for the $\Lambda/\bar{\Lambda}$ and the $\bar{\Lambda}$-enriched simulated samples in the bins of the antiproton kinematics in Fig. II.2.23 and directly compared as a function of the simulated momentum (top) and transverse momentum (bottom) in Fig. II.2.24. As expected, the simulation efficiencies for the two samples are found to be similar at low momentum, where the particles decaying outside the detector region are a negligible fraction. At higher momenta, the $\bar{\Lambda}$-enriched efficiency, only accounting for the parent-particle interactions

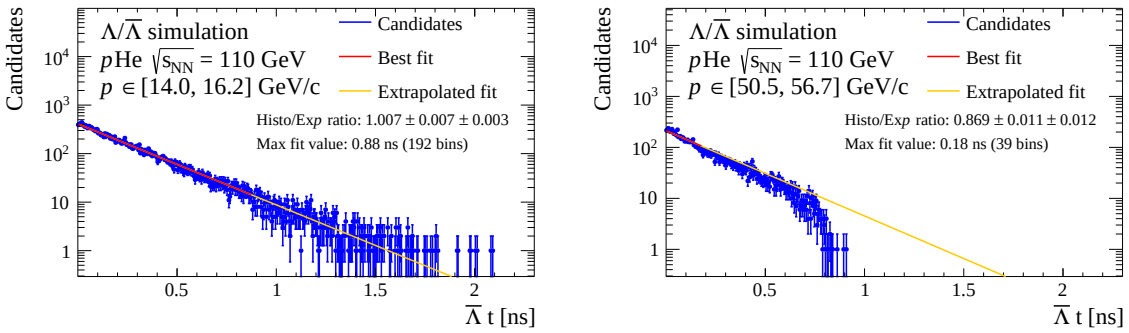


Figure II.2.25: $\overline{\Lambda}$ decay time distribution in the $\Lambda/\overline{\Lambda}$ -enriched simulated sample in an antiproton low (left) and high momentum (right) bin. Because of the candidates that are removed from the generated sample since decaying outside the detector volume, the expected exponential evolution is only found for the former.

with the detector, is almost flat, while it decreases for the $\Lambda/\overline{\Lambda}$ one as a consequence of the candidates leaving the LHCb detector region undecayed.
For the $\Lambda/\overline{\Lambda}$ -enriched simulated sample, the simulation efficiency is checked with the independent approach of Sec. II.1.5.1. As illustrated in Fig. II.2.25 in two far kinematic bins, a cut-off in the $\overline{\Lambda}$ decay-time distribution is observed at high momentum also for this simulated sample. In all considered momentum-transverse momentum intervals, the distribution is modelled as $y = a \cdot e^{-t/\tau}$, being $\tau = 0.26\,\text{ns}$ fixed to the nominal $\overline{\Lambda}$ particle lifetime. The histogram-to-exponential integral ratio, representing the fraction of candidates decaying inside the LHCb region, is measured (in green) and compared to the simulation efficiency (in blue) in Fig. II.2.26 as a function of the antiproton momentum (top) and transverse momentum (bottom). The decay-in-LHCb results, only considering one over the two terms contributing to the simulation efficiency, are always larger and follow the same evolution with the kinematics. Considering that the simulation efficiency for the $\overline{\Lambda}$-enriched simulated sample in Fig. II.2.24, where only the particle interactions with the detector contribute, is almost flat with the kinematics, this is also expected.

II.2.5.3 Reconstruction and selection efficiencies

The efficiency of the decay reconstruction in the detector and of the application of the selection algorithm is measured in this section. This is defined as:

$$\varepsilon^{\Lambda/\overline{\Lambda}(\overline{\Lambda})}_{\text{rec \& sel}}(p,\, p_{\text{T}}) = \frac{N^{\Lambda/\overline{\Lambda}(\overline{\Lambda})}_{\text{rec \& sel}}}{N^{\Lambda/\overline{\Lambda}(\overline{\Lambda})}_{\text{sim}}} \tag{II.2.16}$$

and results from the combination of:

- the pHe collision vertex reconstruction;
- the reconstruction of the antiproton and pion particles in all tracking detectors;
- the $\overline{\Lambda}$ decay vertex reconstruction;
- the efficiency of the trigger and offline selections.

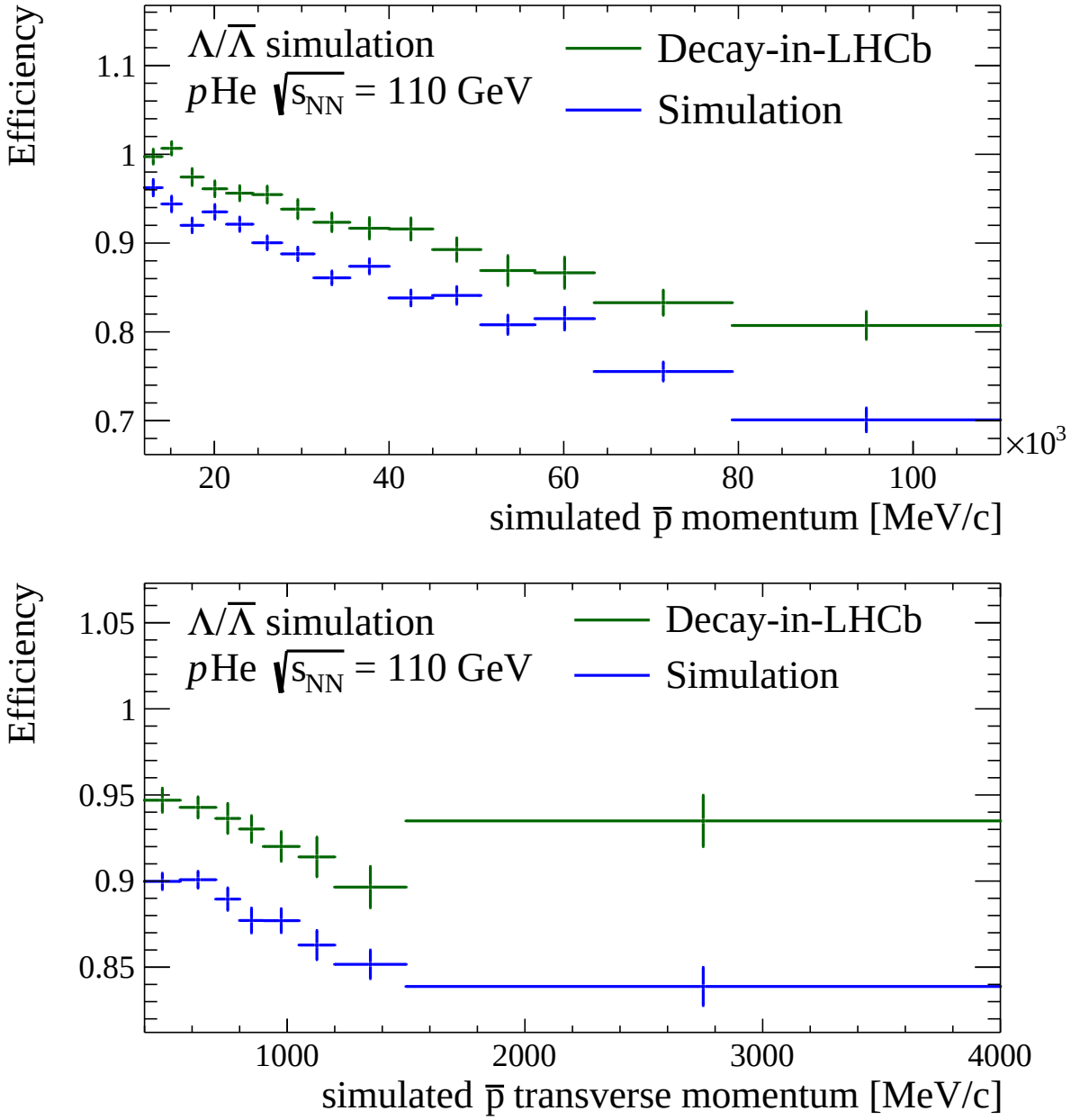


Figure II.2.26: Simulation (in blue) and decay-in-LHCb (in green) efficiencies for the $\Lambda/\bar{\Lambda}$ -enriched simulated sample as a function of the antiproton simulated momentum (top) and transverse momentum (bottom).

The overall efficiency is measured on the $\Lambda/\bar{\Lambda}$ and the $\bar{\Lambda}$-enriched simulated samples. No weighting of the simulated distribution is required, considering that, as observed in Fig. II.2.10, the simulated particle kinematics and the event geometry that most determine the reconstruction efficiency precisely reproduce the data. Small residual discrepancies for the measurement of the efficiency in data and simulation are studied by the LHCb tracking working group with a data-driven technique applied to $J/\psi \to \mu^+\mu^-$ decays reconstructed and selected in pp data [133]. Some weights, indicated in Fig. II.2.27 in regions of the particle momentum and pseudorapidity, are provided and applied to the

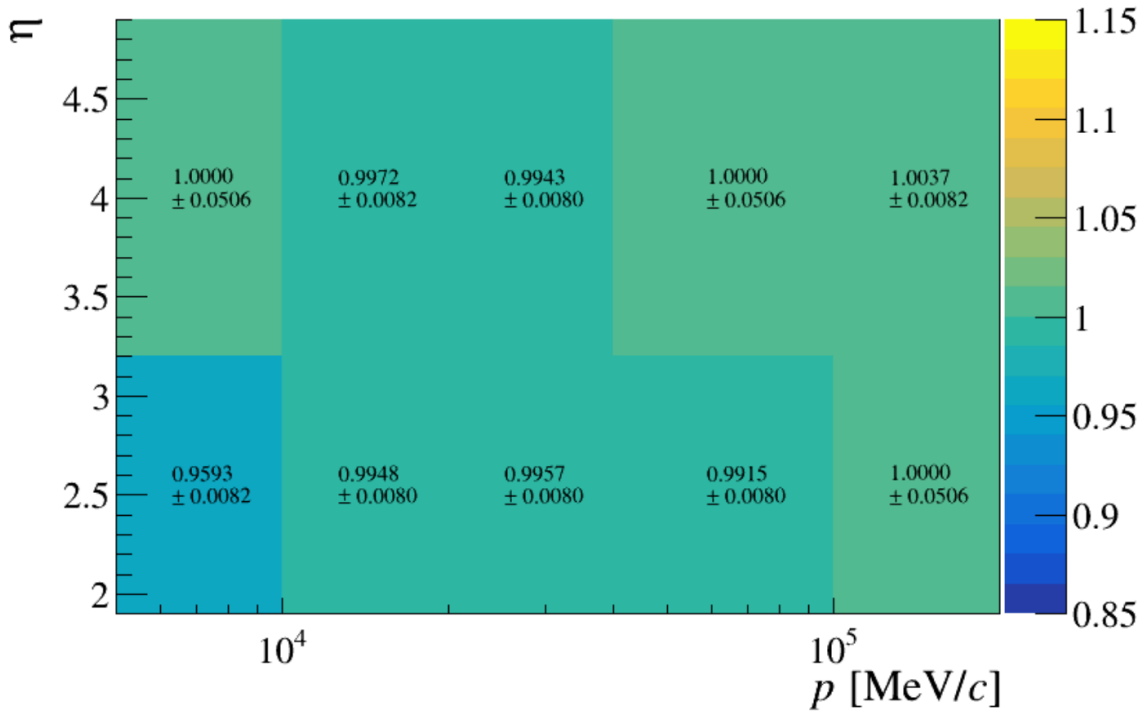


Figure II.2.27: Tracking correction factors as provided by the LHCb tracking group in regions of the particle momentum and pseudorapidity.

antiproton and pion candidates per-event. Since the adopted binning schemes differ, the bin limits are extended to cover the whole range considered for this analysis. Also, following the prescription of the tracking group, a uniform 0.8% relative systematic uncertainty is added. A possible additional contribution due to the differences in the distributions affecting the track reconstruction between the pp and fixed-target samples, notably the PV_z one, and the consequent possible bias in the provided tracking correction factors are discussed in Sec. II.2.6.1. After the application of the correction factors, the number of reconstructed and selected signal events is determined in the $\Lambda/\overline{\Lambda}$ and in the $\overline{\Lambda}$-enriched simulated sample with the same fit procedure applied to data in Sec. II.2.4, so that also signal unmatched candidates are accounted for. By dividing for the number of simulated candidates, the results for the $\Lambda/\overline{\Lambda}$ - (top) and the $\overline{\Lambda}$-enriched (bottom) samples are presented in Fig. II.2.28 in the considered kinematic bins and compared as a function of the antiproton momentum (top) and transverse momentum (bottom) in Fig. II.2.29. A similar efficiency evolution with the antiproton momentum and transverse momentum is found between the two samples. On absolute, larger values are observed for the $\overline{\Lambda}$-enriched simulated sample, which is expected considering the higher background rejection due to the tighter applied cuts at the generator level.

II.2.5.4 Total efficiency

The generation, simulation, reconstruction and selection efficiency terms studied in the previous sections are multiplied according to Eq. II.2.12 to determine the total experimental efficiency. The results are presented in Fig. II.2.30 for the $\Lambda/\overline{\Lambda}$ - (top) and the $\overline{\Lambda}$-enriched (bottom) simulated sample in the kinematic bins and then compared projected onto

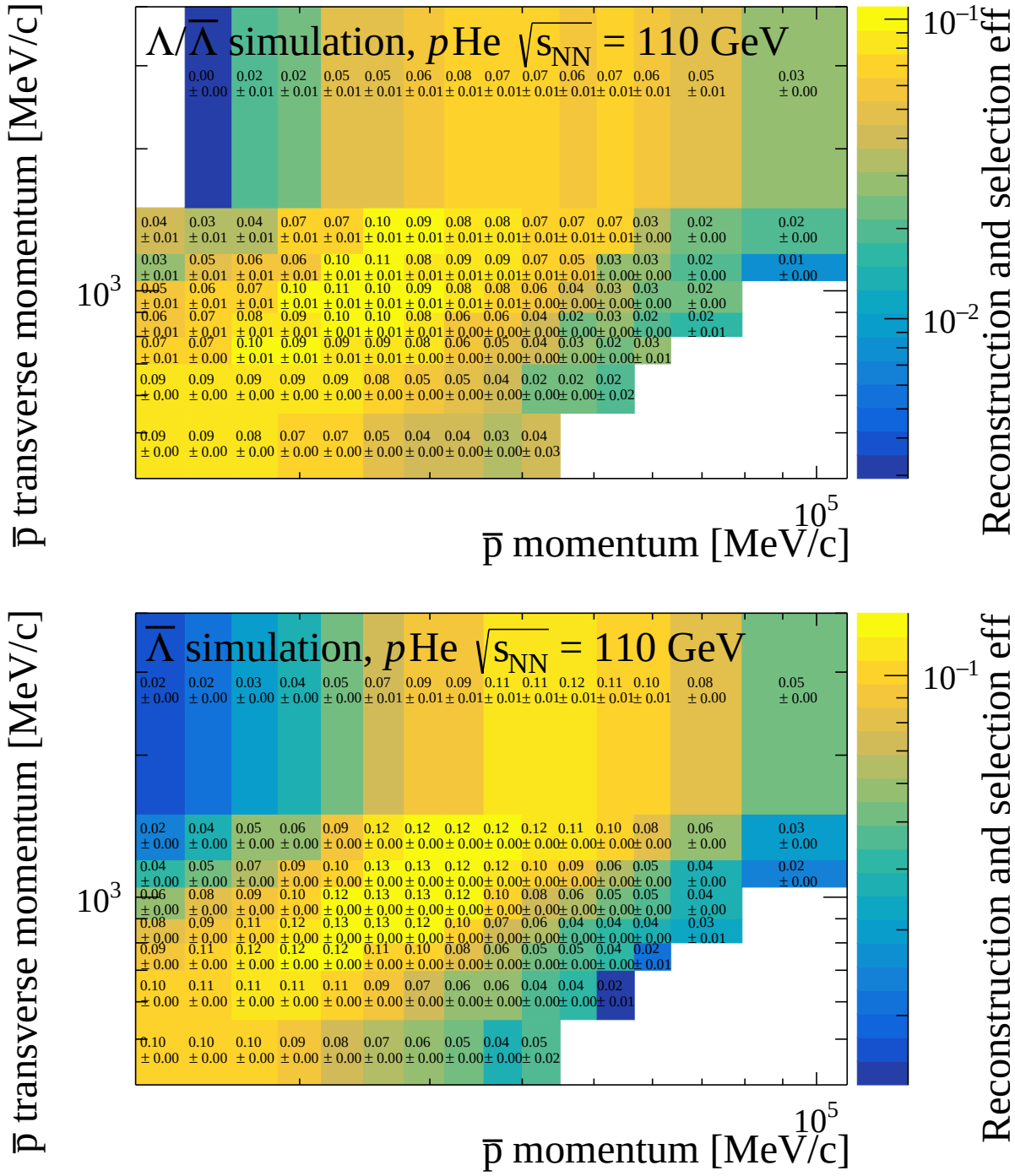


Figure II.2.28: Reconstruction and selection efficiencies for the $\Lambda/\bar{\Lambda}$ - (top) and $\bar{\Lambda}$-enriched (bottom) simulated samples in the considered antiproton momentum-transverse momentum bins.

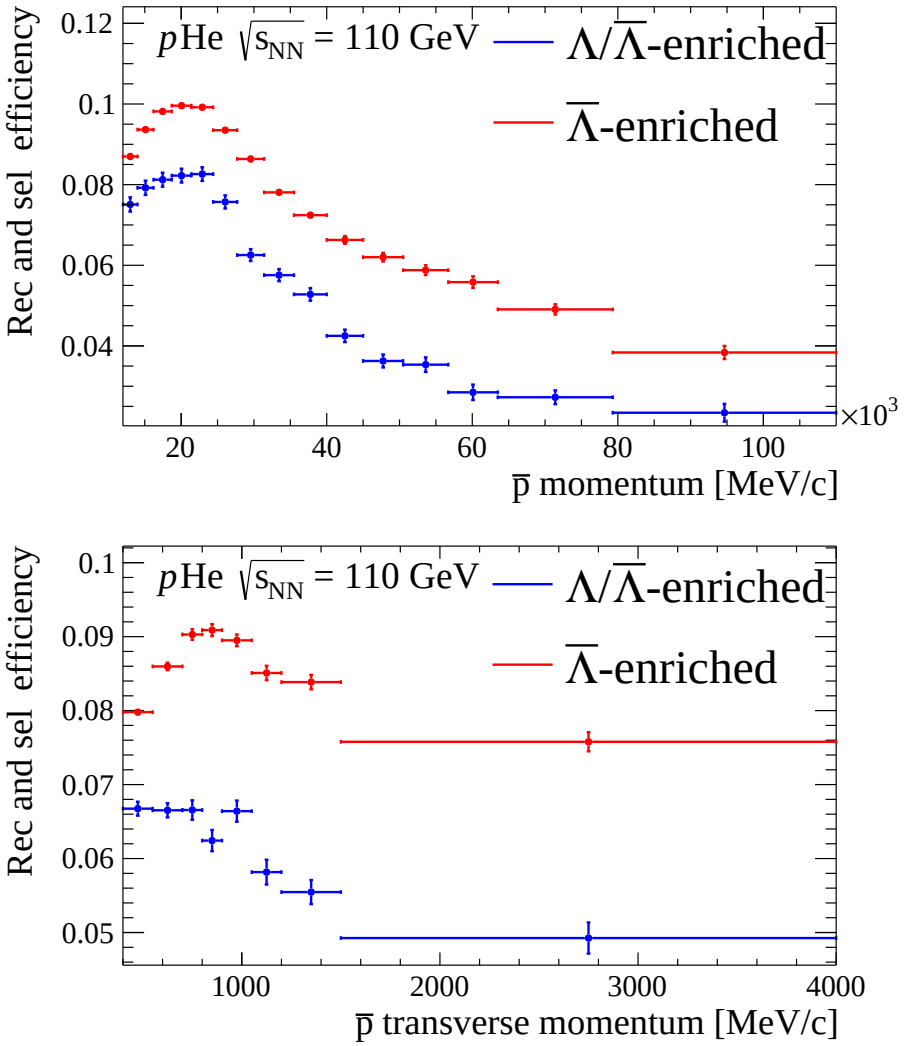


Figure II.2.29: Reconstruction and selection efficiencies for the $\Lambda/\overline{\Lambda}$ - (in blue) and the $\overline{\Lambda}$-enriched (in red) simulated samples as a function of the antiproton simulated momentum (top) and transverse momentum (bottom).

the two axes in Fig. II.2.31. Within the statistical uncertainties, a clear consistency is found between the two profiles. Being the production and selection of the two samples independent, this proves the robustness of the presented efficiency studies. The efficiency values for the $\overline{\Lambda}$-enriched sample, whose statistics is dominant, will be used to calculate the antiproton production cross-section in $\overline{\Lambda}_{\text{prompt}} \to \overline{p}\pi^+$ decays.

II.2.6 Systematic uncertainties

In the following sections, the contributions to the systematic uncertainty are studied:

- the reconstruction of fixed-target particles in the VELO detector, in Sec. II.2.6.1;
- the choice of the parametric function to model the $\overline{\Lambda}$ invariant mass, in Sec. II.2.6.2;
- the interactions of the daughter particles with the detector material, in Sec. II.2.6.3.

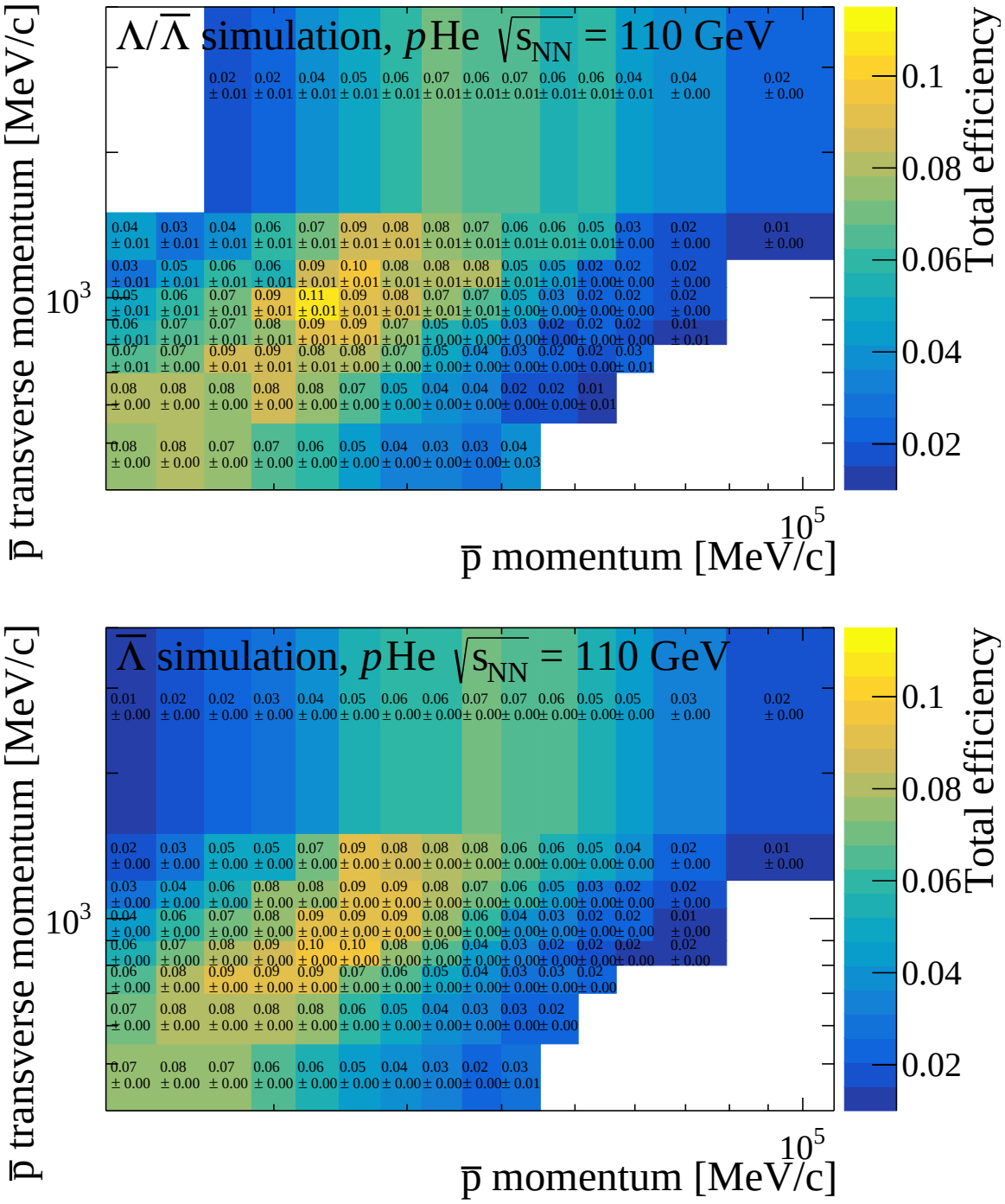


Figure II.2.30: Total experimental efficiency for the $\Lambda/\bar{\Lambda}$ - (top) and $\bar{\Lambda}$-enriched (bottom) simulated samples in the considered antiproton momentum-transverse momentum bins.

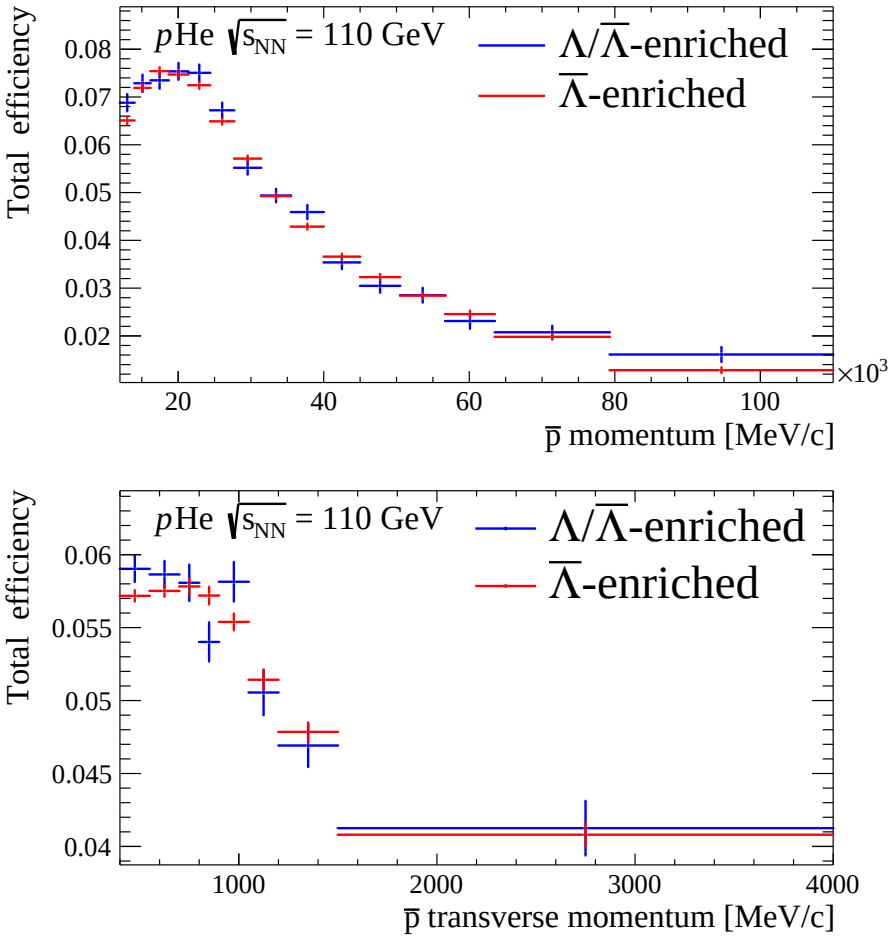


Figure II.2.31: Total efficiency for the $\Lambda/\overline{\Lambda}$ - (in blue) and the $\overline{\Lambda}$-enriched (in red) simulated samples as a function of the antiproton simulated momentum (top) and transverse momentum (bottom). The observed consistency between the two samples proves the robustness of the efficiency studies discussed in the text.

II.2.6.1 Data-driven VELO track reconstruction efficiency

As discussed in Sec. II.2.5.3, the reconstruction and selection efficiencies are measured employing the simulated samples and the correction factors provided by the LHCb tracking group. Since these are determined with a data-driven procedure applied to $J/\psi \to \mu^+\mu^-$ decays in pp data, the differences in the physics and data-taking conditions with the fixed-target samples could induce a systematic uncertainty. In particular, the track reconstruction in the VELO detector, which is expected to be the most dependent on the PV_z, is studied in this section. To investigate whether the track reconstruction efficiency measured in the simulated and data samples agree, the $\overline{\Lambda} \to \overline{p}\pi^+$ decays are reconstructed considering:

- both the antiproton and pion particles associated to tracks reconstructed collecting information from all tracking detectors, *i.e.* as long tracks (LL);
- the antiproton or the pion particle reconstructed as a long track and the other with a special implemented reconstruction algorithm where the VELO segments are

ignored and only downstream trackers are considered (LD).

In all configurations, the decay preselection is common to Sec. II.2.3. Offline, to enhance the selected statistics, looser detachment requirements are set as:

$$\mathcal{F}_{\text{IP}} > 1 \qquad \mathcal{F}_{\chi^2_{\text{IP}}} > 3. \tag{II.2.17}$$

The background contamination is instead reduced by tagging the antiproton or the pion particle with the PID selections:

$$\begin{aligned} \overline{p} &: \text{DLL}_{p,\pi} > 2, \qquad \text{DLL}_{p,K} > 2 \\ \pi &: \text{DLL}_{p,\pi} < 1, \qquad \text{DLL}_{K,\pi} < 1, \end{aligned} \tag{II.2.18}$$

respectively. For each tag particle, one sample has the other particle, the probe, reconstructed as a long track and another one ignoring the VELO information. Correspondences for the same probe in these two are looked for requiring:

- the same event number;
- the same transverse momentum of the tag particle, with a tolerance of 0.01 MeV/c to account for a different resolution in the LL and LD reconstruction;
- an angular distance between the probes in the LL and LD reconstructions such that $\sqrt{(\theta_{\text{LL}} - \theta_{\text{LD}})^2 + (\phi_{\text{LL}} - \phi_{\text{LD}})^2} < 0.8$.

If a tag particle is associated to more than one probe in the LL or LD configurations, the candidates with the lowest angular distance is considered. The tracking efficiency and the related uncertainty are then measured as:

$$\varepsilon_{VELO} = \frac{ll}{ld} \tag{II.2.19}$$

$$\sigma(\varepsilon_{VELO}) = \sqrt{\left|\frac{(1 - 2\varepsilon_{VELO}) \cdot \sigma(ll)^2 + \varepsilon^2_{VELO} \cdot \sigma(ld)^2}{ld^2}\right|} \tag{II.2.20}$$

being:

- $ll \pm \sigma(ll)$ the number of signal events reconstructed in the LL configuration and matched to an LD candidate;
- $ld \pm \sigma(ld)$ the number of signal events reconstructed in the LD configuration, both matched and unmatched.

Indeed, LD candidates where the probe is also reconstructed in the VELO detector will be included in the LL sample. On the contrary, VELO inefficiencies would give a candidate in the LD but not in the LL reconstruction. The uncertainty on the efficiency is obtained by assuming the number of reconstructed and non-reconstructed tracks as independent quantities and propagating the uncertainties. The $ll \pm \sigma(ll)$ and $ld \pm \sigma(ld)$ quantities are obtained by modelling the $\bar{\Lambda}$ candidates invariant mass distribution with a parametric function. The addition of three Gaussian distributions is assumed for the signal, an Argus function for the background. The fit is performed as a function of the probe particle

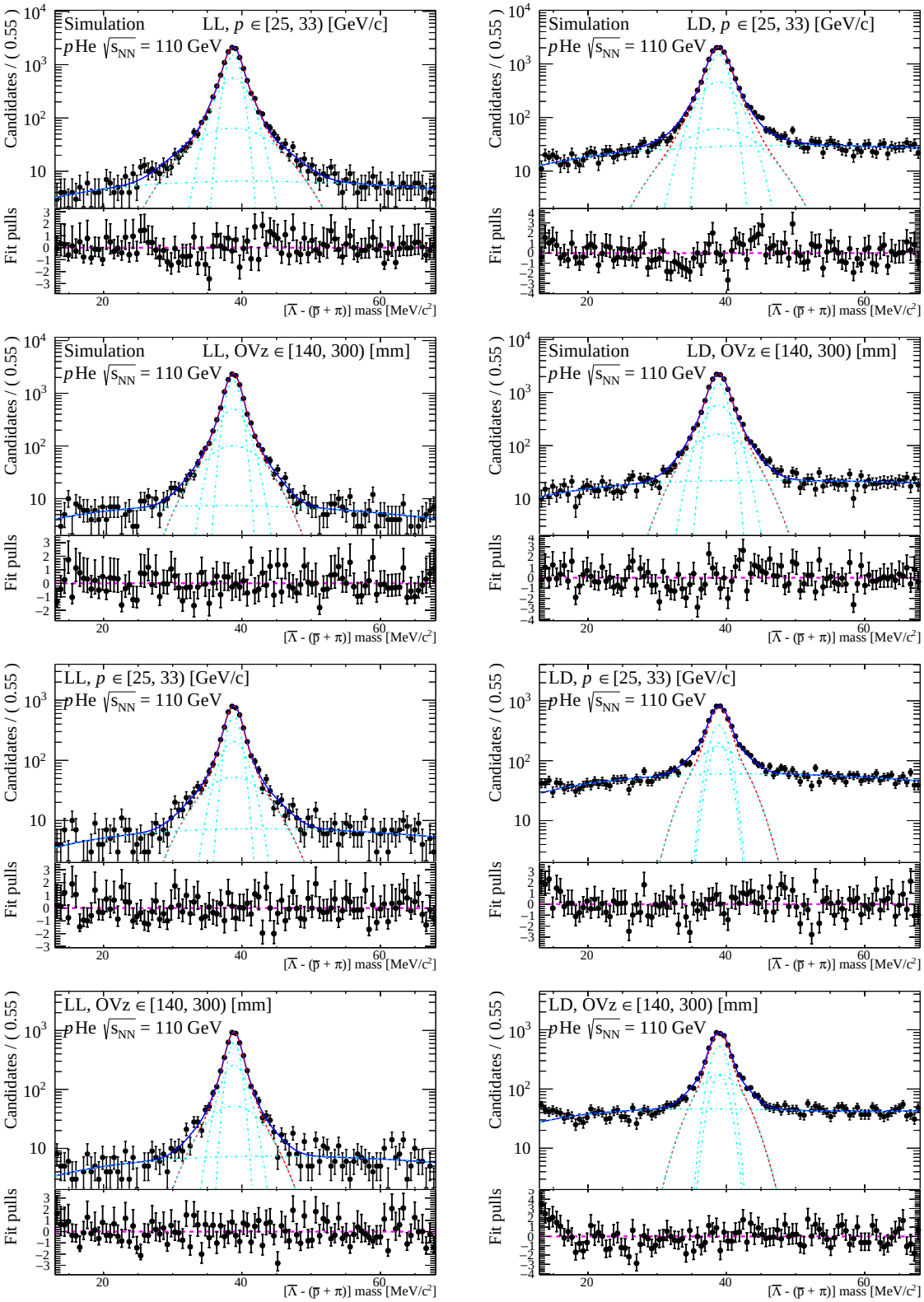


Figure II.2.32: Examples of the fit procedure applied to the $\bar{\Lambda}$ candidates invariant mass in the LL (left) and LD (right) decay reconstruction in the simulated (first and second row) and data (third and fourth row) sample in the same momentum (first and third row) and origin vertex z (second and fourth row) bin. The employed model is the sum of three Gaussian distributions for the signal and an Argus function for the background.

momentum, transverse momentum, OV_z and of the number of reconstructed long tracks in the event, measuring the detector occupancy. Considering the available statistics, only a large binning scheme, with limits determined requiring a similar population in each bin, is applied:

$$p : [12.0, 16.5, 20.5, 25, 33, 110]\,\mathrm{GeV}/c$$
$$p_{\mathrm{T}} : [0.4, 0.52, 0.64, 0.77, 1, 4]\,\mathrm{GeV}/c$$
$$\mathrm{OV}_z : [-700, -30, 140, 300, 450, 700]\,\mathrm{mm}$$
$$N_{\mathrm{long}} : [2, 7, 10, 13, 18, 72].$$

Examples of the fits for the LL (left) and LD (right) configurations considering the antiproton as probe particle and in randomly-chosen bins are presented in Fig. II.2.32. As expected, while the LL peak is well enhanced over the combinatorial background, the higher number of two-tracks combinations induces a larger background contamination in the LD sample. By iterating over all the considered bins and applying in each one Eq. II.2.19, Fig. II.2.33 shows the results for the VELO track reconstruction efficiency as a function of the momentum (top left), transverse momentum (top right), OV_z (bottom left) and number of long tracks (bottom right). In all plots, blue points present the results for the efficiency measured on the simulation, red ones for the data. The evolution is consistent with the expectations: the tracking efficiency decreases at low or high momenta, where spurious effect like the multiple scattering can occur or the intrinsic spectrometer limits the reconstruction. As a function of the OV_z, the efficiency has the maximum upstream and around the nominal interaction point and decreases downstream as a consequence of the lower number of VELO energy deposits used for the track reconstruction. Within the statistical uncertainties, the results on data and simulation are found to be consistent, excluding large bias effects for the efficiency measurement on simulation. No contribution to the systematic uncertainty is hence added.

II.2.6.2 Fit procedure

As discussed in Sec. II.2.4, the number of signal events in data and in simulation is obtained by modelling the $\bar{\Lambda}$ candidates invariant mass distribution with the sum of a Voigtian and two Gaussian distributions for the signal and a polynomial function for the background. The possible contribution to the systematic uncertainty due to the choice of a parametric function is scrutinised here by repeating the fit with the other model that, with an additional Gaussian, was found to correctly describe the signal decays. With this method, the uncertainty related to the signal parametrization only is obtained, but, being the purity of the selected sample larger than 90%, the contribution due to the background is expected to be negligible.

The fit results with the two models are compared as a function of the antiproton momentum and transverse momentum, but with a reduced number of bins with respect to Sec. II.2.4 in order to mitigate the statistical fluctuations. The considered bin limits for this study, defined requiring a similar population in each interval, are:

$$p : [12, 15.2, 18.3, 21.6, 25.4, 30.2, 38.3, 110]\,\mathrm{GeV}/c,$$
$$p_{\mathrm{T}} : [0.4, 0.55, 0.7, 0.9, 4]\,\mathrm{GeV}/c.$$

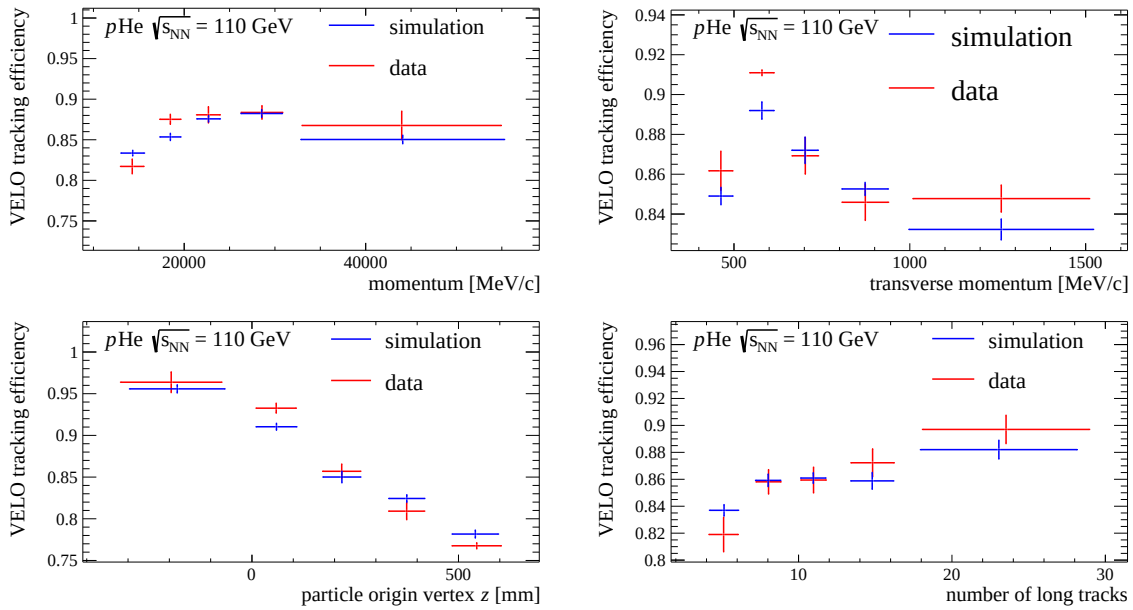


Figure II.2.33: Antiproton VELO track reconstruction efficiency measured on data (in red) and on simulation (in blue) as a function of its momentum (top left), transverse momentum (top right), origin vertex (bottom left) and number of tracks in the event (top right). A good consistency between the two measurements is found.

In each bin, the models with the sum of a Voigtian and two or three Gaussian distributions are applied to the $\overline{\Lambda}$-enriched simulation and to the data sample, with the results for the number of signal events presented in Fig. II.2.34 and II.2.35, respectively. The fit results are found to be consistent between the two fit models within the statistical uncertainty and no contribution is hence added to the systematics uncertainty.

II.2.6.3 Interactions with the detector material

Another contribution to the systematic uncertainty is related to the precision of the simulation to describe the final-state particle interactions with the detector, which is affected by both the description of the detector material and of the relevant cross-sections. Following the prescriptions of the LHCb tracking group, the total systematic uncertainty is estimated to correspond to 10% of the number of interacting particles [134]. This was measured for the antiprotons in Ref. [61], with the results in the kinematic binning scheme considered for that analysis illustrated in top Fig. II.2.36. To extrapolate the result to the kinematic intervals adopted for this analysis, a parametrization of the fraction of interacting particle as a function of the pseudorapidity and the momentum is performed. A polynomial model is assumed and the coefficients are determined with a fit to the distribution as:

$$\overline{p} \text{ fraction}[\%] = 1.31 + 0.68 \cdot \eta + 0.1 \cdot \eta^2 - 0.3 \cdot p\,[\,\text{GeV}/c]. \tag{II.2.21}$$

The results for the binning scheme of this analysis, highlighted with red boxes, are shown in the bottom plot of Fig. II.2.36. In most bins, the found value is below 2%, reaching 4% at the pseudorapidity boundary regions.

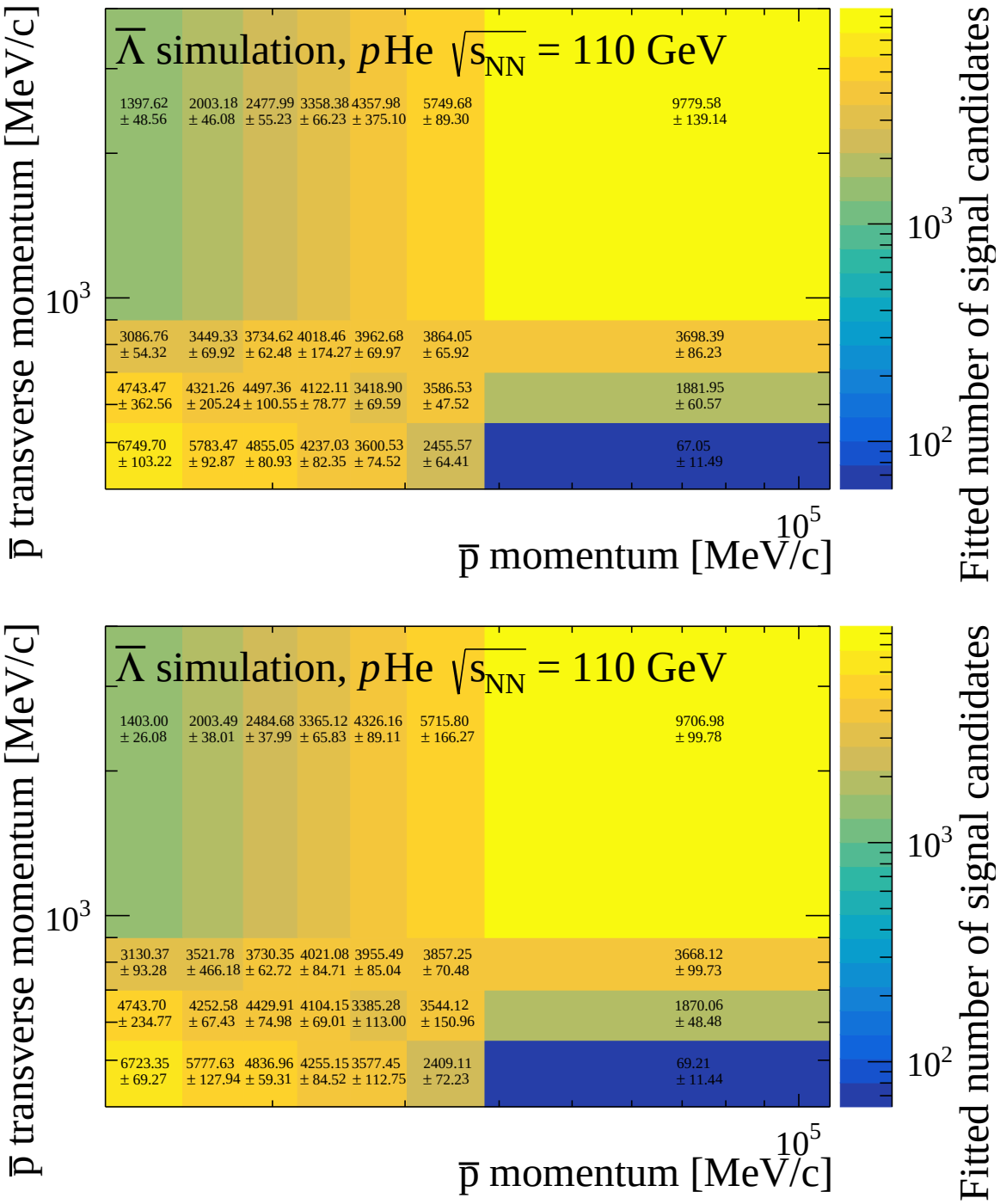


Figure II.2.34: Fitted number of signal events by modelling the $\overline{\Lambda}$ candidates invariant mass distribution with the sum of a Voigtian and two (top) or three (bottom) Gaussians in the $\overline{\Lambda}$-enriched simulated sample.

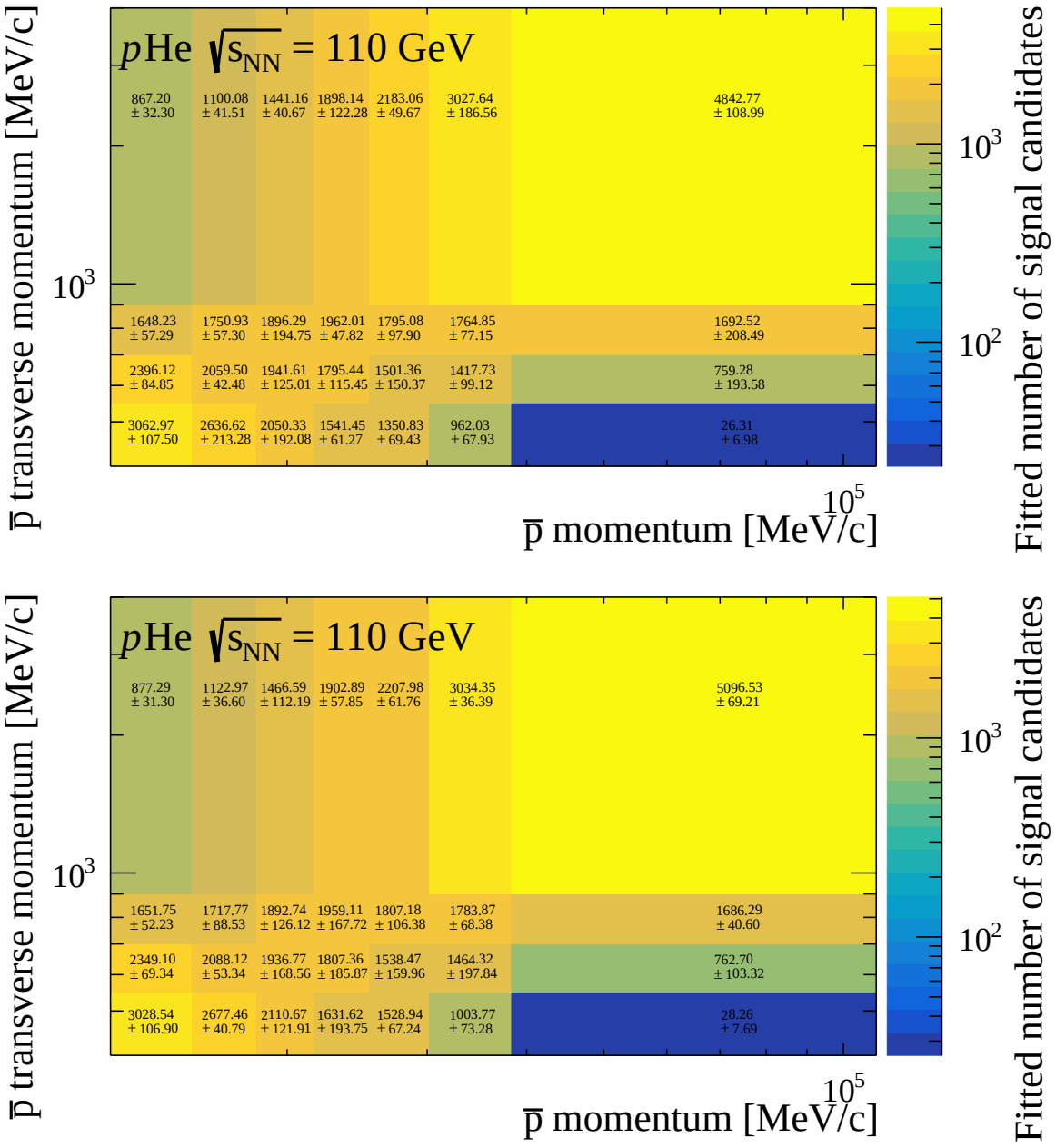


Figure II.2.35: Fitted number signal events by modelling the $\bar{\Lambda}$ candidates invariant mass distribution with the sum of a Voigtian and two (top) or three (bottom) Gaussians in data.

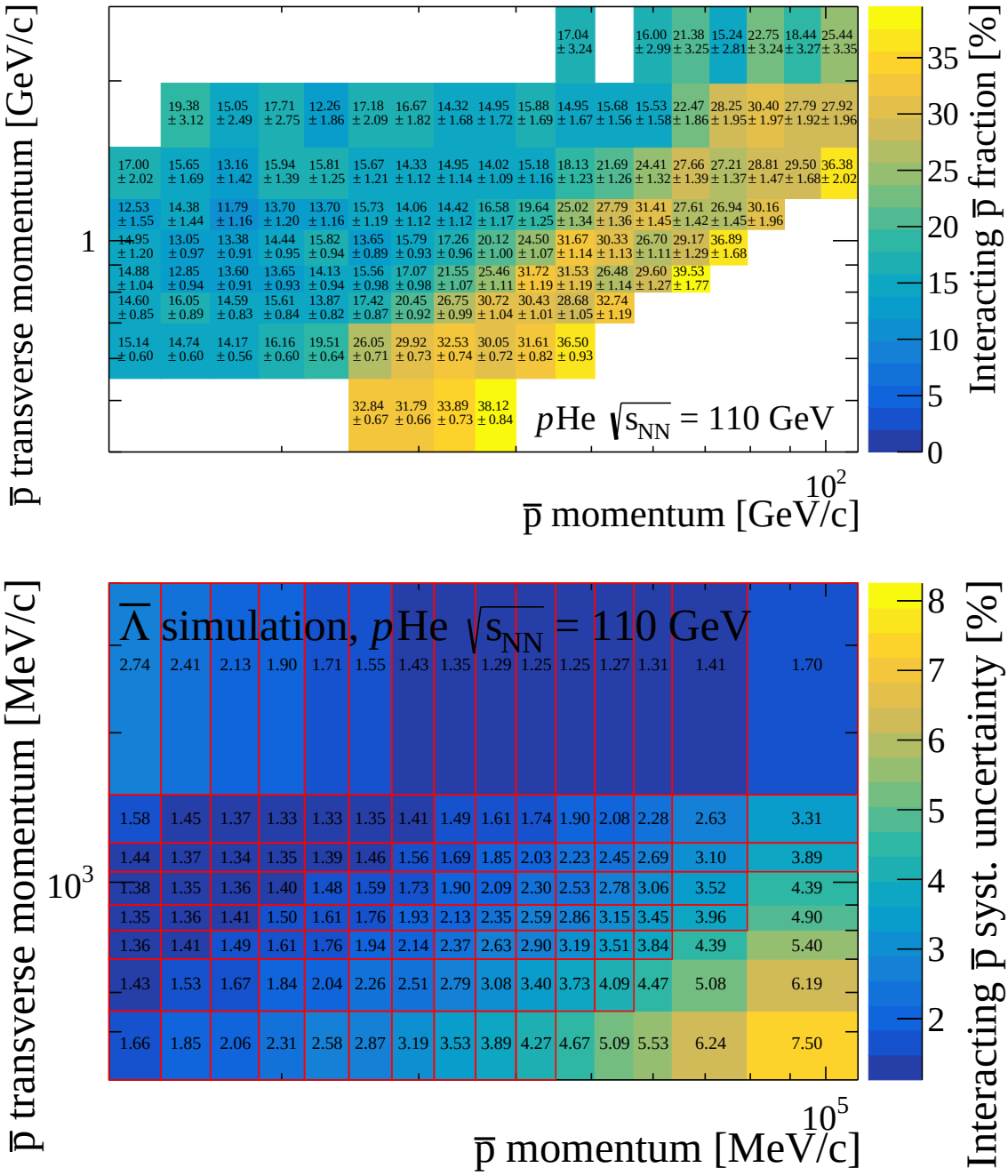


Figure II.2.36: Fraction of interacting antiprotons in the momentum-transverse momentum bins adopted in Ref. [61] (top) and uncertainty assigned for this analysis as described in the text (bottom).

Table II.2.5: Summary of the uncertainties affecting the measurement precision.

Luminosity measurement	6.1%
Material interactions	1.25 - 4.3 % (most bins $< 2\%$)
Simulated statistics	most bins $< 4\%$
Tracking	0.8%
Fit model choice	negligible

II.2.7 Results for the exclusive approach

The fitted number of signal events in the data sample illustrated in Fig. II.2.20, the total efficiency for the $\overline{\Lambda}$-enriched simulation in bottom Fig. II.2.30 and the relevant contributions to the experimental uncertainty, summarised in Tab. II.2.5, are substituted in Eq. II.2.1 to measure $\sigma(p\mathrm{He} \to (\overline{\Lambda}_{\mathrm{prompt}} \to \overline{p}\pi^+)X)$ at $\sqrt{s_{\mathrm{NN}}} = 110$ GeV. The double-differential cross-section results in the considered kinematic binning scheme are presented in Fig. II.2.37 (top) and projected over the momentum (bottom left) and transverse momentum (bottom right) axes. For the projections, the uncertainty originating from the luminosity and the interactions with the detector material are assumed to be fully correlated among kinematic bins, the other contributions as not correlated.

In order to perform a comparison with the inclusive approach, the obtained results are divided by the prompt antiproton ones [61] to measure:

$$R_{\overline{\Lambda}} = \frac{\sigma(p\mathrm{He} \to (\overline{\Lambda}_{\mathrm{prompt}} \to \overline{p}\pi^+)X)}{\sigma(p\mathrm{He} \to \overline{p}_{\mathrm{prompt}}\, X)}. \tag{II.2.22}$$

For this ratio, the uncertainties due to the luminosity measurement and to the antiproton interactions with the detector material, which are the leading ones as listed in Tab. II.2.5, cancel out. The prompt antiproton results are shown in Fig. II.2.38 in the finer kinematic binning scheme adopted in Ref. [61], where also the list of the numerical results is reported. The measured values for the cross-section are presented in the top plot, while the uncertainties uncorrelated and correlated among the kinematic bins are separately shown in the bottom left and in the bottom right, respectively. To measure the $R_{\overline{\Lambda}}$ ratio, the binning scheme of the prompt analysis is adapted to that considered in this chapter. The prompt antiproton production cross-section in exclusive bins $\Delta p \Delta p_{\mathrm{T}}$ that result from the merging of more prompt bins is given by:

$$\frac{d^2\sigma}{dp dp_{\mathrm{T}}} = \frac{1}{\Delta p \Delta p_{\mathrm{T}}} \sum_j \frac{d^2\sigma'}{dp' dp'_{\mathrm{T}}}^{\,j} \cdot \delta_p^j \delta_{pT}^j, \tag{II.2.23}$$

being $\frac{d^2\sigma'}{dp' dp'_{\mathrm{T}}}^{\,j}$ and $\delta_p^j \delta_{pT}^j$ the antiproton production result and the bin widths within the $\Delta p \cdot \Delta p_{\mathrm{T}}$ region. The uncorrelated uncertainty is propagated summed in quadrature, the correlated one linearly (except the luminosity and interaction contributions, which are removed as mentioned above). Exclusive bins containing a prompt bin excluded in the prompt analysis and the $p_{\mathrm{T}} \in [0.4, 0.55]$ GeV/c region where the relative uncertainty

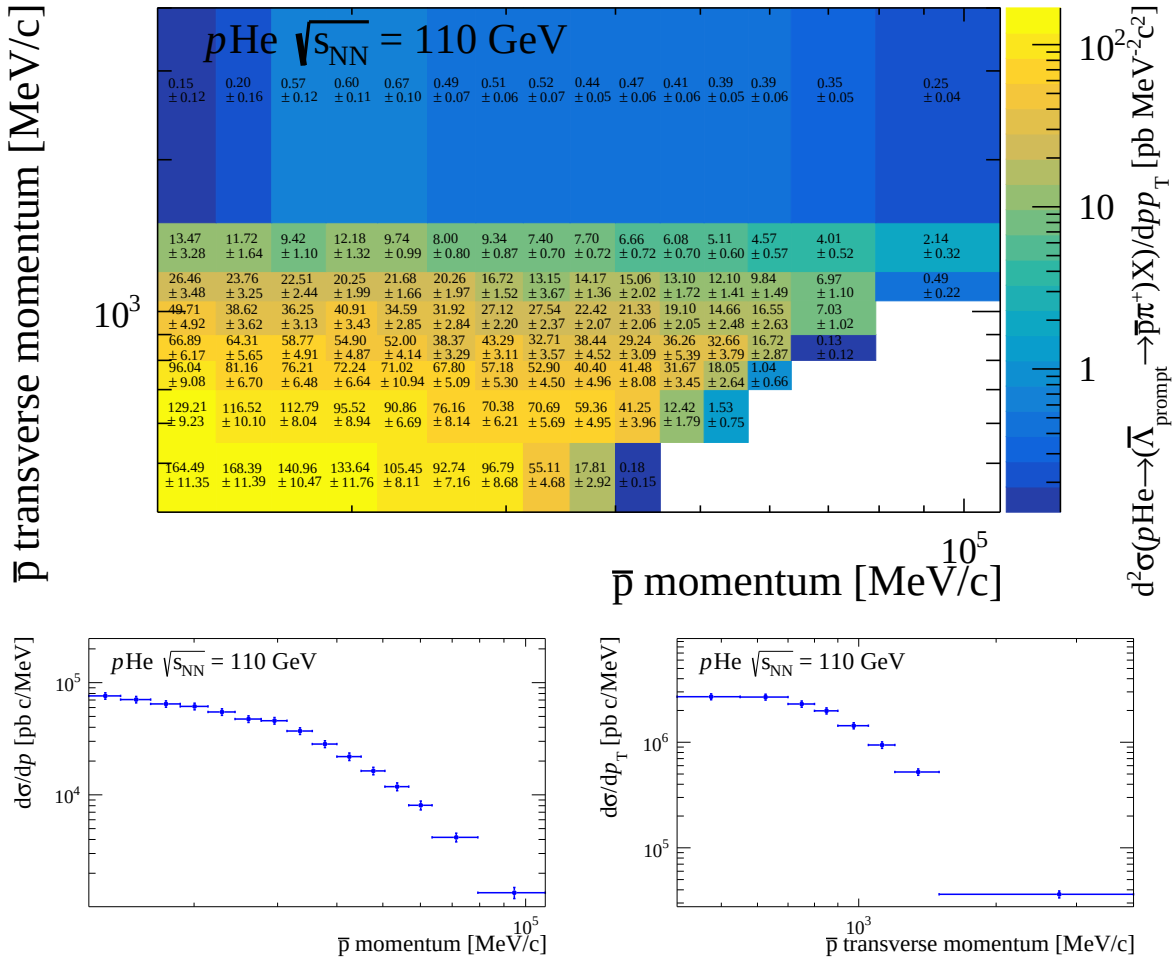


Figure II.2.37: Double differential cross-section for $\bar{\Lambda}_{\text{prompt}} \to \bar{p}\pi^+$ decays in pHe collisions at $\sqrt{s_{\text{NN}}} = 110\,\text{GeV}$ in the considered antiproton kinematic binning scheme (top) and projections over the momentum (bottom left) and transverse momentum (bottom right) axes.

for prompt antiprotons is above 20% are not considered for the $R_{\bar{\Lambda}}$ measurement. The obtained results for the $R_{\bar{\Lambda}}$ ratio in the bidimensional binning scheme of this chapter are presented in top Fig. II.2.39 and listed in Tabs. II.2.6 and II.2.7. The middle and bottom plots of the figure present the projections of the results over the momentum and transverse momentum axes and compare the measurement with the predictions of the EPOS-LHC generator (in green). For the projections, the uncertainties correlated (uncorrelated) among the kinematic bins are summed linearly (in quadrature). The measured $R_{\bar{\Lambda}}$ values are found to show a mild dependence on the antiproton transverse momentum, while they decrease more steeply with the momentum. A similar excess with respect to the EPOS-LHC predictions as in the inclusive approach is found.

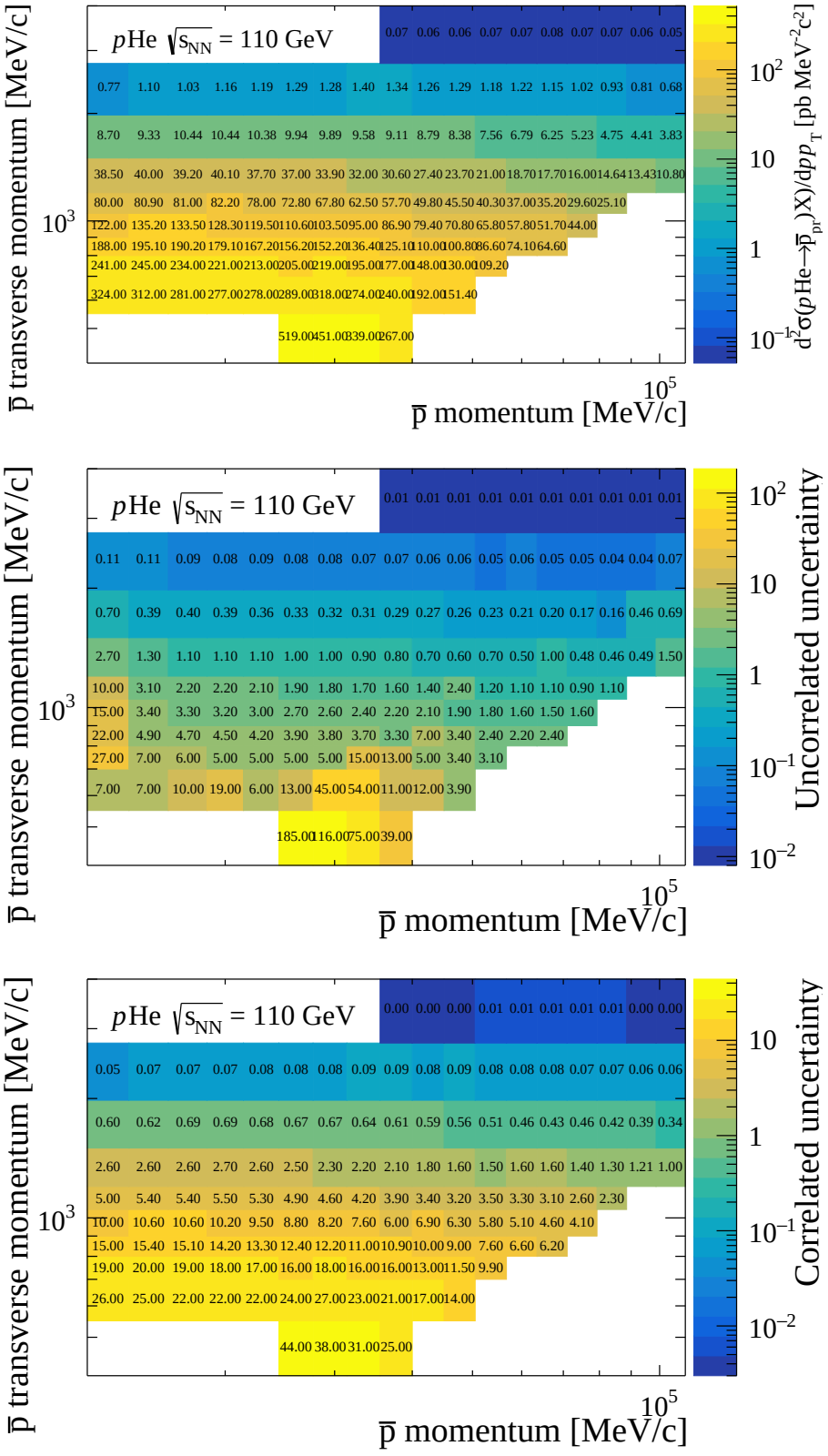


Figure II.2.38: Results for the double differential cross-section for the prompt antiproton production in pHe collisions in the antiproton momentum-transverse momentum binning scheme adopted in Ref. [61]. The top plot presents the numerical cross-section values, the middle (bottom) one the uncorrelated (correlated) uncertainties.

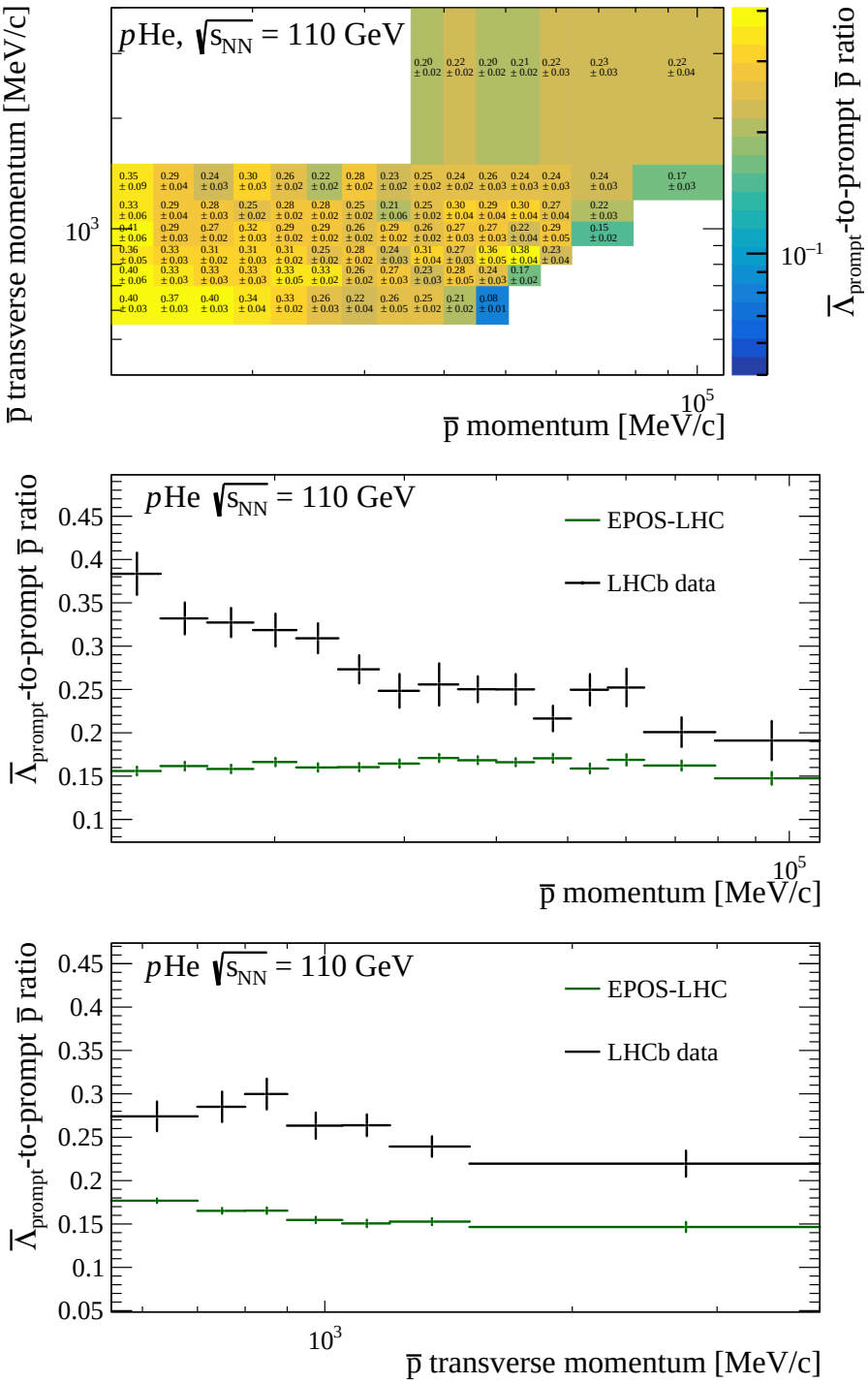


Figure II.2.39: Measurement result for the $\bar{\Lambda}_{\text{prompt}}$-to-prompt antiproton production in the LHCb 2016 pHe data. The top plot shows the numerical results in the binning scheme adopted in this chapter, the middle (bottom) one the projection over the momentum (transverse momentum) axes. For the one-dimensional plot, the prediction of the EPOS-LHC generator is added and a large underestimation of the data is found.

Table II.2.6: Numerical results for the $\overline{\Lambda}_{\text{prompt}}$-to-prompt antiproton ratio in intervals of the antiproton momentum and transverse momentum.

p [GeV/c]	p_{T} [GeV/c]	$R_{\bar{\Lambda}}$	$\delta R_{\bar{\Lambda}}$	p [GeV/c]	p_{T} [GeV/c]	$R_{\bar{\Lambda}}$	$\delta R_{\bar{\Lambda}}$
12.0 - 14.0	0.55 - 0.7	0.40	0.03	18.7 - 21.4	0.9 - 1.05	0.32	0.03
12.0 - 14.0	0.7 - 0.8	0.40	0.06	18.7 - 21.4	1.05 - 1.2	0.25	0.02
12.0 - 14.0	0.8 - 0.9	0.36	0.05	18.7 - 21.4	1.2 - 1.5	0.30	0.03
12.0 - 14.0	0.9 - 1.05	0.41	0.06	21.4 - 24.4	0.55 - 0.7	0.33	0.02
12.0 - 14.0	1.05 - 1.2	0.33	0.06	21.4 - 24.4	0.7 - 0.8	0.33	0.05
12.0 - 14.0	1.2 - 1.5	0.35	0.09	21.4 - 24.4	0.8 - 0.9	0.31	0.02
14.0 - 16.2	0.55 - 0.7	0.37	0.03	21.4 - 24.4	0.9 - 1.05	0.29	0.02
14.0 - 16.2	0.7 - 0.8	0.33	0.03	21.4 - 24.4	1.05 - 1.2	0.278	0.016
14.0 - 16.2	0.8 - 0.9	0.33	0.03	21.4 - 24.4	1.2 - 1.5	0.26	0.02
14.0 - 16.2	0.9 - 1.05	0.29	0.03	24.4 - 27.7	0.55 - 0.7	0.26	0.03
14.0 - 16.2	1.05 - 1.2	0.29	0.04	24.4 - 27.7	0.7 - 0.8	0.33	0.02
14.0 - 16.2	1.2 - 1.5	0.29	0.04	24.4 - 27.7	0.8 - 0.9	0.25	0.02
16.2 - 18.7	0.55 - 0.7	0.40	0.03	24.4 - 27.7	0.9 - 1.05	0.29	0.02
16.2 - 18.7	0.7 - 0.8	0.33	0.03	24.4 - 27.7	1.05 - 1.2	0.28	0.02
16.2 - 18.7	0.8 - 0.9	0.31	0.02	24.4 - 27.7	1.2 - 1.5	0.216	0.019
16.2 - 18.7	0.9 - 1.05	0.27	0.02	27.7 - 31.4	0.55 - 0.7	0.22	0.04
16.2 - 18.7	1.05 - 1.2	0.28	0.03	27.7 - 31.4	0.7 - 0.8	0.26	0.02
16.2 - 18.7	1.2 - 1.5	0.24	0.03	27.7 - 31.4	0.8 - 0.9	0.284	0.019
18.7 - 21.4	0.55 - 0.7	0.34	0.04	27.7 - 31.4	0.9 - 1.05	0.26	0.02
18.7 - 21.4	0.7 - 0.8	0.33	0.03	27.7 - 31.4	1.05 - 1.2	0.247	0.019
18.7 - 21.4	0.8 - 0.9	0.31	0.03	27.7 - 31.4	1.2 - 1.5	0.28	0.02
18.7 - 21.4	0.9 - 1.05	0.32	0.03	31.4 - 35.5	0.55 - 0.7	0.26	0.05

Table II.2.7: Numerical results for the $\overline{\Lambda}_{\mathrm{prompt}}$-to-prompt antiproton ratio in intervals of the antiproton momentum and transverse momentum.

p [GeV/c]	p_{T} [GeV/c]	$R_{\bar{\Lambda}}$	$\delta R_{\bar{\Lambda}}$	p [GeV/c]	p_{T} [GeV/c]	$R_{\bar{\Lambda}}$	$\delta R_{\bar{\Lambda}}$
31.4 - 35.5	0.7 - 0.8	0.27	0.03	45.0 - 50.5	0.8 - 0.9	0.36	0.05
31.4 - 35.5	0.8 - 0.9	0.24	0.03	45.0 - 50.5	0.9 - 1.05	0.27	0.03
31.4 - 35.5	0.9 - 1.05	0.29	0.02	45.0 - 50.5	1.05 - 1.2	0.29	0.04
31.4 - 35.5	1.05 - 1.2	0.21	0.06	45.0 - 50.5	1.2 - 1.5	0.26	0.03
31.4 - 35.5	1.2 - 1.5	0.231	0.019	45.0 - 50.5	1.5 - 4.0	0.20	0.02
35.5 - 40.0	0.55 - 0.7	0.25	0.02	50.5 - 56.7	0.7 - 0.8	0.17	0.02
35.5 - 40.0	0.7 - 0.8	0.23	0.03	50.5 - 56.7	0.8 - 0.9	0.38	0.04
35.5 - 40.0	0.8 - 0.9	0.31	0.04	50.5 - 56.7	0.9 - 1.05	0.22	0.04
35.5 - 40.0	0.9 - 1.05	0.26	0.02	50.5 - 56.7	1.05 - 1.2	0.30	0.04
35.5 - 40.0	1.05 - 1.2	0.25	0.02	50.5 - 56.7	1.2 - 1.5	0.24	0.03
35.5 - 40.0	1.2 - 1.5	0.25	0.02	50.5 - 56.7	1.5 - 4.0	0.21	0.02
35.5 - 40.0	1.5 - 4.0	0.20	0.02	56.7 - 63.5	0.8 - 0.9	0.23	0.04
40.0 - 45.0	0.55 - 0.7	0.21	0.02	56.7 - 63.5	0.9 - 1.05	0.29	0.05
40.0 - 45.0	0.7 - 0.8	0.28	0.05	56.7 - 63.5	1.05 - 1.2	0.27	0.04
40.0 - 45.0	0.8 - 0.9	0.27	0.03	56.7 - 63.5	1.2 - 1.5	0.24	0.03
40.0 - 45.0	0.9 - 1.05	0.27	0.03	56.7 - 63.5	1.5 - 4.0	0.22	0.03
40.0 - 45.0	1.05 - 1.2	0.30	0.04	63.5 - 79.3	0.9 - 1.05	0.15	0.02
40.0 - 45.0	1.2 - 1.5	0.24	0.02	63.5 - 79.3	1.05 - 1.2	0.22	0.03
40.0 - 45.0	1.5 - 4.0	0.22	0.02	63.5 - 79.3	1.2 - 1.5	0.24	0.03
45.0 - 50.5	0.55 - 0.7	0.082	0.012	63.5 - 79.3	1.5 - 4.0	0.23	0.03
45.0 - 50.5	0.7 - 0.8	0.24	0.03	79.3 - 110.0	1.2 - 1.5	0.17	0.03
45.0 - 50.5	0.8 - 0.9	0.36	0.05	79.3 - 110.0	1.5 - 4.0	0.22	0.04

II.3 Analysis conclusions

After the first determination of the prompt antiproton production in pHe collisions at $\sqrt{s_{\rm NN}} = 110\,\text{GeV}$ performed by the LHCb experiment in its fixed-target configuration, another production mechanism, accounting for 25-40% of prompt antiprotons, is that of the anti-hyperon decays $\overline{\Lambda} \to \overline{p}\pi^+$ (with a promptly- or $\overline{\Xi}$-produced $\overline{\Lambda}$ particle) and $\overline{\Sigma}^- \to \overline{p}\pi^0$. Limited experimental data were available for the anti-hyperon relative contribution and the dedicated measurement at LHCb has been documented in this part.
Two different approaches have been followed and compared. In the former, called inclusive and presented in Chap II.1, all antiprotons produced in pHe collisions are reconstructed and selected and the distinction between prompt and produced in the anti-hyperons decays is obtained by their different impact parameter distribution. The measured detached-to-prompt antiproton ratio:

$$R = \frac{\sigma(p\text{He} \to (\overline{\Lambda} \to \overline{p}\pi^+)\, X) + \sigma(p\text{He} \to (\overline{\Sigma}^- \to \overline{p}\pi^0)\, X)}{\sigma(p\text{He} \to \overline{p}_{\text{prompt}}\, X)} \tag{II.3.1}$$

is compared as a function of the antiproton momentum and transverse momentum with the predictions of the most common theoretical models used in CR physics. While the R evolution with the kinematics is found to be correctly reproduced by these models, a larger absolute value is measured.

An independent measurement for the quantity of antiprotons produced in $\overline{\Lambda}_{\text{prompt}} \to \overline{p}\pi^+$ decays is described in Chap. II.2. This is complementary to the inclusive one, being the antiprotons selected without any requirement on the PID information coming from the RICH system. The results obtained for the cross-section are divided by the prompt antiproton production from Ref. [61], measuring:

$$R_{\overline{\Lambda}} = \frac{\sigma(p\text{He} \to (\overline{\Lambda}_{\text{prompt}} \to \overline{p}\pi^+)X)}{\sigma(p\text{He} \to \overline{p}_{\text{prompt}}\, X)}, \tag{II.3.2}$$

where most of the systematic uncertainty, notably the leading one due to the measurement of the luminosity of the pHe data sample, cancels out.

To directly compare the results of the two approaches, the inclusive binning scheme in the $p \in [12, 50.5]\,\text{GeV}/c$, $p_{\text{T}} \in [0.55, 1.2]\,\text{GeV}/c$ region, where the prompt, inclusive

and exclusive measurements are all available, is considered. The prompt and exclusive cross-sections are rebinned according to Eq. II.2.23, while the inclusive bins outside that kinematic region are discarded and the projections are repeated. Figures II.3.1 compare the R (in red) and $R_{\bar{\Lambda}}$ (in blue) measurements as a function of the antiproton momentum (top) and transverse momentum (bottom). As expected, the $R_{\bar{\Lambda}}$ values, only considering one term among all anti-hyperon contributions, are consistently lower than R and follow a similar evolution with the antiproton kinematics. The ratio between the results obtained with the inclusive and exclusive approaches represents the fraction of detached antiprotons originating in a decay of a promptly-produced $\bar{\Lambda}$ baryon. It is presented in Fig. II.3.2 as a function of the antiproton momentum (top) and transverse momentum (bottom) and compared to the prediction according to the EPOS-LHC generator. Such a prediction, largely based on isospin symmetry arguments, is expected to be more robust than that for the detached-to-prompt antiproton ratio. In most bins, the measurement is found to be consistent with the prediction within the experimental uncertainties, with a slight indication of a steeper dependence on the momentum. The observed overall consistency between the two curves validates the followed approaches.

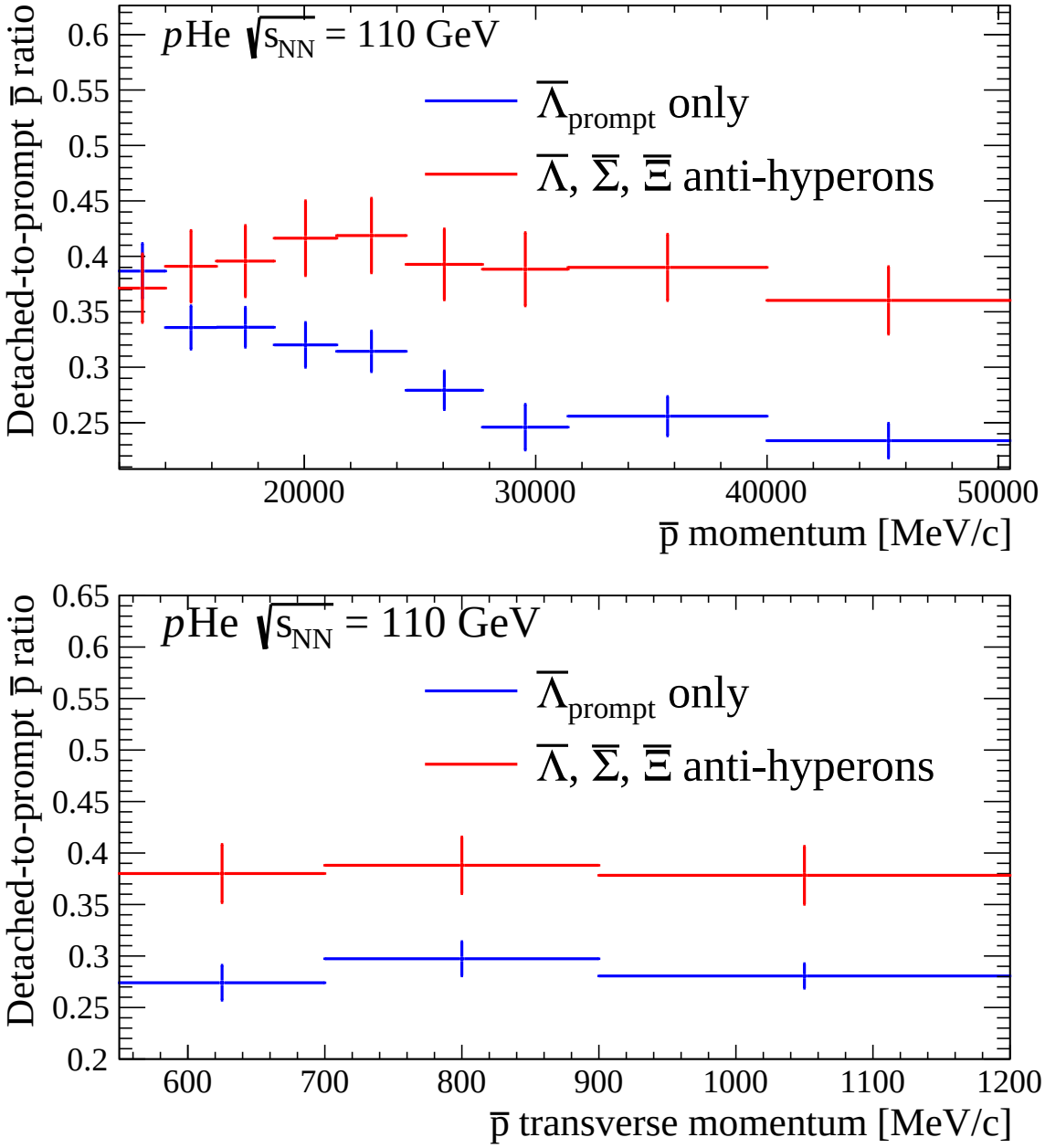


Figure II.3.1: Detached-to-prompt (R) and $\overline{\Lambda}_{\text{prompt}}$-produced-to-prompt ($R_{\overline{\Lambda}}$) antiproton ratios measured in the analysis discussed in this document with the inclusive (in red) and exclusive (in blue) approach, respectively, as a function of the antiproton momentum (top) and transverse momentum (bottom).

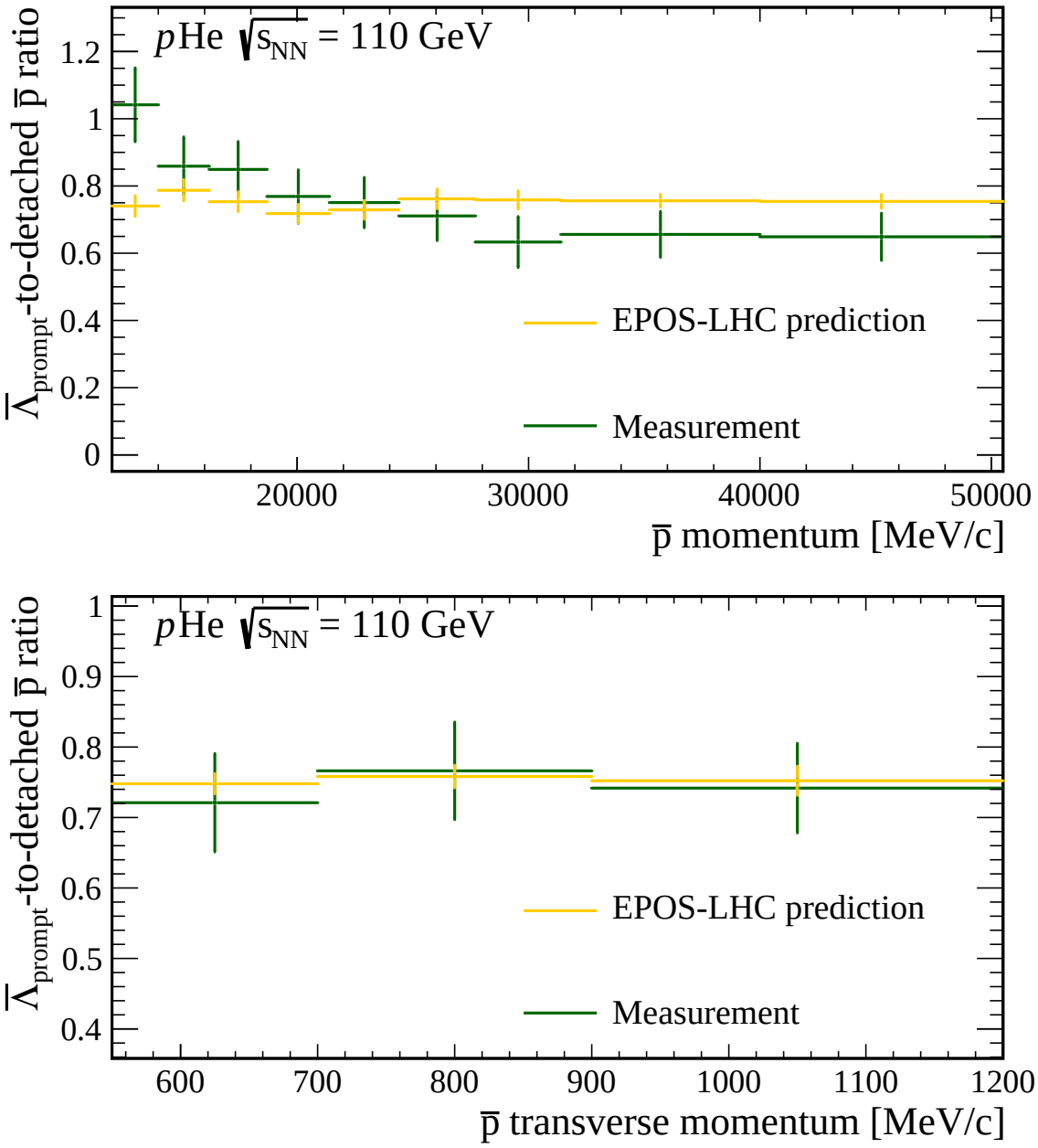


Figure II.3.2: EPOS-LHC prediction (in yellow) and measurement (in green) for the fraction of detached antiprotons originating in a promptly-produced $\bar{\Lambda}$ decay as a function of the antiproton momentum (top) and transverse momentum (bottom). The measurement is obtained as the ratio between the the inclusive and the exclusive results.

Part III

A machine-learning approach to particle identification

One of the leading uncertainties for the analyses of fixed-target samples is the Particle IDentification (PID) in the LHCb detector. While for pp data high-statistics calibration channels are used to validate PID analyses, for fixed-target samples, mostly collected in short dedicated periods, their statistics is insufficient. The exploitation of the pp calibration channels for SMOG data is also not suitable, considering their poor overlap in the distributions of the quantities affecting the PID detector responses. For the prompt antiproton production analysis, as discussed in Sec. III.1.1, the PID is the next-to-leading contribution to the uncertainty.

During my *Ph.D.* programme, I conceived and developed a new approach to the modelling of PID in fixed-target data employing machine-learning techniques. As described in Sec. III.1.2, the PID classifiers, built by combining the response of the dedicated detectors, depend on several variables, such as those describing the particle kinematics, the event geometry and the data-taking conditions. In the proposed model, the PID classifiers of probe particles in calibration channels are described as a Gaussian Mixture Model (GMM). The parameters of all the multinormal functions are determined by neural networks fed with the relevant variables. Iteratively, the PID dependence on these variables, also taking into account their correlation, is learned. The model can be used to predict the PID classifier distributions for samples obtained with different data-taking conditions and applied selections. Section III.1.3 contains an example of the proposed method. The calibration channels for the pion, kaon and antiproton particle species are reconstructed and selected in the 2017 pNe data and a GMM is trained and validated for each of them. Section III.1.4 presents three applications of the proposed approach: the generalization of the trained model to two lower-statistics pHe and pAr samples, where the approach is compared to models based on the detailed LHCb simulation, the study of the different PID response for prompt and detached antiprotons in pHe and the understanding of a long-standing tension between two calibration samples for kaon PID response in pp data.
All the research activities documented in this part have been performed as part of my *Ph.D.* programme and are published in Ref. [135].

III.1.1 Limitations of SMOG particle identification

Particle IDentification (PID) at high-energy physics experiments typically relies on a set of detectors offering enough redundancy to identify the particle species in a wide range of features, notably their momentum. The description of the PID classifier probability density function (pdf) as a function of the quantities affecting the response of these detectors is crucial for physics analyses. For charged hadrons at the LHCb experiment, the identification mostly relies on the RICH system, whose response depends on:

- the particle momentum, determining the Cherenkov photon emission angle according to Eq. I.3.6;
- the particle propagation in the RICH system, which is a function of its origin vertex, its direction and the quality of its reconstruction;
- the total number of hits in the RICH detectors and of the reconstructed tracks.

To exemplify the PID performance on fixed-target data, the strategy followed in the measurement of the antiproton production cross-section in pHe collisions [61], introduced in Sec. I.4.2.2, is here discussed in more details. For that analysis, the selected sample of negatively-charged particles contains a mixture of pions, kaons and antiprotons, being the abundances of all the other species negligible. Rather than applying a selection, antiprotons are distinguished from the other negatively-charged particles by considering the $\mathrm{DLL}_{p,\pi}$-$\mathrm{DLL}_{p,K}$ distribution in pHe data, illustrated in one kinematic interval in the left plot of Fig. III.1.1. To measure the antiproton abundance, a fit with the addition of one template for each particle species and their relative abundances as free parameters is performed. The template distributions are drawn from the pHe simulation or from the $\bar{\Lambda} \to \bar{p}\pi^+$, $K^0_S \to \pi^+\pi^-$ and $\phi \to K^+K^-$ calibration channels reconstructed and selected in data with no PID requirements. A fourth category of tracks reconstructed from energy deposits left from more than one single particle, called ghosts in the following and accounting for 1.6% of the total selected particles, is added. The simulation provides the same experimental conditions as the data and adequate statistics, but biases in the results could arise because of the unavoidable imperfections in the description of the detector response. To mitigate this effect, the simulation is weighted with respect to the antiproton transverse momentum and the number of hits in the SPD detector in order to reproduce the data distributions. With the applied weights, the right plot of Fig. III.1.1 shows a projection of a fit in one kinematic interval with simulated templates onto a combination of the DLL variables where the three particle species can be distinguished. Imperfections in the description of the pHe data can be seen, as a consequence of the small size of the calibration samples and the consequent large statistical uncertainty affecting the evaluated weights. For $p > 30\,\mathrm{GeV}/c$, calibration channels reconstructed and selected in pp collisions have been hence considered. These are abundantly produced, but, as a consequence of their different distribution in the quantities affecting the response of the RICH detectors, need to be weighted as well. In particular, for the calibration channels reconstructed from pp data, the occupancy in the detector is much larger than pHe and there is no statistics for $|z| > 100\,\mathrm{mm}$. Their application to the pHe data was hence found to be poorly efficient. As listed in Tab. I.4.1, the resulting systematic uncertainty on the cross-section measurement was the next-to-leading one after the luminosity contribution.

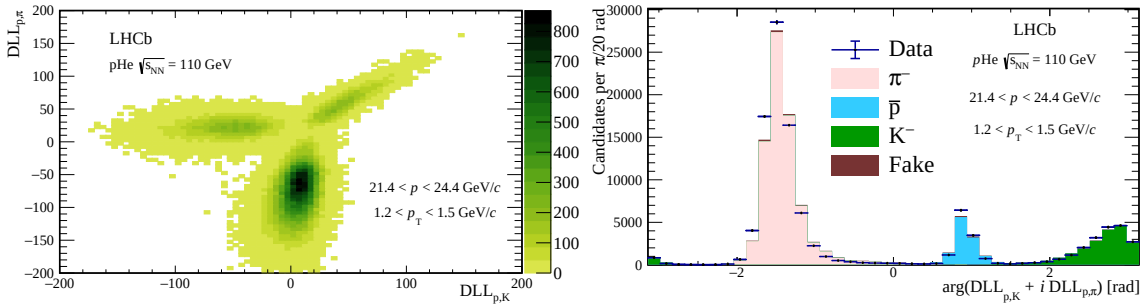


Figure III.1.1: Example of the PID performance on fixed-target data. The left plot shows the DLL$_{p,\pi}$-DLL$_{p,K}$ distribution for all negatively-charged prompt particles selected in the pHe data. To distinguish the antiprotons from kaons and pions, a fit to the pHe data is performed, with the template distributions drawn for all particle species on simulation or on calibration channels reconstructed and selected in the pHe data and with their relative abundances as free parameters. The right plot shows an example of the fit projected onto arg(DLL$_{p,K}$ + i DLL$_{p,\pi}$). Imperfections in the description of data can be seen, as a consequence of the poor calibration statistics. Figures from Ref. [61].

III.1.2 Machine-learning approach

The core idea of the proposed model is to learn from a high-statistics data sample the marginal pdf of a PID classifier x as a function of some relevant experimental features, denoted with $\underline{\theta}$. According to the distributions of the same variables in smaller samples, a PID template distribution can be generated, overcoming the limitation of the poor calibration statistics.
During the training, the x pdf in a calibration channel is modelled as a Gaussian Mixture Model (GMM), the sum of N_g normal distributions $\mathcal{G}$:

$$x_p \sim \sum_{j=1}^{N_{g,p}} \alpha_{j,p}(\underline{\theta}) \mathcal{G}(x, \mu_{j,p}(\underline{\theta}), \sigma_{j,p}(\underline{\theta})), \tag{III.1.1}$$

denoting p the considered particle species. The choice of the normal distribution ensures a fast convergence of the model and, with N_g large enough, any smooth profile with exponential-like tails, as it is the case for the PID classifiers built at LHCb, can be described. In order to generalise to a PID bidimensional distribution, as for the physics-case described in the previous section, the GMM representation can be extended to a vector target $\underline{x}$ by replacing the normal with multinormal distributions:

$$\underline{x}_p \sim \sum_{j=1}^{N_{g,p}} \alpha_{j,p}(\underline{\theta}) \frac{\exp(-\frac{1}{2}(\underline{x}_p - \underline{\mu}_{j,p}(\underline{\theta}))^T \Sigma_{j,p}^{-1}(\underline{\theta}) (\underline{x}_p - \underline{\mu}_{j,p}(\underline{\theta})))}{2\pi\sqrt{\det(\Sigma_{j,p}(\underline{\theta}))}} \tag{III.1.2}$$

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}. \tag{III.1.3}$$

For each multinormal distribution in the GMM, all the free parameters, *i.e.* the central values $\underline{\mu}$, the standard deviations (σ_1, σ_2), the correlation coefficient ρ and the relative weight in the combination α show a non-trivial evolution as a function of $\underline{\theta}$, which is what the machine-learning model has to learn. Each parameter is determined with a set of

Multi Layer Perceptron (MLP) Neural Network (NN) algorithms fed with the feature values $\underline{\theta}$. The NN internal parameters are iteratively determined by training on a set of calibration data n_p. The loss function, part of the novelty of the described approach, is defined as the opposite of the maximum-likelihood:

$$\mathcal{L} = -\sum_{i=1}^{n_p} w_i \log \left[\sum_{j=1}^{N_{g,p}} \alpha_{j,p}(\underline{\theta}_i) \mathcal{G}(\underline{x}_i, \mu_{j,p}(\underline{\theta}_i), \sigma_{j,p}(\underline{\theta}_i)) \right]. \tag{III.1.4}$$

In case the calibration channel is affected by a large background contamination, the *sPlot* [130] technique can be used to subtract it. The resulting weights for the signal hypothesis, w_i, appear in this case as multiplicative factors in the loss function [136–138]. By minimising $\mathcal{L}$ over a large number of processing iterations, epochs, the GMM parameters are adapted to the n_p calibration data and $\underline{x}_p(\underline{\theta})$ is learned.
Before the minimization of the loss function, $\underline{x}$ and $\underline{\theta}$ are processed with the scikit-learn [139] MinMaxScaler and QuantileTransformer algorithms, respectively. Since the model aims to predict the $\underline{x}$ pdf, the former only applies a scaling to the [0, 1) range, more easily digested by the NN. The latter, instead, is a non-linear transformation mapping the $\underline{\theta}$ distribution in the calibration data to a Gaussian. By equalising the range and, for the features, the functional form of the pdf, the numerical evaluation of the derivatives in the loss minimization is eased, resulting into a significant speed-up of the training phase. With the processed variables, an easy NN model with only few layers and a low number of nodes activated with a *tanh* function can be considered. The minimization is performed in two steps using a mini-batches gradient descent. The $\underline{x}$ pdf is firstly assumed to be independent on $\underline{\theta}$ and the NN parameters are determined with a low number of epochs. Secondly, the obtained results are used as input of the second step, where $\underline{x}$ ($\underline{\theta}$) is learned with a larger number of epochs.

The model has been developed with widely-used machine-learning libraries such as scikit-learn [139] and Keras [140] with its TensorFlow [141] backend, offering state-of-the-art NN functionalities such as the error backward propagation [142] and the auto-differentiation [143]. Also, we profited from the versatility between the CPU and GPU TensorFlow backends, testing the model performance on both a server with four AMD Opteron(TM) 6238 processors and an NVIDIA TU102GL [Quadro RTX 6000/8000] GPU card. The configuration of the model only requires to define:

- the number of multinormal distributions to be composed in the GMM, N_g;
- the NN structure, *i.e.* the number of its layers and nodes for each layer;
- the minimization parameters, *i.e.* the number of iterations, the initial step in the parameters and the number of calibration events for each iteration.

III.1.3 Model application on SMOG data

The general approach described in the previous section is applied as a benchmark application, summarised by the scheme in Fig. III.1.2, on LHCb fixed-target data. A GMM model is built and trained to describe the DLL distribution of calibration channels reconstructed

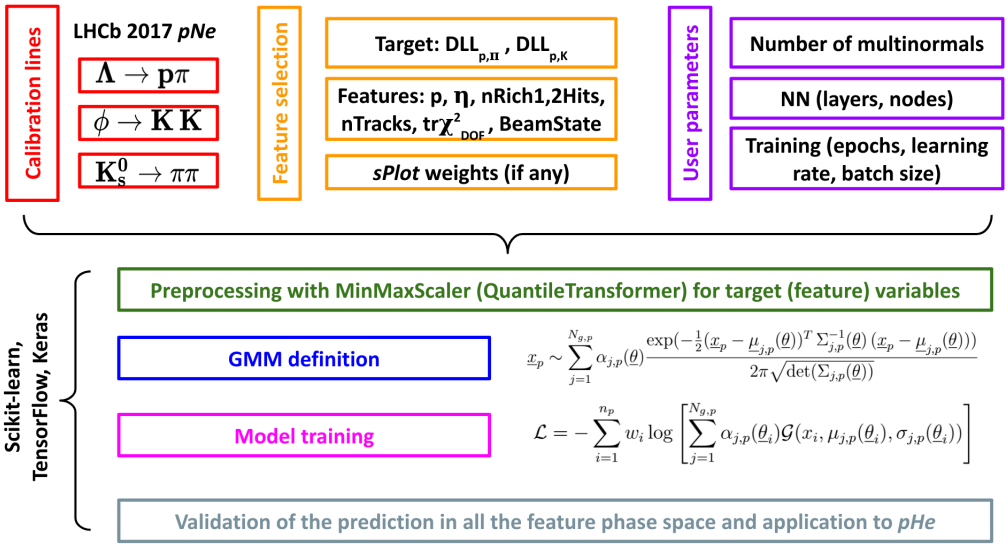


Figure III.1.2: Scheme of the benchmark application of the developed machine-learning approach to PID of fixed-target samples.

and selected in the 2017 pNe data and is then generalised to the 2016 pHe and the 2015 pAr samples. The selection of the calibration channels in the pNe data is detailed in Sec. III.1.3.1, the choice of the features in Sec. III.1.3.2. The training and validation of one model for each calibration channel are described in Sec. III.1.3.3 and III.1.3.4, respectively. The generalization to the pHe and pAr samples and the comparison of the proposed model with the LHCb detailed simulation is included in Sec. III.1.4.1.

III.1.3.1 Selection of the calibration channels

Calibration channels for the PID response to pions, kaons and antiprotons are reconstructed and selected in the pNe sample, collected in 2017 with $\sqrt{s_{\rm NN}} = 68\,\text{GeV}$ and offering the highest statistics among all Run2 fixed-target datasets, as shown in Fig. I.4.3. Though the energy and collision system vary among SMOG samples, there is a large overlap for the distributions of the features affecting the response of the PID detectors. This is exemplified by Fig. III.1.3, comparing the normalised distributions of the momentum in the $p \in [12, 110]$ range (top left), the pseudorapidity (top right), the number of hits in the RICH1 detector (middle left) and the OV$_z$ (middle right) for negative kaons in $\phi \to K^+K^-$ decays in the pp (in red), pHe (in blue) and pNe (in green) data. A larger phase-space overlap with pHe kaons is found for pNe than for pp data, as also reflected in the DLL$_{K,\pi}$ and DLL$_{p,K}$ comparisons in the bottom plots.

The $\Lambda \to p\pi^-$ ($\overline{\Lambda} \to \overline{p}\pi^+$)[1] and $K^0_S \to \pi^+\pi^-$ decays are considered to calibrate the PID response of (anti)protons and pions, respectively. Their selection in the pNe data, thanks to their narrow mass peaks and to the kinematic descriptions of the decays, achieves a larger purity than 99%. Concerning kaons, the $D^0 \to K^-\pi^+$ decay is typically considered for pp PID calibration. The charm production in fixed-target is too poor and the $\phi \to K^+K^-$ process is used instead. For all calibration channels, the

[1]Considering the presented benchmark application, the $\overline{\Lambda}$ decay will be considered to train a model.

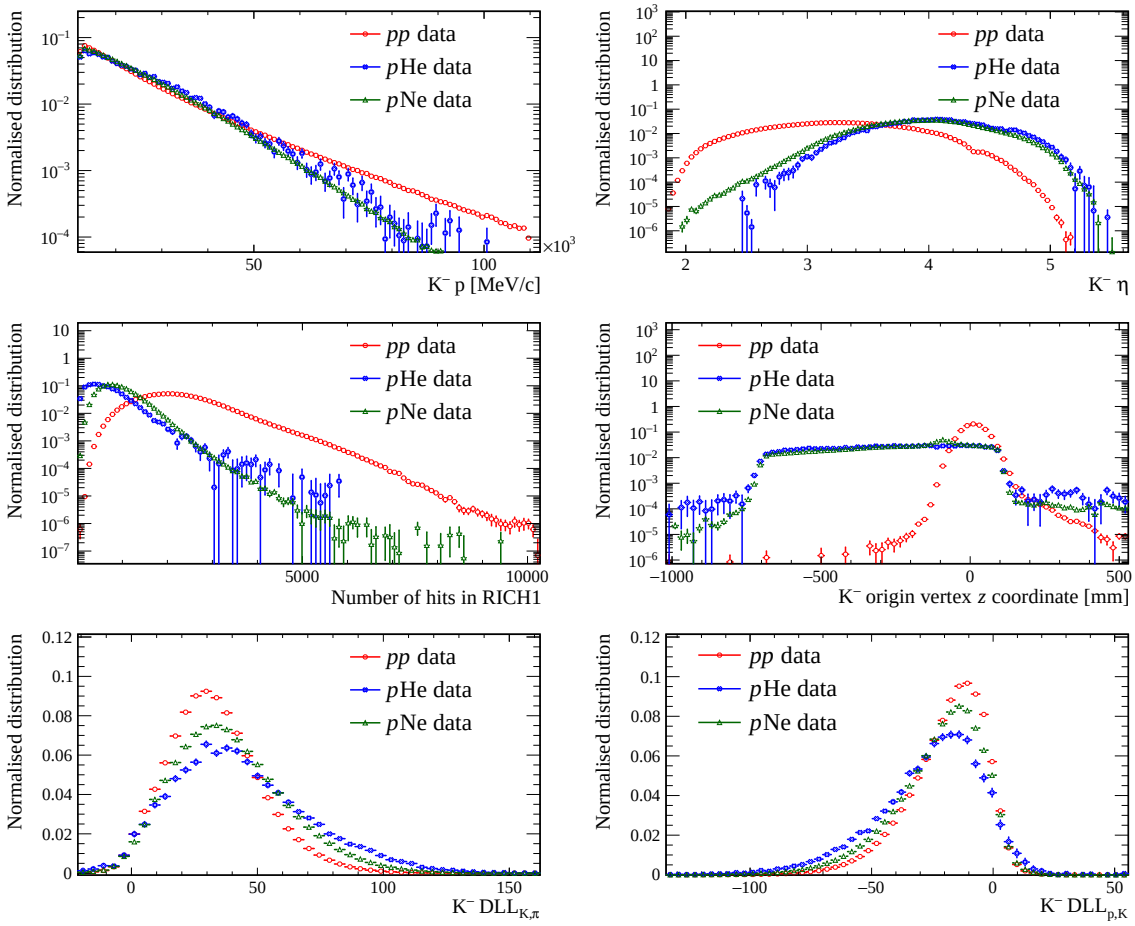


Figure III.1.3: Normalised distributions for the negatively-charged kaon momentum (top left), pseudorapidity (top right), number of hits in the RICH1 detector (middle left), origin vertex z coordinate (middle right), $\mathrm{DLL}_{K,\pi}$ (bottom left) and $\mathrm{DLL}_{p,K}$ (bottom right) in $\phi \to K^-K^+$ decays reconstructed and selected in the 2016 pHe (in blue), 2017 pp (in red) and pNe (in green) data samples.

Table III.1.1: PID calibration channels first selection stage in the 2017 pNe data.

Decay		Selection
$K^0_S \to \pi^+\pi^-$	π^+, π^- :	$p > 2\,\mathrm{GeV}/c$, track $\chi^2/\mathrm{ndf} < 5$, $\chi^2_{\mathrm{IP}} > 25$
	comb. :	$M < 1\,\mathrm{GeV}/c^2$, DV $\chi^2 < 16$
	K^0_S :	$\lvert M - M(K^0_S)\rvert < 50\,\mathrm{MeV}/c^2$, $\mathrm{DV}_z < 2200\,\mathrm{mm}$, $\bar{\Lambda}$ veto
$\bar{\Lambda} \to \bar{p}\pi^+$	$\bar{p}$, π^+ :	$p > 2\,\mathrm{GeV}/c$, track $\chi^2/\mathrm{ndf} < 5$, $\chi^2_{\mathrm{IP}} > 25$
	comb. :	$M < 1.5\,\mathrm{GeV}/c^2$, DV $\chi^2 < 16$
	$\bar{\Lambda}$:	$\lvert M - M(\bar{\Lambda})\rvert < 25\,\mathrm{MeV}/c^2$, $\mathrm{DV}_z < 2200\,\mathrm{mm}$, K^0_S veto
$\phi \to K^+K^-$	K^+, K^-:	$p_{\mathrm{T}} > 380\,\mathrm{MeV}/c$, track $\chi^2/\mathrm{ndf} < 5$
	K^+ :	$GhostProb < 0.025$, K probability > 0.75
	comb. :	$\lvert M - M(\phi)\rvert < 40\,\mathrm{MeV}/c^2$, DV $\chi^2 < 16$
	ϕ :	$\lvert M - M(\phi)\rvert < 20\,\mathrm{MeV}/c^2$, $\mathrm{DV}_z < 2200\,\mathrm{mm}$

Table III.1.2: PID calibration channels second selection stage in the 2017 pNe data.

Decay		Selection
Event-level		$N_{\rm trk}(\eta < 0) = 0$
$K^0_{\rm S} \to \pi^+\pi^-$	π^- :	$p > 12\,\text{GeV}/c$, $p_{\rm T} > 400\,\text{MeV}/c$, RICH signals
	π^- :	beam POCA $z \in [-700, 100]$ mm
	Arm.:	$\left\lvert \left(\frac{p^{\pi}_{Transv}}{206\,\text{MeV}/c}\right)^2 + \left(\frac{\alpha}{0.83}\right)^2 - 1 \right\rvert < 0.1$
	$K^0_{\rm S}$:	$M \in [450, 540]\,\text{MeV}/c^2$
$\overline{\Lambda} \to \overline{p}\pi^-$	$\overline{p}$:	$p > 12\,\text{GeV}/c$, $p_{\rm T} > 400\,\text{MeV}/c$, RICH signals
	$\overline{p}$:	beam POCA $z \in [-700, 100]$ mm
	Arm.:	$\left\lvert \left(\frac{p^{p}_{Transv}}{101\,\text{MeV}/c}\right)^2 + \left(\frac{\alpha - 0.69}{0.18}\right)^2 - 1 \right\rvert < 0.1$
	$\overline{\Lambda}$:	$M \in [1100, 1150]\,\text{MeV}/c^2$
$\phi \to K^+K^-$	K^- :	$p > 12\,\text{GeV}/c$, $p_{\rm T} > 400\,\text{MeV}/c$, RICH signals
	K^- :	beam POCA $z \in [-700, 100]$ mm
	ϕ :	$M \in [1010, 1028]\,\text{MeV}/c^2$

selection is performed in two steps, summarised in Tabs. III.1.1 and III.1.2. Particles reconstructed combining the information from all LHCb tracking detectors and with a track fit $\chi^2/\text{ndf} < 5$ are associated to the final-state hypotheses and accepted when satisfying loose kinematic thresholds. For the $\overline{\Lambda}$ and $K^0_{\rm S}$ decays, the final-state particles detachment with respect to the pNe collision vertex is imposed requiring a minimum $\chi^2_{\rm IP}$ of 25. The tracks are then paired to form the two-body combinations and the invariant mass is evaluated with Eq. I.3.5. By selecting an interval around the parent-particle nominal mass, most of the random track pairs are removed. The combination vertex is also required to be reconstructed with low χ^2 and to be located upstream the LHCb magnet. Similarly to Sec. II.2.3, the contamination between the $\overline{\Lambda}$ and $K^0_{\rm S}$ calibration channels, originating from the antiproton misidentification as a pion and vice versa, is removed. The two-track combinations invariant mass is re-evaluated under the negatively-charged final-state correct mass hypothesis and a region around the $K^0_{\rm S}$ or $\overline{\Lambda}$ nominal mass values is vetoed. For the ϕ calibration channel, given the larger background contamination, one of the two kaons (the positively-charged one in Tab. III.1.1 since we are considering negatively-charged probes) is selected with tighter criteria to increase the fraction of signal candidates. A possible systematic effect due to the correlation between the two kaons, resulting from the low mass difference between the parent and the sum of the two final-state particles, is discussed in Sec. III.1.4.3. The tagged kaon is required to present a low probability to be reconstructed from spurious energy deposits (*GhostProb*) and a high one to be recognised as a kaon. The mass peaks for the three calibration channels after the first selection stage are shown in the left column of Fig. III.1.4.

A more detailed description of the decays is then performed to reduce their background contamination. All events with one or more tracks with negative pseudorapidity are removed in order to mitigate the contamination from the unexpected pp collisions

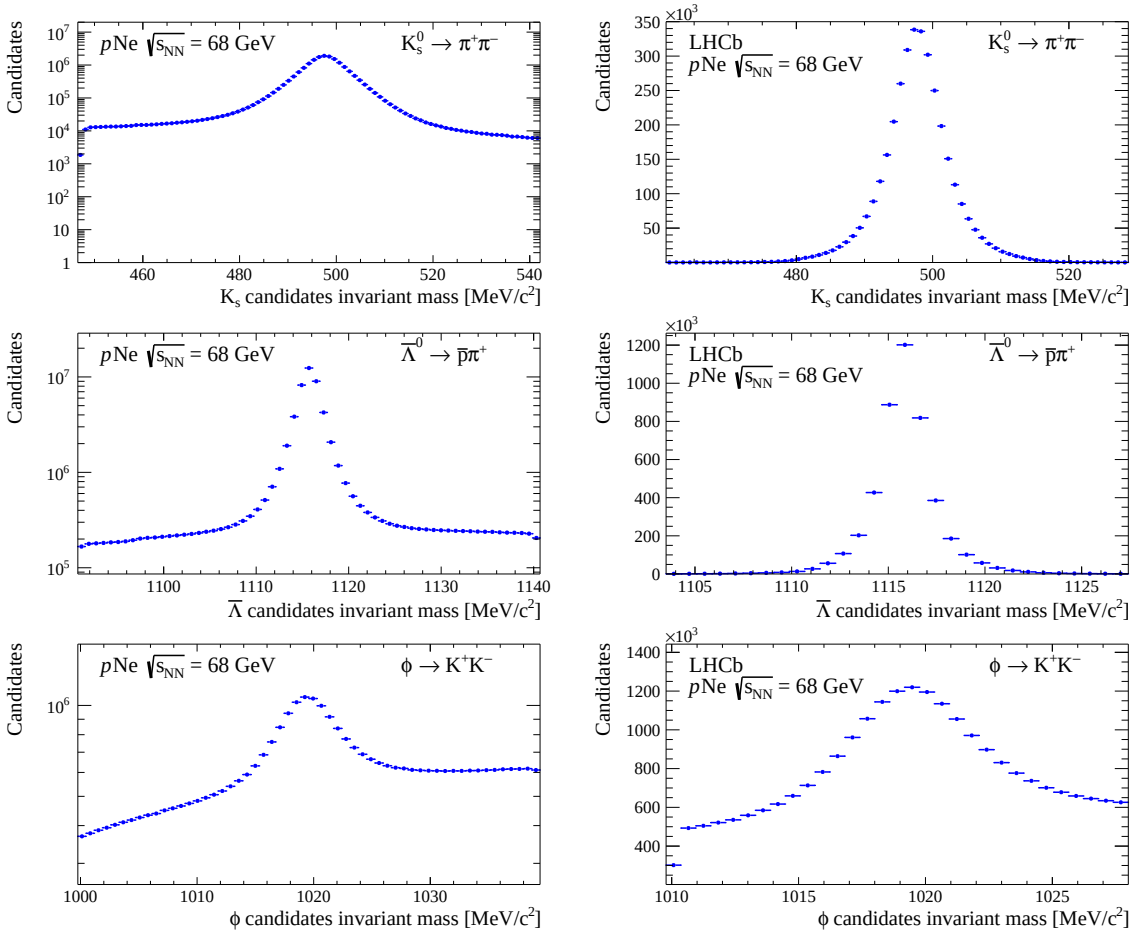


Figure III.1.4: Invariant mass distributions for the K^0_S (top), $\overline{\Lambda}$ (middle) and ϕ (bottom) calibration channels in the 2017 pNe data after the first (left) and second (right) selection steps.

originating from the protons bunch-by-bunch migrations, as discussed in Sec. I.4.2. The extent of this effect is exemplified by the left plot of Fig. III.1.5, showing the PV_z distribution for $\overline{\Lambda} \to \bar{p}\pi^+$ candidates selected in the pNe data. Over the continuous distribution due to the pNe collisions, an evident peak centred around the nominal collision point is observed. For these candidates, as illustrated in the right plot, the number of tracks with negative pseudorapidity averages to larger values than the upstream region and their contribution is thus suppressed requiring no backward tracks in the event. The residual background, considering that the presented approach learns the PID dependence on all the variables and takes into account their correlations, is not expected to affect the results but only to decrease the overlap between the calibration and application phase-spaces. The mass windows and the kinematic thresholds are tightened with respect to the first selection, with the minimum momentum and transverse momentum set to 12 GeV/c and 400 MeV/c, respectively. All probe particles are imposed to be associated with signals in the RICH system. For the K^0_S and $\Lambda/\overline{\Lambda}$ calibration channels, the discrimination between the signal and the background is achieved with the analysis in the Armenteros-Podolanski plot [131], defined as in Eq. II.1.26. The results are presented for the K^0_S (left) and the $\Lambda/\overline{\Lambda}$ (right) decays in Fig. III.1.6. The blue markers indicate all candidates selected by the requirements of the first stage and a

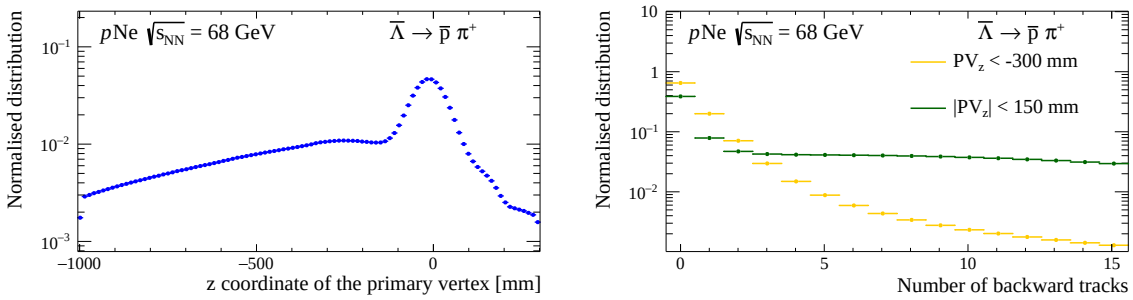


Figure III.1.5: PV_z distribution (left) for $\overline{\Lambda} \to \overline{p}\pi^+$ candidates in the 2017 pNe data sample after the first selection. The right plot compares the normalised distributions for the number of tracks with negative pseudorapidity in two different PV_z regions. The found contamination from pp collisions can be reduced by imposing $N_{trk}(\eta < 0) = 0$.

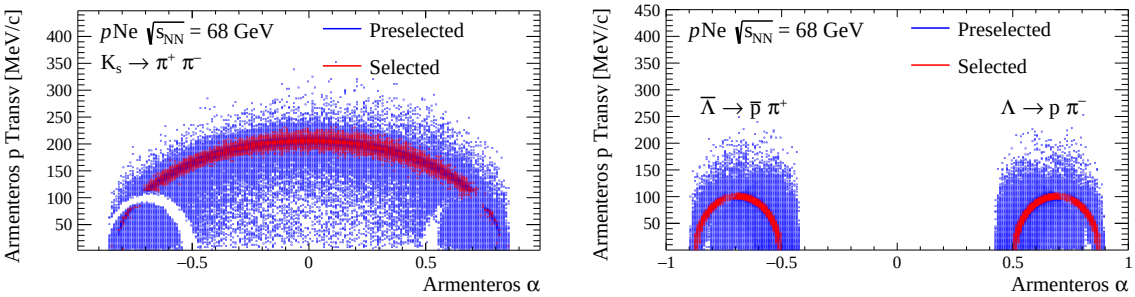


Figure III.1.6: Armenteros-Podolanski plot for the $K^0_S \to \pi^+\pi^-$ (left) and the $\Lambda \to p\pi^-$ and $\overline{\Lambda} \to \overline{p}\pi^+$ (right) decays after their first (in blue) and the second (in red) selection step. Clear elliptic patterns around the expected theoretical profiles are selected, achieving a larger purity than 99%.

smooth distribution is found. The applied Armenteros-Podolanski selection, indicated in Tab. III.1.2, only accepts the red points and a larger purity than 99% is reached. For the ϕ decay, the statistical subtraction of the background is performed via the *sPlot* technique. The invariant mass distribution for the ϕ candidates is modelled with the sum of a Voigtian distribution, the convolution of a Gaussian and a Breit-Wigner, for the signal and a first-order polynomial function for the background, as shown in the left plot of Fig. III.1.7. With the assumed parametric models, weights for each event to belong to the signal or to the background category are evaluated and validated in the right plot of Fig. III.1.7. This compares the normalised distributions of a combination of the K^- $\mathrm{DLL}_{p,\pi}$ and $\mathrm{DLL}_{p,K}$ variables where the pion, kaon and proton species can be distinguished before (in blue) and after (in red) the application of the *sPlot* weights for the signal hypothesis. The dominant left-most peak, corresponding to the pion background contamination, is suppressed. The mass peaks for the three calibration channels after the second selection stage are presented in the right column of Fig. III.1.4 and, when comparing to the left one, a clear enhancement of the signal is obtained.

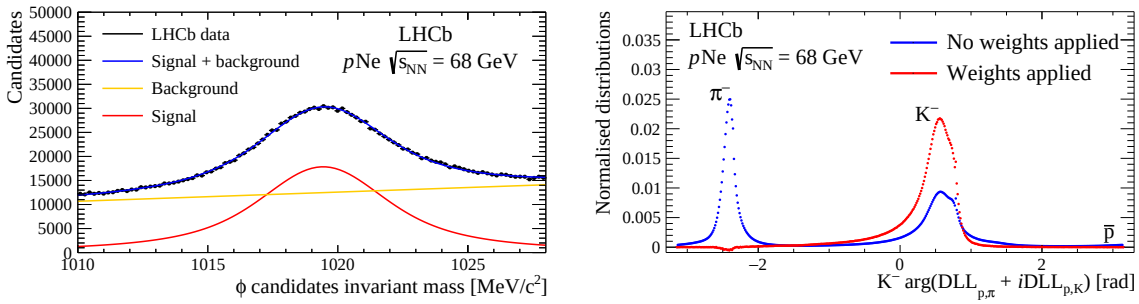


Figure III.1.7: Application of the *sPlot* technique to the $\phi \to K^-K^+$ decays selected in the 2017 pNe sample. The left plot shows the modelling of the ϕ candidates invariant mass with the sum of a Voigtian, the convolution of a Gaussian and a Breit-Wigner, for the signal and a first-order polynomial function for the background. The right plot validates the calculated weights for the signal hypothesis by comparing the normalised distributions of $\arg(\mathrm{DLL}_{p,\pi} + i\mathrm{DLL}_{p,K})$ before (in blue) and after (in red) their application. The left-most peak, corresponding to pions, is largely suppressed.

Table III.1.3: List of the feature candidates in descending importance according to the KS ordering parameter, measured as explained in the text. The variables highlighted in bold are included in the trainings of the models.

Variable	Max KS	Variable	Max KS	Variable	Max KS
$\boldsymbol{p}$	0.64	p_z	0.64	$\boldsymbol{\eta}$	0.54
p_T	0.51	$\boldsymbol{yz}$ **slope**	0.38	**track ndf**	0.34
$\boldsymbol{xz}$ **slope**	0.34	**nTracks**	0.34	$\mathbf{N_{RICH2}}$ **hits**	0.33
N_{SPD} hits	0.32	$\mathbf{N_{RICH1}}$ **hits**	0.28	**track** $\boldsymbol{\chi^2}$**/ndf**	0.26

III.1.3.2 Feature selection

The set of features affecting most the response of the RICH detectors need to be determined. This is a crucial point of the approach, provided that enough quantities must be considered to ensure a complete description of the RICH response, but also that is desirable to keep the complexity of the model low enough to ensure a fast tuning of the hyperparameters. To order the variables, a method based on the Kolmogorov-Smirnov (KS) distance is applied. For each feature candidate, the $\mathrm{DLL}_{p,\pi}$ and $\mathrm{DLL}_{p,K}$ variables for pions in the pHe data, selected with $\arg(\mathrm{DLL}_{K,\pi} + i\,\mathrm{DLL}_{p,\pi}) < -1$, are compared in equally-populated intervals. The maximum measured KS distance between all possible histogram pairs is considered as an ordering parameter: the larger it is, the more important is the feature candidate in determining the DLL variables. Figure III.1.8 shows two examples of the procedure with the particle pseudorapidity (top) and the y coordinate of the primary vertex (bottom). The projections of the bidimensional distribution onto the $\mathrm{DLL}_{p,\pi}$ (left) and $\mathrm{DLL}_{p,K}$ (right) axes in intervals of the respective feature candidate are superimposed and, as expected and confirmed by the measured KS distance, a larger dependence on the pseudorapidity is found. By iterating over a set of possible feature candidates, Tab. III.1.3 lists the KS results in descending order. The leading features reflect:

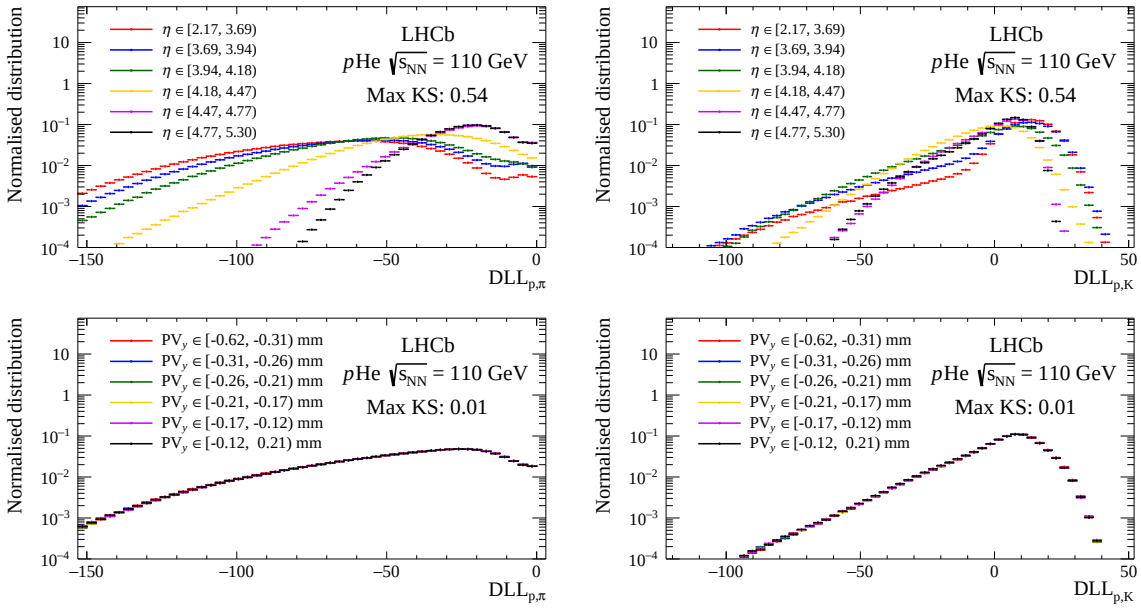


Figure III.1.8: Distributions of the $\mathrm{DLL}_{p,\pi}$ (left) and $\mathrm{DLL}_{p,K}$ (right) variables for pions in the pHe sample in intervals of their pseudorapidity (top) or y coordinate of the primary vertex (bottom). The maximum Kolmogorov-Smirnov distance values are indicated to exemplify the ordering procedure explained in the text.

- the particle kinematics, such as for the momentum p, its longitudinal (p_z) and transverse (p_T) components and the pseudorapidity η;
- the occupancy in the detector, expressed as the number of hits in the SPD, RICH1 and RICH2 detectors or as the number of reconstructed tracks;
- the quality of the track reconstruction, exemplified by the χ^2/ndf or the number of degree of freedom (ndf) of its fit;
- the geometry of the event, as for the track slopes with respect to the z axis.

All variables listed in Tab. III.1.3, when not largely or analytically correlated to others (as highlighted in bold), are included in the model. For example, the number of hits in the RICH1 detector, despite the lower KS value with respect to N_{SPD}, is considered since casually connected with the PID. The transverse momentum, which can be obtained from the momentum and the pseudorapidity, is discarded instead. Variables relating the different geometry of the training and application particles are added. Indeed, while the models are applied to prompt pions, kaons and antiprotons, whose OV_z distribution coincides by definition with the selected PV_z range ($\mathrm{PV}_z \in [-700, 100]$ mm for pHe, $\mathrm{PV}_z \in [-200, 100]$ mm for pAr), the training final-state particles originate in a decay. To account for the different distributions of the OV coordinates, the tracks associated to the training particles are extrapolated to the beam and the coordinates of the intersection (called beam POCA in the following) are added to the set of features. Also, to only train on the interesting region for the application, the requirement beam POCA $z \in [-700, 100]$ mm is added for all calibration channels, as indicated in Tab. III.1.2.

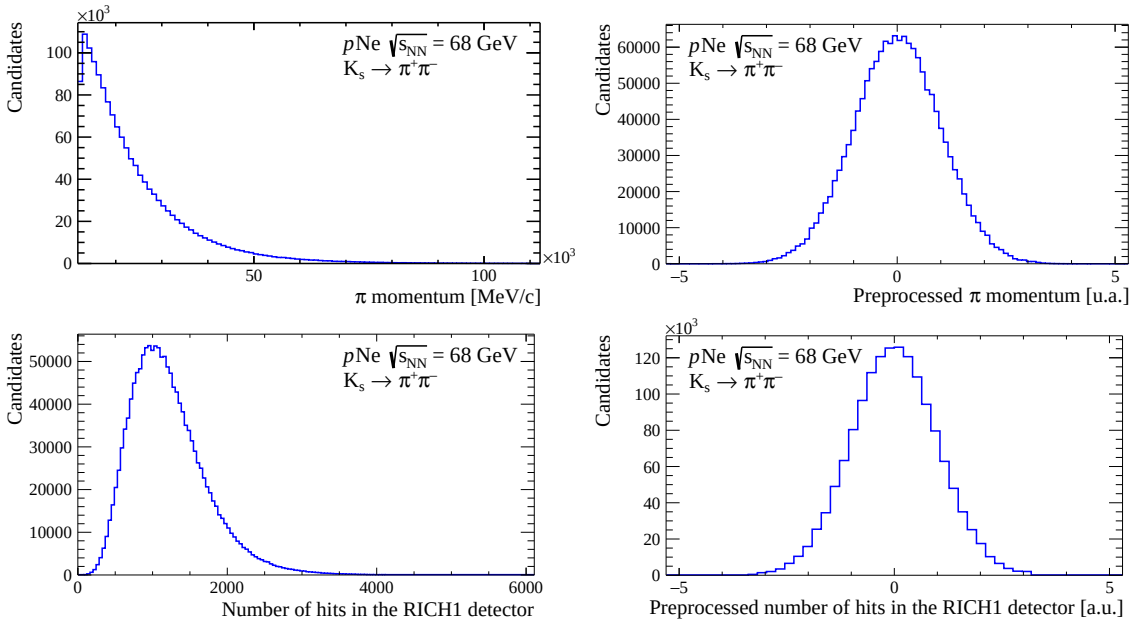


Figure III.1.9: Momentum (top) and number of hits in the RICH1 detector (bottom) distributions before (left) and after (right) the application of the QuantileTransformer preprocessing algorithm for the negative pions in reconstructed and selected $K^0_S \to \pi^+\pi^-$ decays in the 2017 pNe data.

Table III.1.4: Input user-defined parameters to train one model for each calibration channel.

Input parameter	$K^0_S \to \pi^-\pi^+$	$\overline{\Lambda} \to \overline{p}\pi^+$	$\phi \to K^-K^+$
Number of Gaussians	64	20	64
Number of NN nodes	128	128	128
Starting learning rate	10^{-3}	10^{-4}	$5 \cdot 10^{-6}$
Batch size [events]	10000	10000	20000

III.1.3.3 Model preprocessing and training

Once identified the set of features affecting the RICH detector responses, one model is trained for each calibration sample. As mentioned above, a preprocessing of all variables is firstly performed. The results for the π^- particle momentum (top) and for the number of hits in the RICH1 detector (bottom) for reconstructed and selected $K^0_S \to \pi^+\pi^-$ decays are shown in Fig. III.1.9. Despite the different input functional form, both features are mapped to normal distributions. The GMM is then built with the input parameters for the three calibration channels indicated in Tab. III.1.4. A larger complexity is required for the K^0_S and ϕ decays given that, as a consequence of the threshold effects involved in the classifier definition, multimodal PID distributions are found. For the ϕ calibration channel, because of the application of the *sPlot* weights w_i, only a fraction of the events induces a significant variation of the loss function. A larger batch size is hence set in order to counterbalance the larger statistical uncertainty.
The parameters of the GMM are randomly initialised according to uniform distributions in the ranges:

Table III.1.5: Number of parameters to train one model for each considered calibration channel.

	$K^0_S \to \pi^-\pi^+$	$\overline{\Lambda} \to \overline{p}\pi^+$	$\phi \to K^-K^+$
Training events $[\cdot 10^6]$	3.5	4.6	11 (after *sPlot*)
Training parameters	122112	87792	122112

- $\left[\langle \underline{x} \rangle - \frac{1}{2}\sqrt{\langle \underline{x}^2 \rangle - \langle \underline{x} \rangle^2}, \langle \underline{x} \rangle + \frac{1}{2}\sqrt{\langle \underline{x}^2 \rangle - \langle \underline{x} \rangle^2}\right]$ for the mean values $\underline{\mu}$, indicating by $\langle \cdot \rangle$ the average. For the ϕ calibration channel, the background-subtracted distribution is considered;
- $\left[\frac{1}{10}\sqrt{\langle \underline{x}^2 \rangle - \langle \underline{x} \rangle^2}, \sqrt{\langle \underline{x}^2 \rangle - \langle \underline{x} \rangle^2}\right]$ for the widths $\sigma_{1,2}$;
- $[0, 2\pi]$ for the correlation angle, defined as $\omega = arcsin(\rho)$;
- $[0, 1]$ for the component weights α_j.

The first training stage is performed with this initialization only. Then, for each parameter of the GMM, the output of a NN fed with the feature values and built with the parameters in Tab. III.1.4 is added. The number of training parameters is indicated in Tab. III.1.5 for the three calibration decays and is $\mathcal{O}$ (10^5), depending on the chosen complexity for the NN and the GMM. For the K^0_S calibration line, with the training sample size indicated in Tab. III.1.5 and the model chosen parameters in Tab. III.1.4, the corresponding training time is measured in approximately 10 hours on a NVIDIA K80 GPU. Also with different input configurations, the loss minimization procedure converges in few hours with both the CPU and GPU TensorFlow backends. The loss function evolution with the number of epoch is illustrated in Fig. III.1.10 for the pion (top left), proton (top right) and kaon (bottom) final-state particles. For all calibration channels, the loss firstly present a steep decrease, followed by a smoother one and then a gentle oscillation towards the minimum.

In contrast to other machine-learning problems, notably the classification ones, the overtraining is not a key aspect to monitor in this case considering that the model learns a pdf and is then applied to other samples than the training one. Yet, in corner regions of the hyper-space, statistical fluctuations could induce fast variations of the model parameters to reproduce single data entries or clusters. To exclude this effect, the evolution of the model parameters as a function of the feature variables is monitored and shown for example in Fig. III.1.11 for the $\overline{\Lambda} \to \overline{p}\pi^+$ calibration channel. The smooth evolution of all the curves, corresponding to the multinormal components in the GMM, indicates the regularity of the training. Also, the learning rate choice is validated by the absence of abrupt variations of the curves.

III.1.3.4 Model validation

For each calibration channel, the model has to learn how to reproduce the $\mathrm{DLL}_{p,\pi}$ and $\mathrm{DLL}_{p,K}$ variables in the whole considered multi-dimensional feature hyper-space. To verify this, all models are separately validated in equally-populated bins of all pairs of the considered features. In each interval, calibration events are randomly chosen and a $\mathrm{DLL}_{p,\pi}$-$\mathrm{DLL}_{p,K}$ template is generated according to their feature distribution. The model is

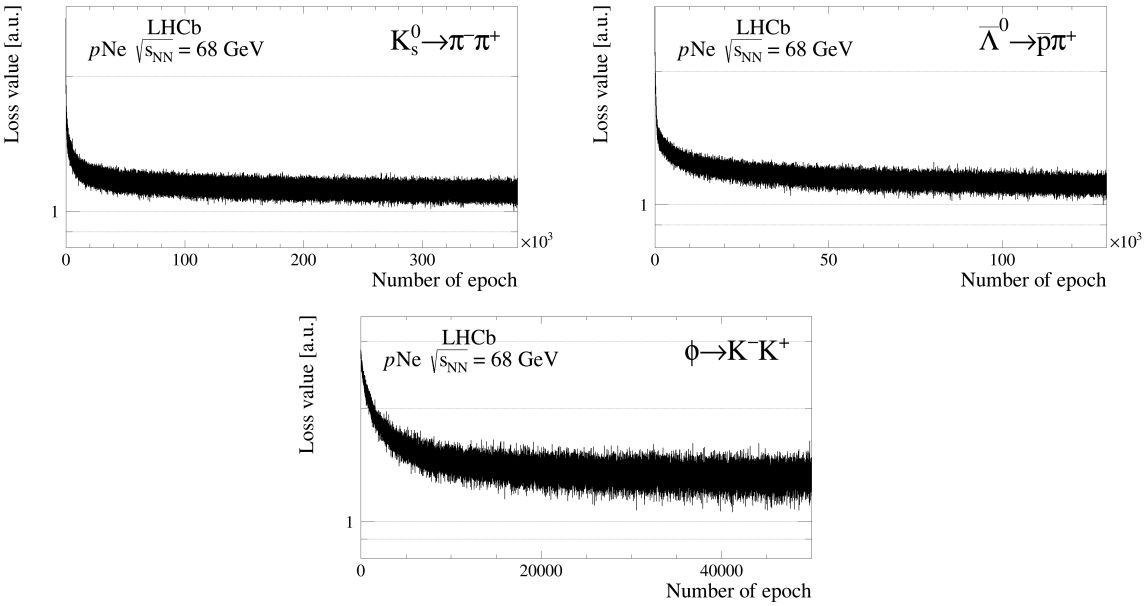


Figure III.1.10: Evolution of the loss function with the number of epoch for the training of the models for (top left) the pion, (top right) the proton and (bottom) the kaon calibration channels.

validated if the two distributions coincide in all the feature hyper-space regions. Examples of the comparison in the K^0_S, $\bar{\Lambda}$ and ϕ calibration channels considering different feature pairs are presented in Fig. III.1.12, III.1.13 and III.1.14, respectively. In all figures, the bidimensional comparison (first row) and the projections onto the $\text{DLL}_{p,\pi}$ and $\text{DLL}_{p,K}$ considering the full (second row) or intervals of the projected variable (third and fourth row) are shown and an excellent match between the data (in red) and the distributions generated by the trained models (in blue) is always found. Figure III.1.15 shows another example of the validation for the $K^0_S \to \pi^+\pi^-$ decay but in a feature pair interval presenting low calibration data statistics. The match between the two distributions also in this region demonstrates the model capability to draw a smooth template based on the available information, mitigating the limitation due to the scarce calibration data, and without significant overtraining.

III.1.4 Applications of the approach

Three applications of the proposed approach are discussed in the following. In Sec. III.1.4.1, the three trained samples are employed to produce templates for the PID classifiers on pHe and pAr data. Their combination is fitted to the data with the relative abundance as free parameters and the resulting quality of the fit is compared with that of a model based on the LHCb detailed simulation. In Sec. III.1.4.2, the PID response difference between prompt and detached antiprotons in pHe data is discussed to confirm the results obtained in Sec. II.1.6.4. Section III.1.4.3 studies a long-standing tension between two calibration samples used to model the particle identification response to kaon mesons in pp data.

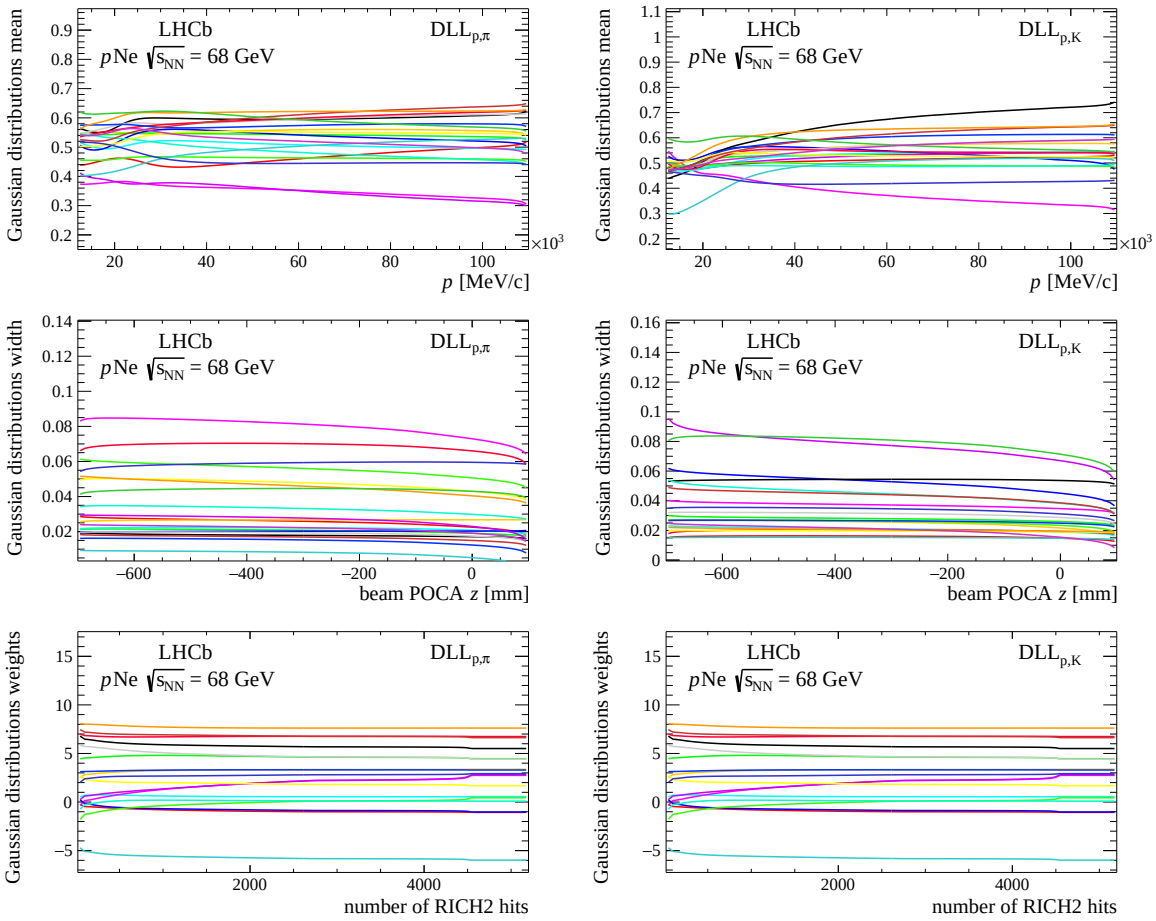


Figure III.1.11: Check against a possible overtraining effect for the $\overline{\Lambda} \to \overline{p}\pi^+$ calibration channel. The evolution of (top) the Gaussian mean values, (middle) widths and (bottom) weights for (left) the DLL$_{p,\pi}$ and (right) the DLL$_{p,K}$ variables are presented as a function of the antiproton momentum, beam POCA z and number of hits in the RICH2 detector, respectively. The different colours represent the Gaussian components included in the model.

III.1.4.1 PID model for *p*He and *p*Ar data

The models trained on the calibration channels reconstructed and selected in the 2017 pNe data are applied to lower-statistics SMOG samples, the 2016 pHe and the 2015 pAr with $\sqrt{s_{\rm NN}} = 110$ GeV. For the former, the machine-learning-based approach performance is compared to the strategy adopted in Ref. [61] and discussed in Sec. III.1.1. With the trained models, templates for the DLL$_{p,\pi}$ and DLL$_{p,K}$ classifiers are generated for the pion, kaon and antiproton particle species with the same procedure as the validation but based on the pHe features. The templates drawn in the weighted simulation and predicted by the data-based models are composed with the relative abundances as free parameters and fitted to the pHe data in intervals of the antiproton momentum and transverse momentum. In both cases, the initialization for the particle abundances and the template for ghost particles are taken from the simulation. Examples of the fit to the pHe data in randomly-chosen kinematic intervals with the simulated and predicted templates are presented in Figs. III.1.16 and III.1.17, respectively, while their direct comparison is presented in Fig. III.1.18. In all kinematic intervals, the predicted templates

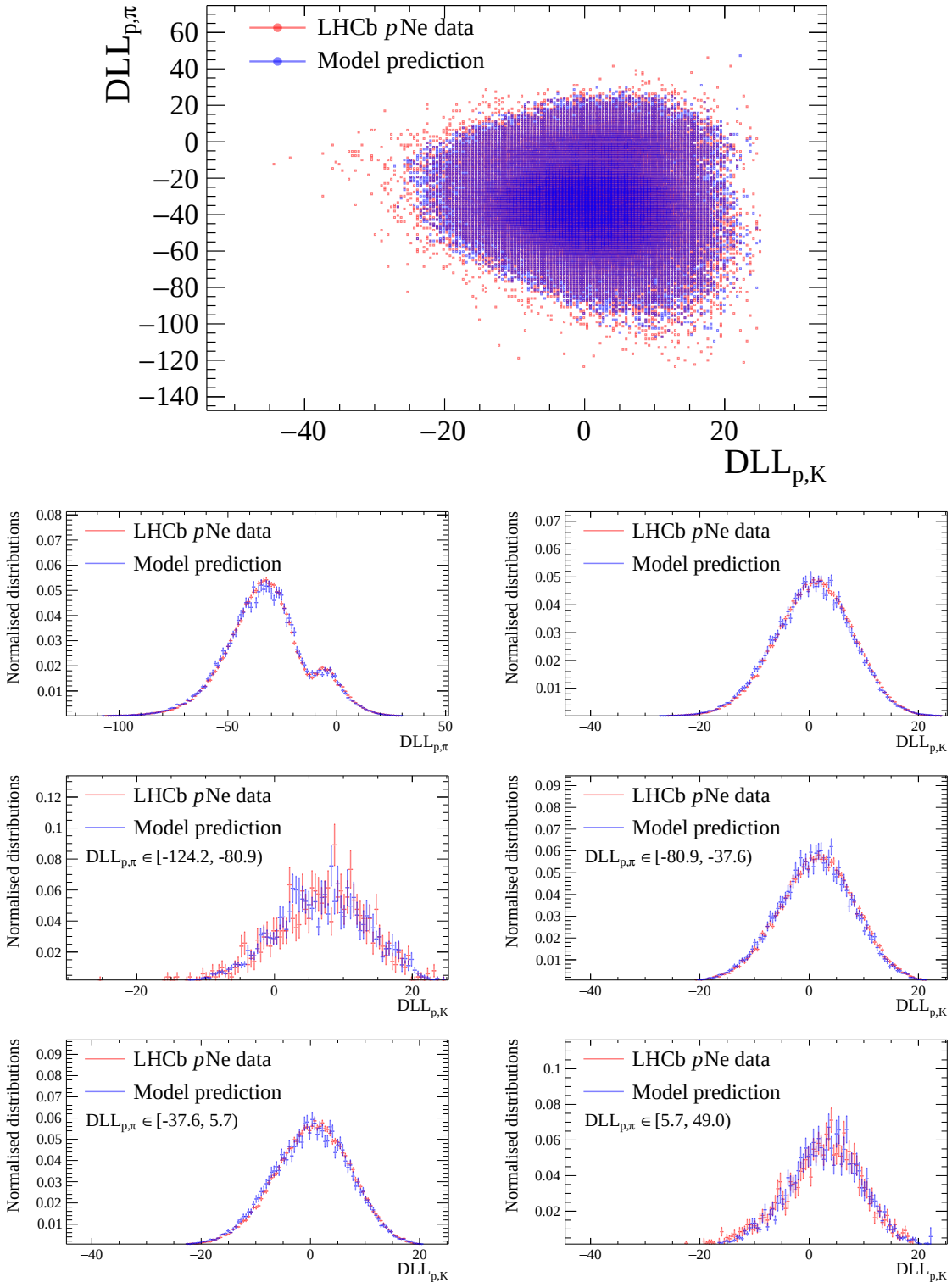


Figure III.1.12: Comparison for the $K^0_S \to \pi^-\pi^+$ channel in the $p \in [12.0, 15.5)\,\text{MeV}/c$ - $\eta \in [4.1, 4.4)$ region between the $\text{DLL}_{p,\pi}$-$\text{DLL}_{p,K}$ distributions in (red) the pNe data and (blue) predicted with the trained model (top plot), its projections onto the two axes (second row) and onto the $\text{DLL}_{p,K}$ axis in intervals of the $\text{DLL}_{p,\pi}$ variable (third and fourth rows).

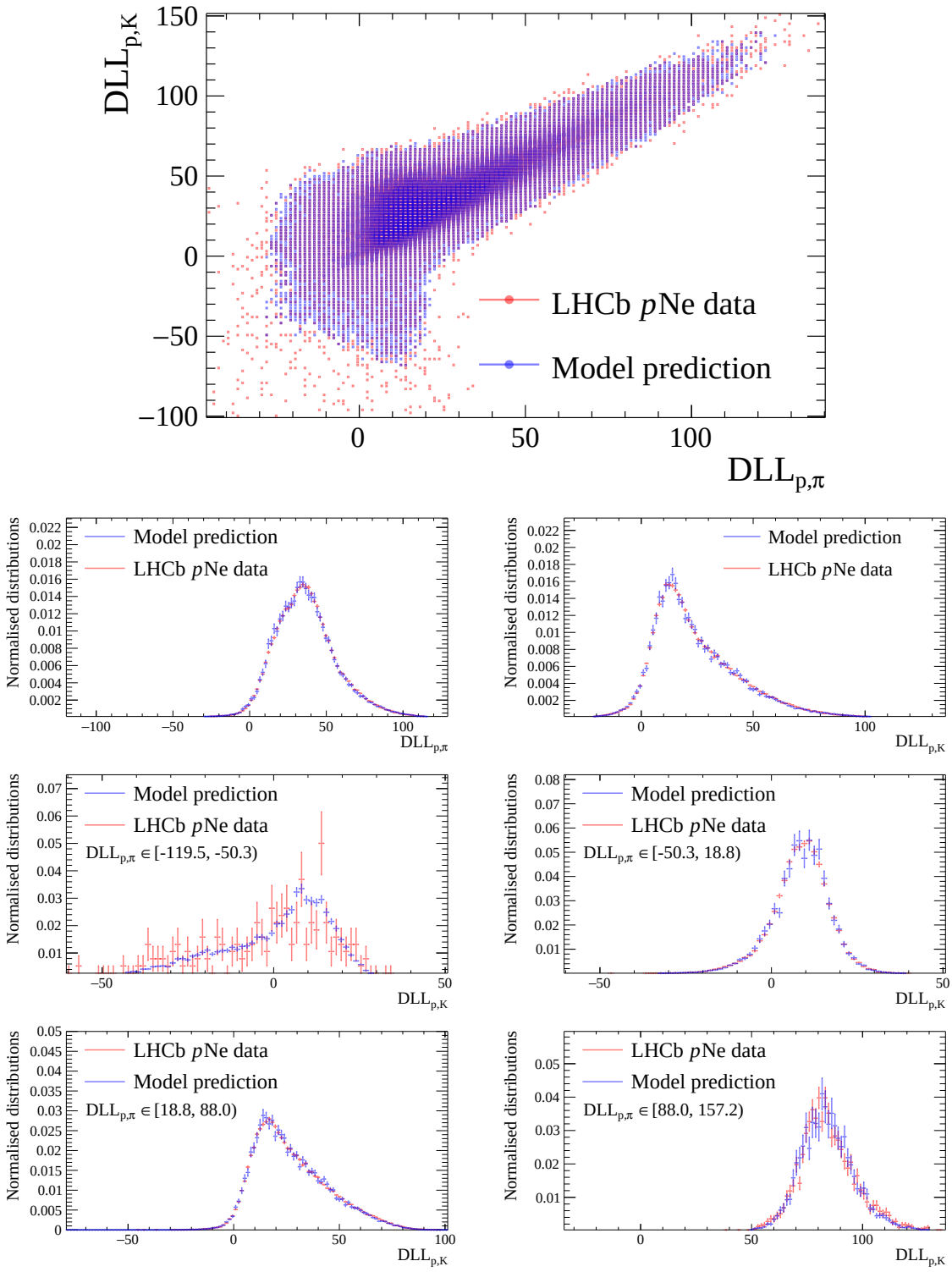


Figure III.1.13: Comparison for the $\overline{\Lambda} \to \overline{p}\pi^+$ channel in the beam POCA $z \in [-700, -430.8]$ mm - $N_{RICH1} \in [21, 688)$ region between the DLL$_{p,\pi}$-DLL$_{p,K}$ distributions in (red) the pNe data and (blue) predicted with the trained model (top plot), its projections onto the two axes (second row) and onto the DLL$_{p,K}$ axis in intervals of the DLL$_{p,\pi}$ variable (third and fourth rows).

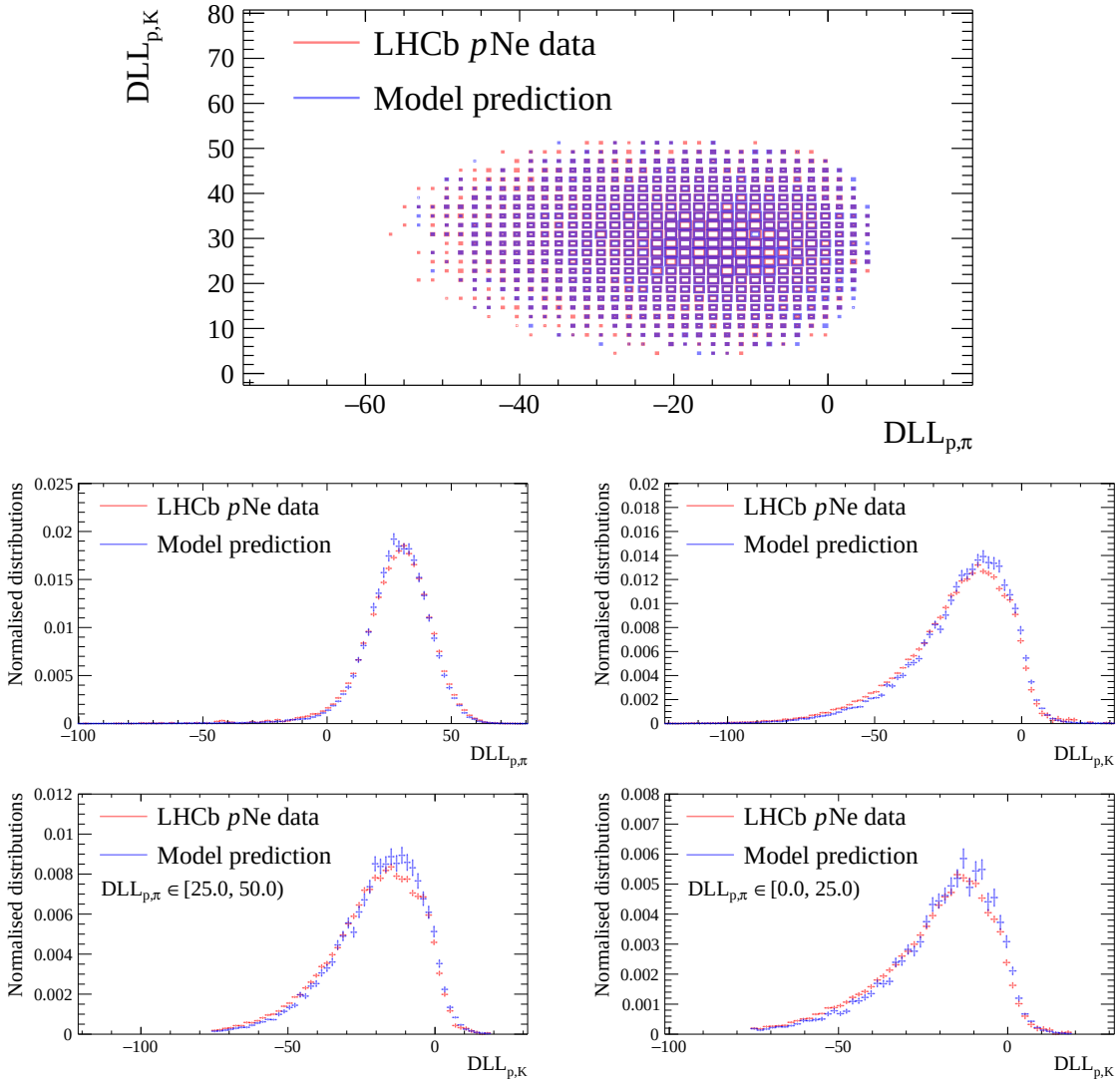


Figure III.1.14: Comparison for the $\phi \to K^-K^+$ channel in the track fit $\chi^2/\text{ndf} \in [1.5, 4)$ - $\eta \in [1.8, 3.7)$ region between the $\text{DLL}_{p,\pi}$-$\text{DLL}_{p,K}$ distributions in (red) the pNe data and (blue) predicted with the trained model (top plot), its projections onto the two axes (second row) and onto the $\text{DLL}_{p,K}$ axis in intervals of the $\text{DLL}_{p,\pi}$ variable (third row).

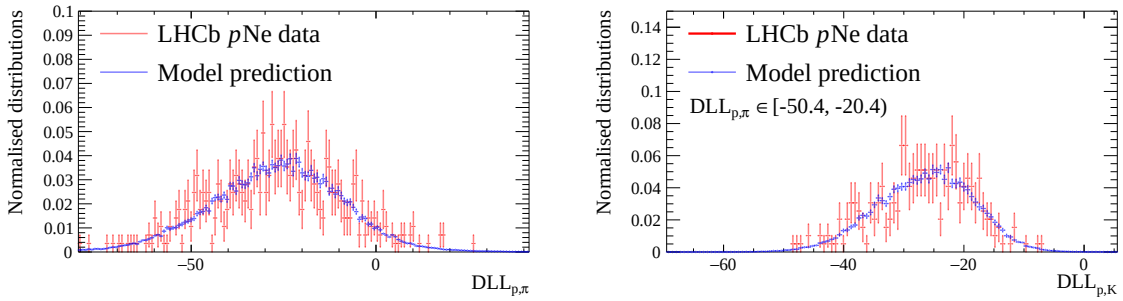


Figure III.1.15: Model validation for the $K^0_S \to \pi^-\pi^+$ calibration channel in the $p \in [70, 110)$ GeV/c - $\eta \in [3, 4)$ region. Despite the low calibration statistics (in red), a smooth template is generated by the trained model (in blue) with the available information.

offer an excellent description of the pHe data, equal or better than the model based on the LHCb detailed simulation.
To verify that the simulation-based PID model is not limited by the weighting procedure applied to the pHe data, a second application is performed on the 2015 pAr sample. In this case, a more sophisticated weighting of the simulation employing a Boosted Decision Tree (BDT) algorithm [144] is applied to efficiently take into account more variables than the pHe sample and their correlations. Templates are generated for the pAr sample and combined to be fitted. The resulting comparison of the fit with simulated (in purple) or data-based (in blue) templates, shown in Fig. III.1.19, still favours the latter.
In order to quantitatively measure the improvement in the DLL variables description achieved by the data-based model, the KS distance between the data PID distribution and the simulated or predicted templates is measured on both samples. Figure III.1.20 shows the resulting KS difference between the simulation and the data-based approach for the pHe (top) and the pAr (bottom) samples. Most intervals in the considered kinematic phase-space show a positive value, indicating a better performance of the machine-learning approach in describing the data.

III.1.4.2 PID for prompt and detached antiprotons in pHe

With the benchmark application presented in the previous section, the machine-learning-based model is demonstrated to offer a better description of the PID classifiers than the LHCb detailed simulation. For fixed-target analyses, suffering from the scarce statistics of calibration data, it is expected to be useful in a large variety of use-cases. In particular, for the first periods of the upgraded fixed-target system or in future short special runs such as the proposed one with oxygen-oxygen or proton-oxygen collisions [145], the low statistics of calibration channels is expected to affect the precision of the PID studies. Considering that the new system will provide a luminosity measurement with an expected $1-2\%$ uncertainty, PID is expected to be the leading uncertainty.
For all analyses counting a particle species with a template fit as in the measurement of the prompt antiproton production in pHe collisions, the discussed machine-learning methodology can generate more precise templates than the simulation, improving the quality of the fit to data. The efficiency of an applied PID selection requirement, typically measured on simulation, can be verified with the model. As an example, in the non-prompt antiproton production in pHe collisions measurement, the different PID efficiency for detached and prompt antiprotons is assigned as a relative 0.9% contribution to the systematic uncertainty. The estimation of the contribution relies on the simulation of the RICH detector response, which can be checked with this data-based model. Since we only consider antiprotons for this application, the $\overline{\Lambda}$ decay in pNe data is used as a calibration channel. The comparison between the normalised distributions for all features included in the model in the pNe data (in red) and for simulated prompt (in orange) and detached (in green) antiprotons in pHe is shown in Figs. III.1.21 and III.1.22. Some differences, especially in the detector occupancy and in the event geometry, arise. With the same parameters as in Tab. III.1.5, a data-based model is trained on the $\overline{\Lambda}$ pNe calibration channel and used to predict the $\mathrm{DLL}_{p,\pi}$ and $\mathrm{DLL}_{p,K}$ variables for prompt and detached antiprotons in the pHe data, according to the respective feature distributions. The PID

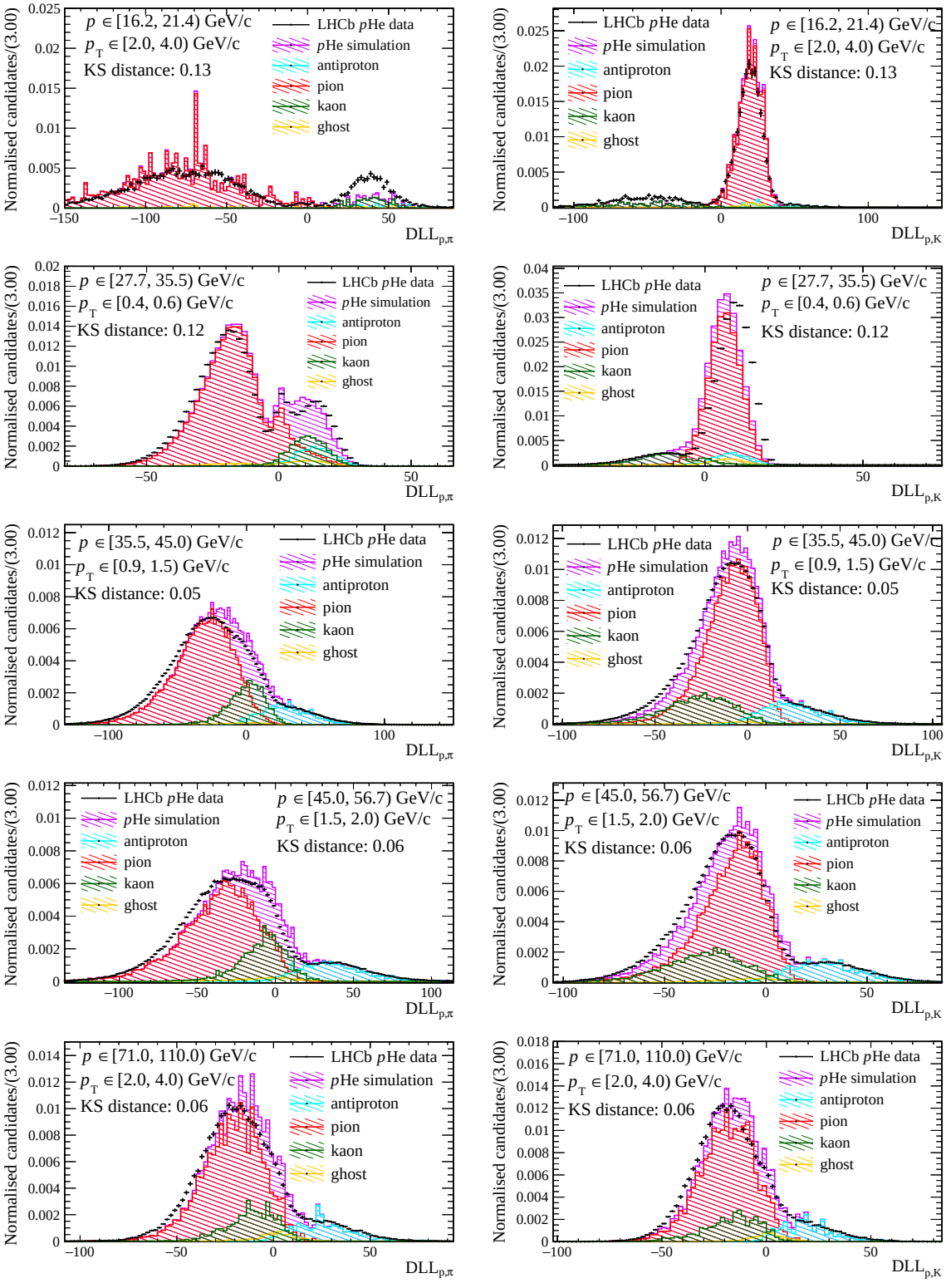


Figure III.1.16: Projections onto the (left) $\mathrm{DLL}_{p,\pi}$ and (right) $\mathrm{DLL}_{p,K}$ axes of the fit to pHe data employing simulation-based templates in five momentum, transverse momentum intervals.

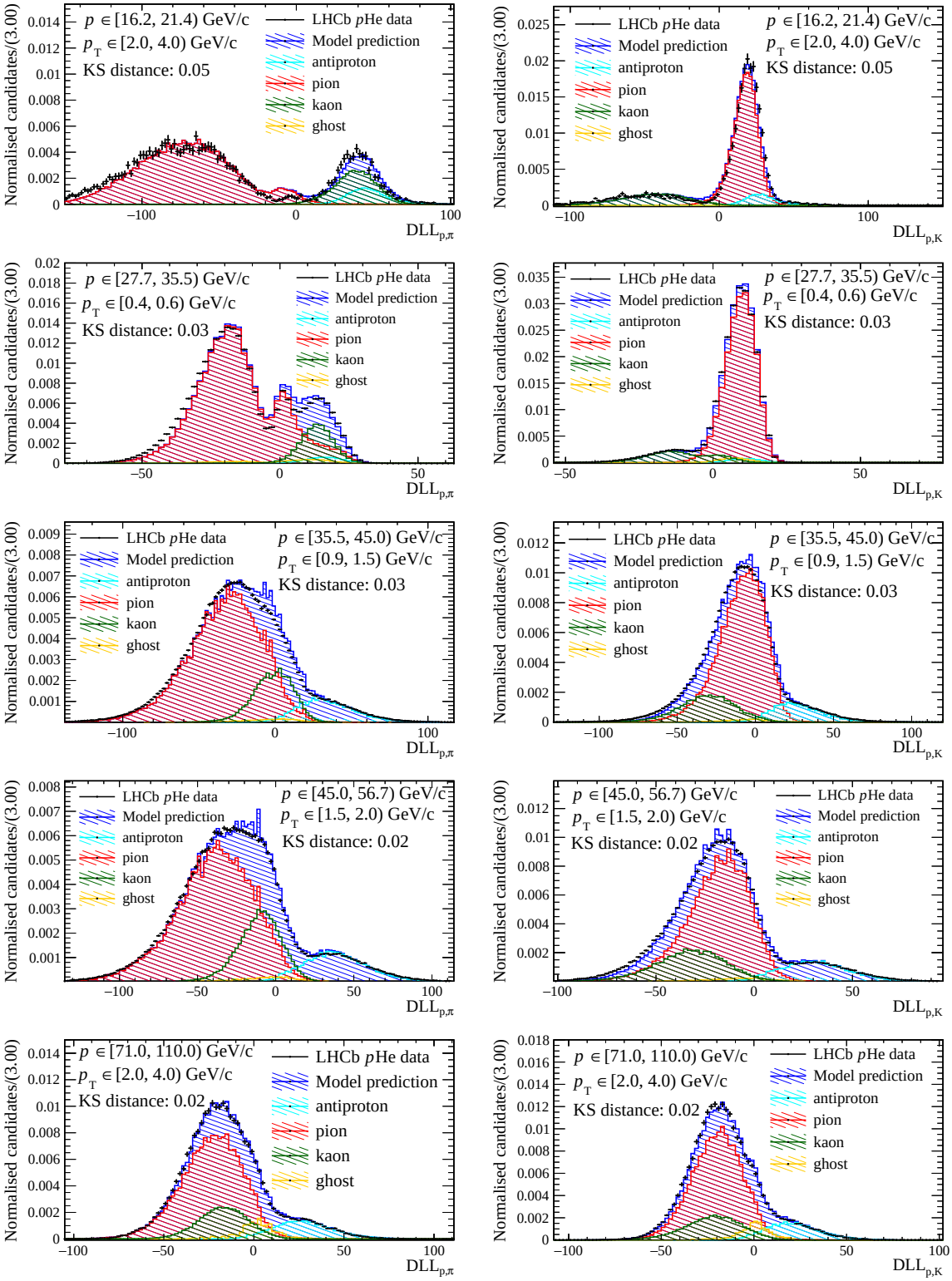


Figure III.1.17: Projections onto the (left) $\mathrm{DLL}_{p,\pi}$ and (right) $\mathrm{DLL}_{p,K}$ axes of the fit to pHe data employing the data-driven templates in five momentum, transverse momentum intervals.

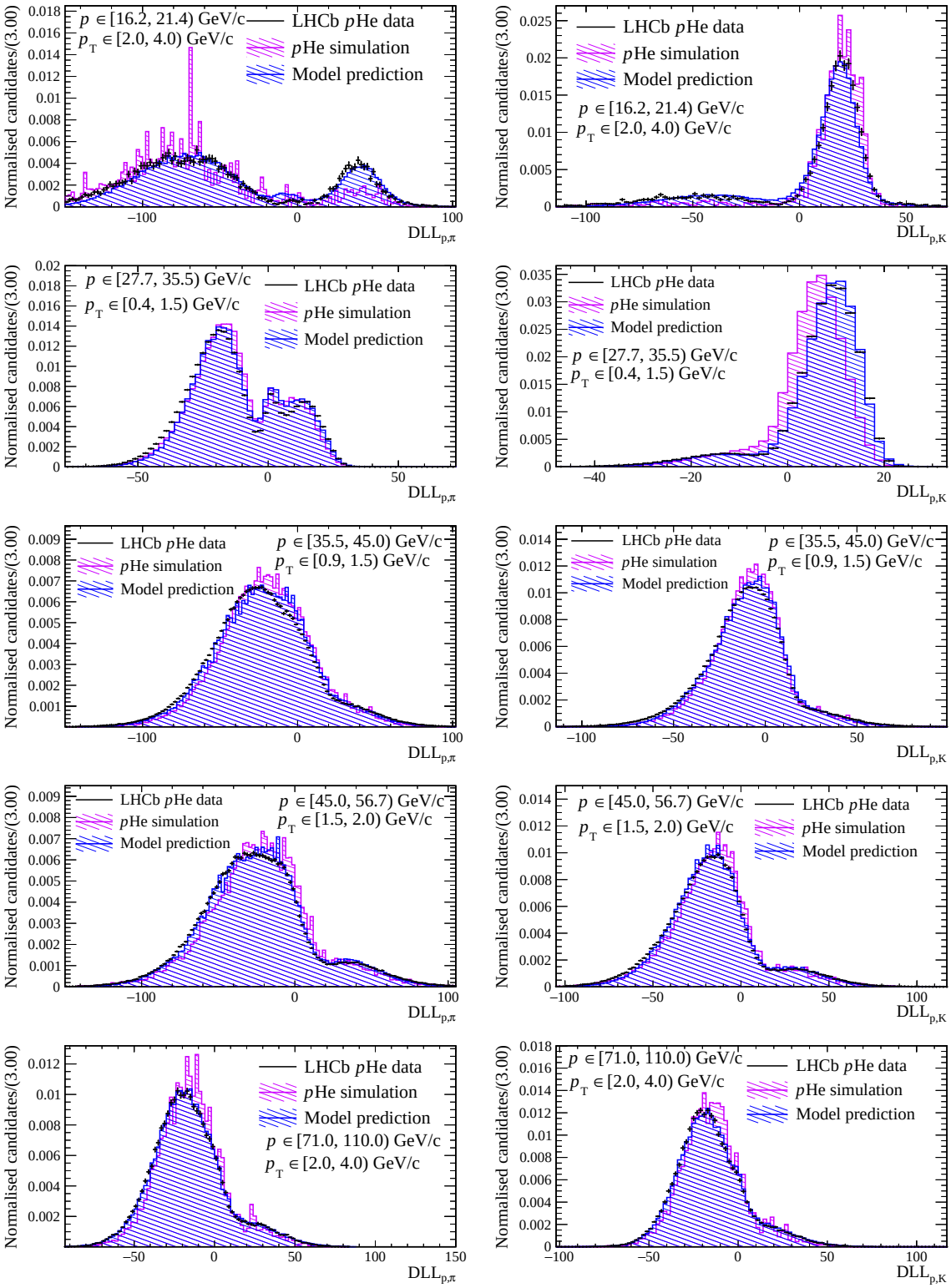


Figure III.1.18: Comparison of the (left) $\mathrm{DLL}_{p,\pi}$ and (right) $\mathrm{DLL}_{p,K}$ distributions in pHe data modelled with simulated- (violet) and data-based (blue) templates in five momentum, transverse momentum intervals.

Table III.1.6: PID efficiency for the requirements applied in the antiproton selection obtained from simulation and from the machine-learning-based prediction for the same simulated prompt and detached antiproton samples.

$\overline{p}$ category	$\varepsilon_{PID}^{\text{sim}}$ [%]	$\varepsilon_{PID}^{\text{pred}}$ [%]	$\varepsilon_{PID}^{\text{sim}}/\varepsilon_{PID}^{\text{pred}}$
Prompt $\overline{p}$	65.70 ± 0.14	65.3 ± 0.2	1.006 ± 0.002
Detached $\overline{p}$	60.0 ± 0.5	60.48 ± 0.02	0.993 ± 0.007

efficiency is then computed as:

$$\varepsilon_{PID}^{\text{sim (pred)}} = \frac{\int_{\text{DLL}_{p,\pi}>20,\,\text{DLL}_{p,K}>10} f^{\text{sim (pred)}}(\text{DLL}_{p,\pi}, \text{DLL}_{p,K})\, d\text{DLL}_{p,\pi} d\text{DLL}_{p,K}}{\int f^{\text{sim (pred)}}(\text{DLL}_{p,\pi}, \text{DLL}_{p,K})\, d\text{DLL}_{p,\pi} d\text{DLL}_{p,K}}, \quad \text{(III.1.5)}$$

denoting $f^{\text{sim (pred)}}$ the bidimensional distribution of the $\text{DLL}_{p,\pi}$-$\text{DLL}_{p,K}$ variables in the simulation (predicted by the model). These are illustrated in the left (right) columns of Figs. III.1.23 and III.1.24 for prompt and detached antiprotons, respectively. In both cases, the generated template are much smoother than the data. By applying the PID selection requirements, the efficiency results are listed in Tab. III.1.6 and indicate that, once the differences in the input feature distributions are all taken into account, the simulation well reproduces the RICH response. This confirms that the different PID efficiency between prompt and detached $\overline{p}$ is well explained by the different kinematic distributions, that are reliably reproduced by the simulation.

III.1.4.3 Understanding kaon calibration samples in *pp* data

Also the studies of PID systematic effects in *pp* data can profit from the discussed model. For example, kaons PID in *pp* data can be described considering the $\phi \to K^-K^+$ (in red) and the $D^0 \to K^-\pi^+$ (in green) decay as a calibration channel. Even when considering narrow kinematic bins, differences are observed between the kaon PID distribution in the two decays. Considering the 2017 *pp* data sample, this is illustrated in one kinematic interval in Fig. III.1.25. A possible explanation for this difference could be the bias induced in the $\phi \to K^-K^+$ decay originating from the low mass difference between the parent and the two kaon masses and hence the hyper-space proximity between the two kaons. When tagging one of the two kaons with tight requirements, the PID distribution of the probe could be distorted.

In order to understand this effect, a GMM is trained on the negatively-charged kaons produced by $D^0 \to K^-\pi^+$ decays including the same features as those of the benchmark demonstration. According to the distributions of the same variables for K^- particles in $\phi \to K^-K^+$ decays, its $\text{DLL}_{K,\pi}$-$\text{DLL}_{p,K}$ distribution is generated (in orange). If the correlation between the two kaons is relevant, the model application, not taking it into account, will fail. On the other side, if the observed difference in the PID response of the two channels is due to some other quantity, a consistency between the data and the generated template is expected. The result is presented in Fig. III.1.25 superimposed to the PID distribution of kaons originating in D^0 and ϕ decays. The prediction of the model is found to match the data for $\phi \to K^-K^+$. Hence, most of the observed discrepancy between the two channels is demonstrated not to be due to the two-kaon correlation. This solves a long-standing puzzle in the LHCb PID studies.

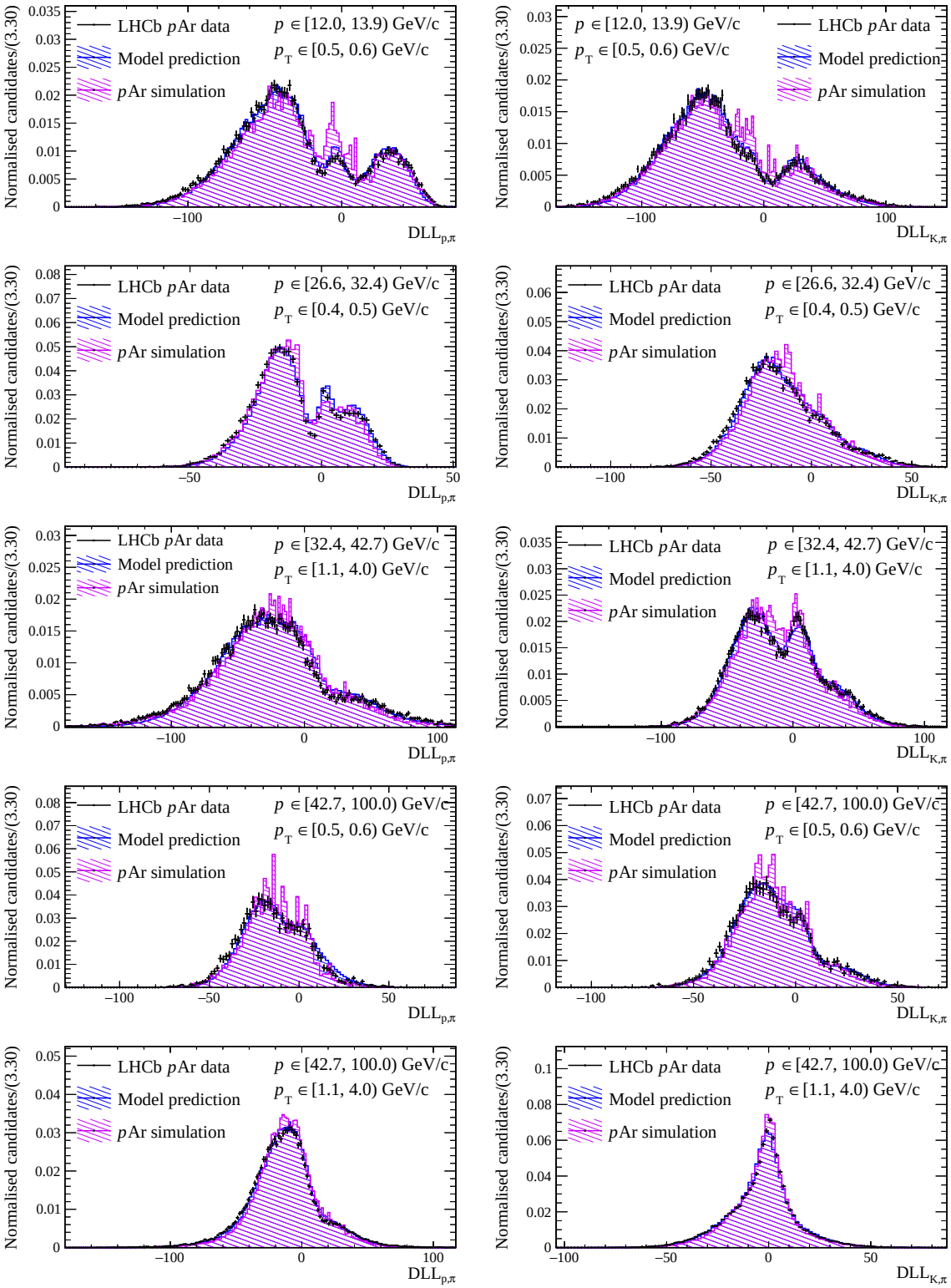


Figure III.1.19: Comparison of the (left) $\mathrm{DLL}_{p,\pi}$ and (right) $\mathrm{DLL}_{p,K}$ distributions in pAr data modelled with simulated- (violet) and data-based (blue) templates in five momentum, transverse momentum intervals.

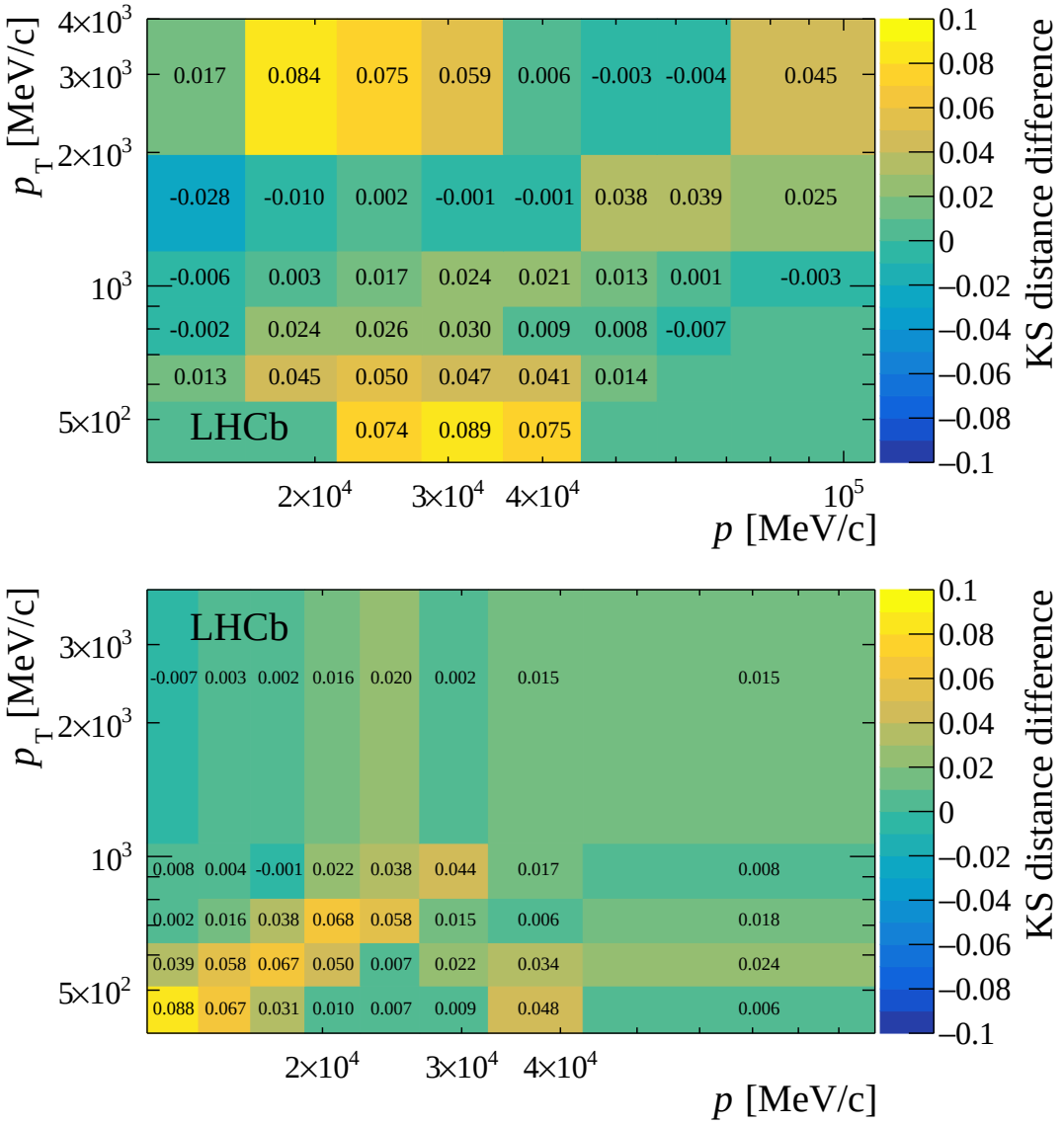


Figure III.1.20: Difference of the Kolmogorov-Smirnov distance for the modelling of the (top) pHe or (bottom) pAr PID classifiers distribution with templates drawn with the detailed LHCb simulation and with the machine-learning approach. Positive bin values, most in the considered kinematic phase-space, indicate a better quality of the latter in describing the data.

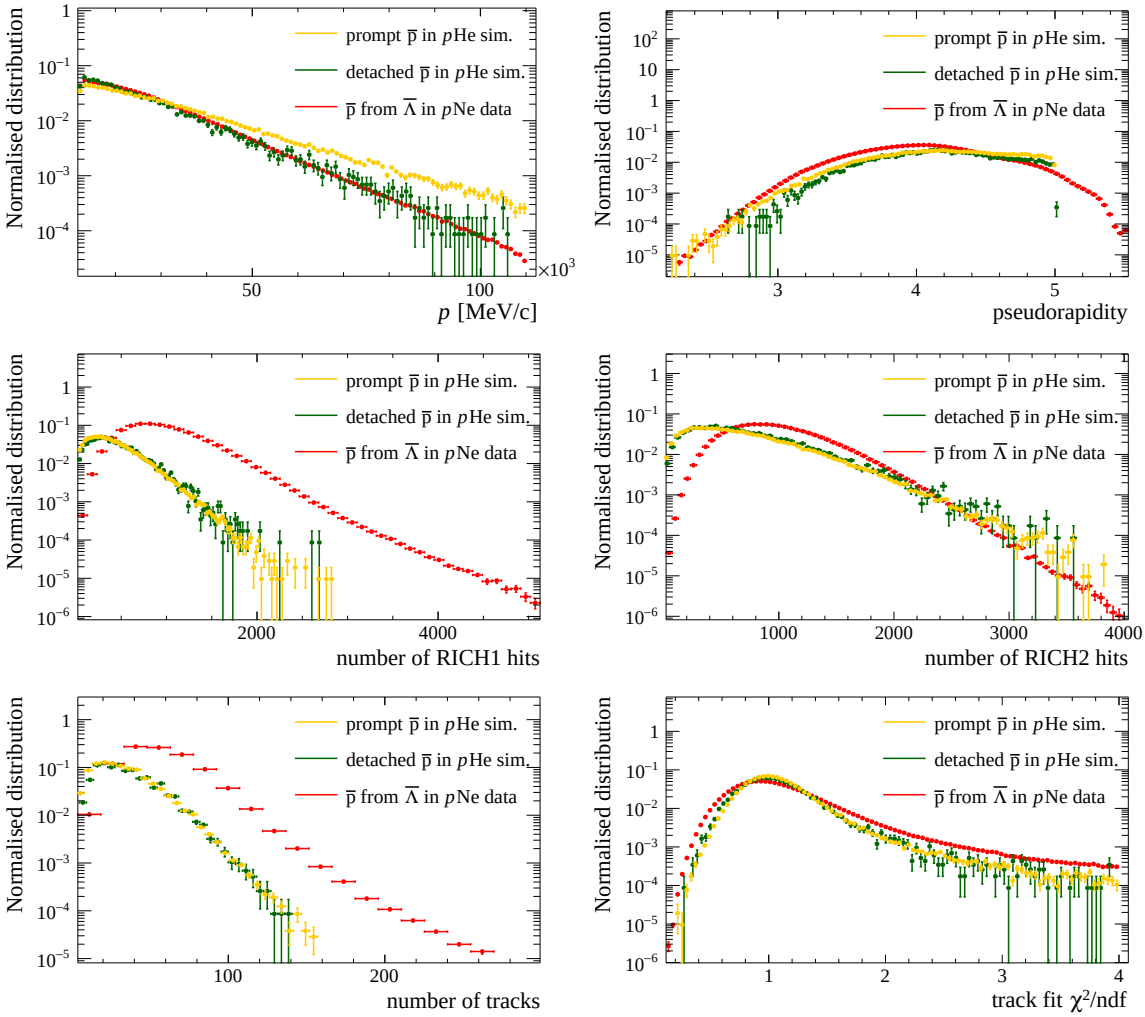


Figure III.1.21: Comparisons between the normalised distributions of (in order) the momentum, pseudorapidity, number of hits in the RICH1, RICH2 and reconstructed tracks and of the track fit χ^2/ndf for antiprotons from $\overline{\Lambda} \to \overline{p}\pi^+$ decays in the pNe data sample (in red) and for prompt (in orange) and detached (in green) antiprotons in the pHe simulation.

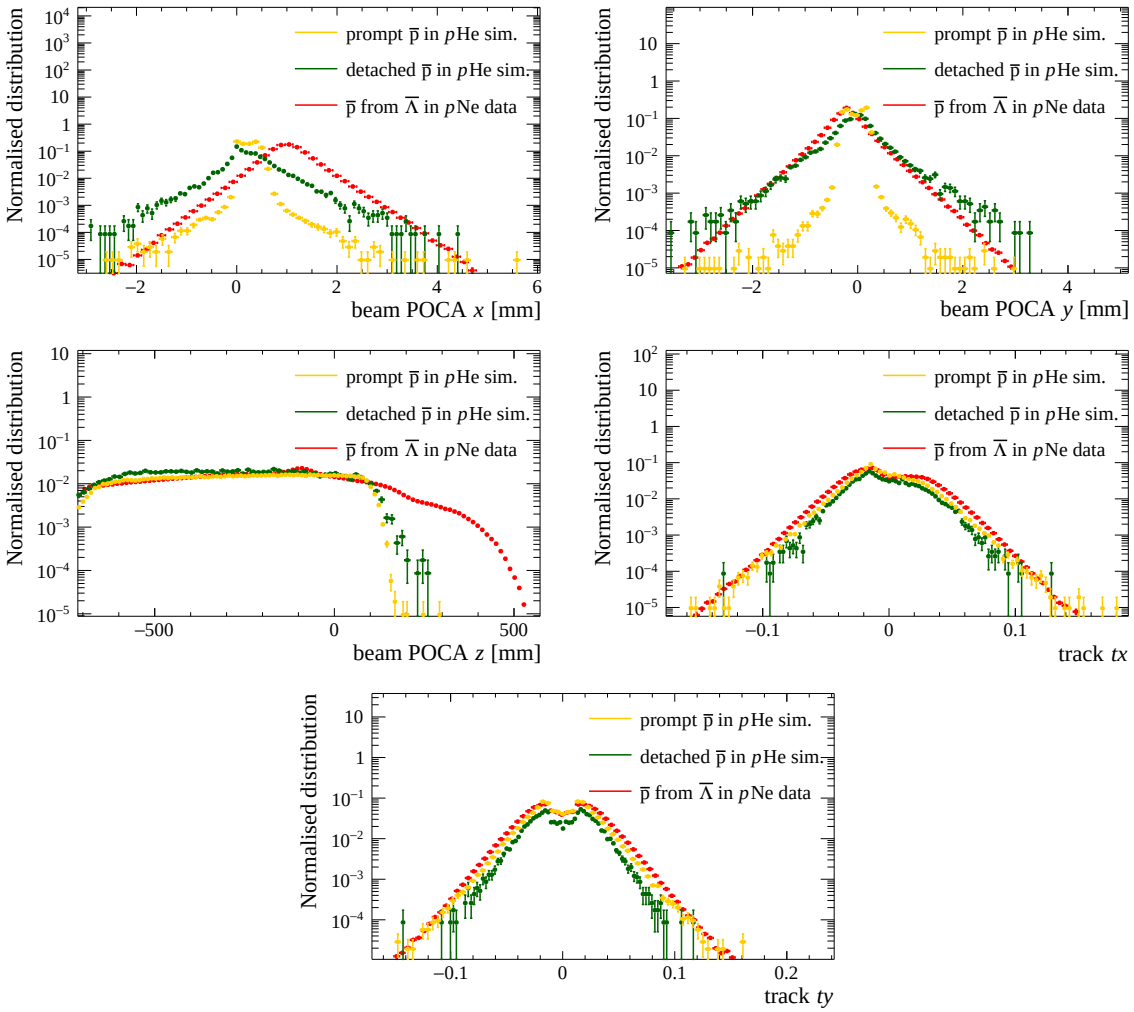


Figure III.1.22: Comparisons between the normalised distributions of (in order) the x,y,z coordinates and the tx and ty slopes for antiprotons from $\bar{\Lambda} \to \bar{p}\pi^+$ decays in the pNe data sample (in red) and for prompt (in orange) and detached (in green) antiprotons in the pHe simulation.

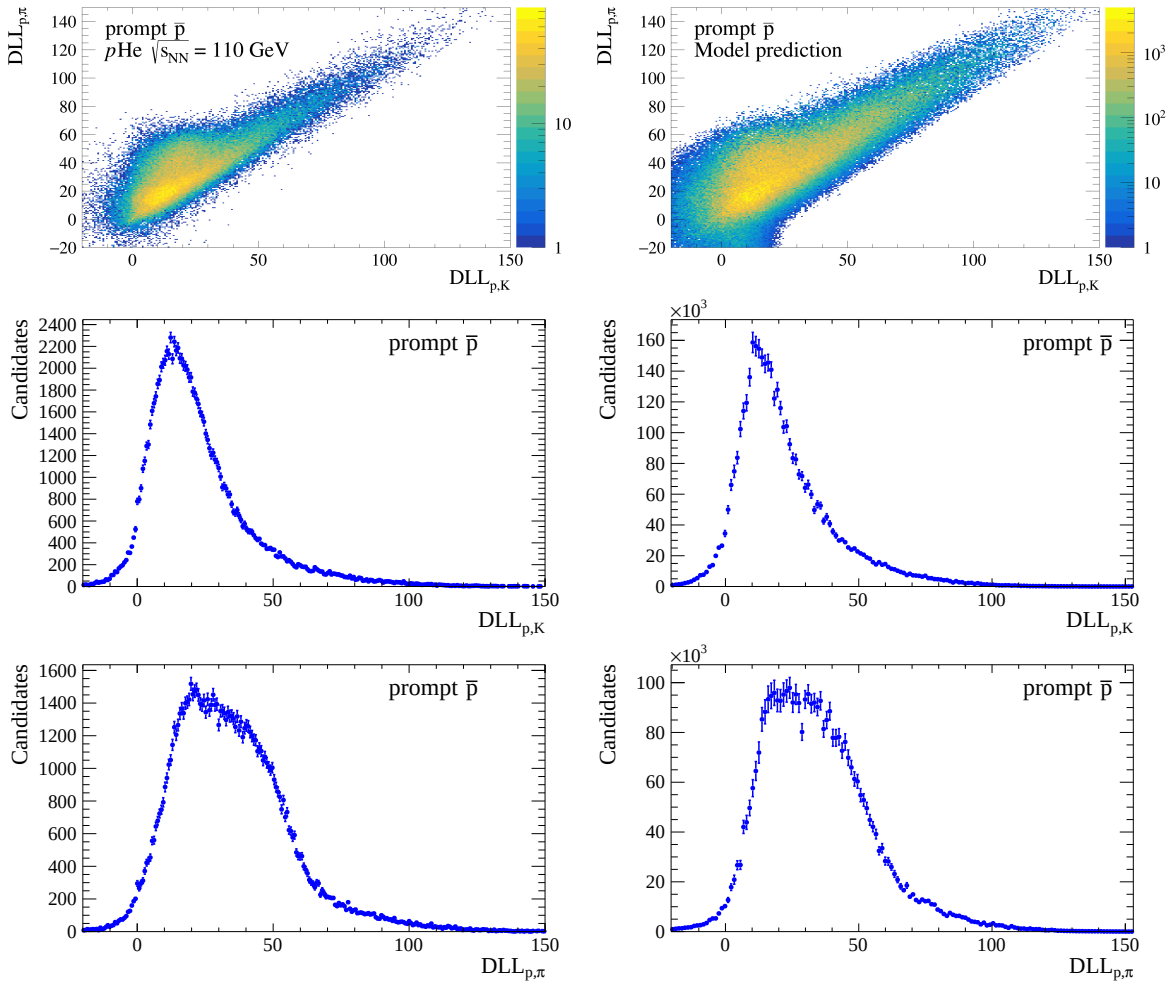


Figure III.1.23: Comparison between the prompt antiprotons $\mathrm{DLL}_{p,\pi}$ and $\mathrm{DLL}_{p,K}$ distributions in the pHe simulation (left) and predicted by the machine-learning model (right).

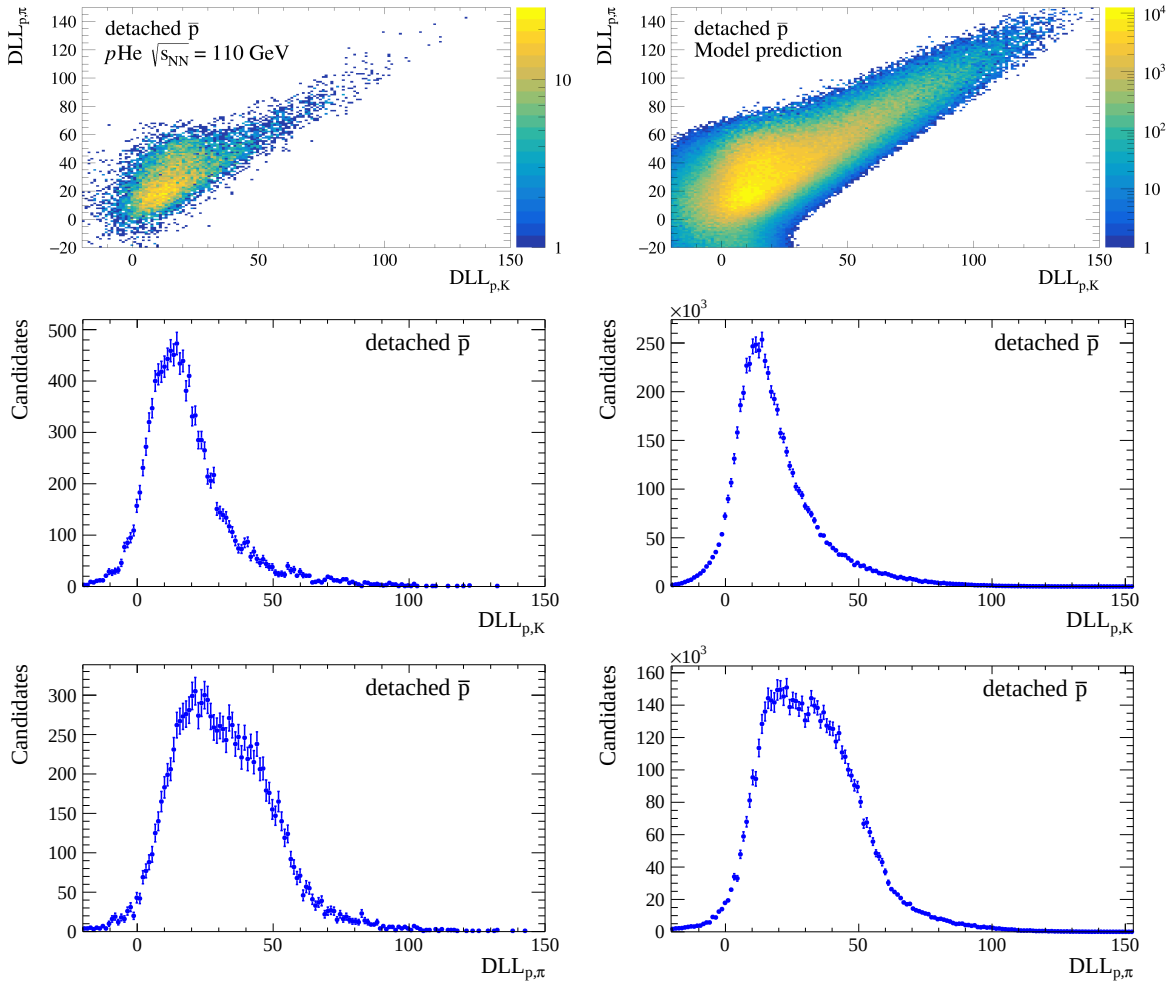


Figure III.1.24: Comparison between the detached antiprotons $\mathrm{DLL}_{p,\pi}$ and $\mathrm{DLL}_{p,K}$ distributions in the pHe simulation (left) and predicted by the machine-learning model (right).

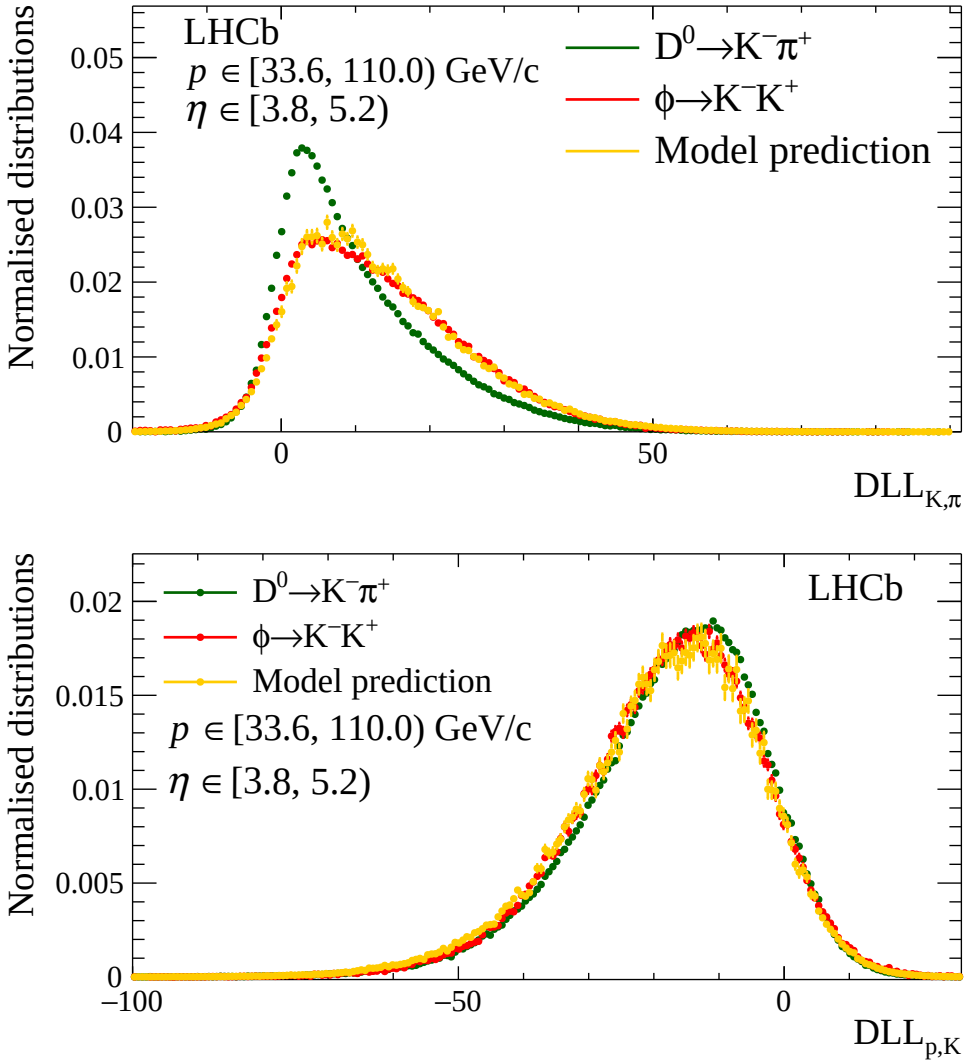


Figure III.1.25: Distributions of (top) the $\mathrm{DLL}_{K,\pi}$ and (bottom) $\mathrm{DLL}_{p,K}$ variables for kaons from the D^0 and ϕ calibration channels in 2017 pp data, restricted to the kinematic region where the two samples differ most. The prediction of the machine-learning-based model trained on D^0 and applied to the ϕ data is also shown and found to match with the distributions for kaons produced in ϕ decays. This indicates the possible systematic effect due to the correlation between the two final-state particles in the $\phi \to K^-K^+$ decay is negligible.

Part IV

Preparation to Run3 LHCb fixed-target physics

With the end of the LHC Run2 period of data-taking, the LHCb experiment started a major upgrade with the replacement of most of the detectors and the definition of an innovative strategy to acquire data and perform a full real-time event reconstruction. In parallel, an upgrade of the fixed-target system called SMOG2, mainly consisting in the installation of a 20 cm-long confinement cell for the gas, was proposed and approved by the LHCb collaboration to operate from 2022. A limited and well-defined region for the beam-gas collisions will allow the effective areal density of the gas target to be increased by up to two orders of magnitude and also non-noble gases such as hydrogen, deuterium and oxygen could be exploited. Thanks to the separation between the beam-beam and beam-gas interaction regions, a simultaneous and continuous LHCb operation in collider and fixed-target mode will be explored. Together with the more accurate control of the gas flow offered by an improved gas injection system, an unprecedented fixed-target physics programme will be performed, foreseeing to probe the nucleon and nuclear structure through heavy flavour and Drell-Yan production and more studies of light flavour production of interest to CR physics.

This part describes the activities I was responsible for during my *Ph.D.* programme to include the data acquisition with SMOG2 within the real-time framework being deployed for the Run3 LHCb. I tested and optimised the event reconstruction algorithms on simulated beam-gas events, aiming to use a unique setup to efficiently reconstruct them together with the beam-beam collisions. Chapter IV.1 is dedicated to the upgrade of the LHCb experiment and illustrates the Run3 detectors and data acquisition strategy. Chapter IV.2 presents the SMOG2 project, covers the hardware conception and installation and the related prospects for the physics opportunities. Chapter IV.3 discusses the reconstruction performance, particularly focussing on my activities to compare beam-beam and beam-gas data. These have been acknowledged in some LHCb technical papers describing the preparation to the upgrade, such as [146], and others are in preparation.

IV.1 The LHCb upgrade

In this chapter, the ongoing upgrade of the LHCb detector, introduced in Sec. IV.1.1, is described. Section IV.1.2 covers the upgraded detector structure, Section IV.1.3 the hardware trigger replacement with a real-time fully-software data processing strategy. De facto, with the LHC Run3 start in 2022, a new experiment will be running.

IV.1.1 LHCb upgrade overview

As discussed in Chapter I.3, even if initially conceived for precision measurements of b and c quark decays, LHCb evolved already during Run2 to a general-purpose experiment extending its physics reach. While no direct detection of physics beyond the SM has been achieved so far, hints for deviations from the predictions of the SM, possibly related to contributions from new particles, have been reported by the LHCb collaboration [147–149]. More data and an improvement of the experimental performance will possibly shed light on such deviations [150, 151], currently limited by the statistical uncertainty. The operational instantaneous luminosity will be increased from 2022 by a factor five, from the Run2 maximum value of $\mathcal{L} \simeq 4 \times 10^{32}\,\mathrm{cm}^{-2}s^{-1}$ to $\mathcal{L} \simeq 2 \times 10^{33}\,\mathrm{cm}^{-2}s^{-1}$. With these conditions, over the full LHC Run3 and Run4 periods of data-taking, a total of $50\,\mathrm{fb}^{-1}$ integrated luminosity is expected to be recorded. Accordingly, the rate of per-event collisions will increase to $\nu \sim 7.6$. As a consequence:

- to maintain the same performance despite the increased occupancy and pile-up conditions and thus the higher complexity for the event reconstruction, a finer granularity and a faster readout for all detectors are required;
- because of the harsher environment, all detectors must be resistant to the radiation. The upgraded VELO detector, for example, is expected to accumulate a maximum integrated flux of $\sim 8 \times 10^{15}\,1\,\mathrm{MeV}n_{\mathrm{eq}}/\mathrm{cm}^2$ [152], the innermost sensors of the UT and SciFi detectors of $\sim 3 \times 10^{14}\,1\,\mathrm{MeV}n_{\mathrm{eq}}/\mathrm{cm}^2$ and $10^{14}\,1\,\mathrm{MeV}n_{\mathrm{eq}}/\mathrm{cm}^2$, respectively [153];
- despite the increase in data acquisition channels, the quantity of material crossed by the particles needs to be kept limited;

- considering the increase in the flow of the acquired data and in the complexity of the low-level trigger selections, a strategy to sustain a $\sim 40\,\text{MHz}$ detector read-out is needed (though, considering the rate of visible collisions in LHCb, the rate will be $\sim 30\,\text{MHz}$).

The upgrade of the LHCb experiment has been conceived to overcome all these challenges and is exemplified by Fig. IV.1.1. Regarding the detector, illustrated in the top figure and discussed in Sec. IV.1.2, the complete tracking system and the RICH optics will be replaced, while the M1 muon chamber, the PS and the SPD detectors will be removed. Also, as presented by the graphs in the top-left corner of the figure, the full electronics and data acquisition will be upgraded. Concerning the data processing, as shown in the bottom figure and discussed in Sec. IV.1.3, the hardware trigger level will be removed, as its efficiency for hadron and electromagnetic signatures could not withstand the luminosity upgrade. The LHCb experiment will be the first one equipped with a fully-software detector readout, calibration and alignment and with a real-time event reconstruction and selection, designed to operate at the upgrade rate.

IV.1.2 The LHCb upgrade detectors

While the general LHCb detector structure will be maintained, most of its detectors will face major replacements. The VELO detector strip sensors will be replaced by pixels, offering an increased granularity. The tracking chambers, TT and IT-OT, will be superseded by the Upstream Tracker (UT) and the Scintillating Fibre (SciFi) detectors, employing silicon and scintillator fibres, respectively. Following the removal of the hardware trigger level, the detectors feeding it, the PS, SPD and the M1 muon chamber will be dismounted, also to reduce the material budget. Several other improvements to the support structures, to the mechanics and to the readout electronics, not described in the following, will also take place.

IV.1.2.1 The tracking system

The upgrade of the VELO detector [152] is exemplified by Fig. IV.1.2 and by the comparison of some relevant parameters with respect to Run2 in Tab. IV.1.1. As illustrated in the top plots, the z positions of the VELO stations will change from Run2 (in black) to Run3 (in red), profiting from the removal of the pile-up hits modules, the left-most black ones, mounted in Run2 for trigger and luminosity measurement purposes. The upgraded positions are optimised imposing that all particles in the detector acceptance, shown in yellow in the top right figure, traverse at least four modules. The middle figures present the modules shape and dimensions. The semi-circular Run2 geometry is replaced by an L-shaped one and the minimum sensor distance from the beam is lowered from the 8.2 mm to 5.1 mm. Together with the replacement of the strips with 41 million 55 μm $\times$ 55 μm silicon pixels, this is expected to significantly improve the reconstruction performance and the resolution on the IP determination. The radiation hardness is improved thanks to micro-channel cooling of evaporative CO_2 circulating in capillaries surrounding the sensors and keeping them at a stable temperature $\sim$ -25°. The supporting structures have also been improved. As for the LHC Run2 data-taking period, all VELO modules are included in a RF box separating the LHC and the detector

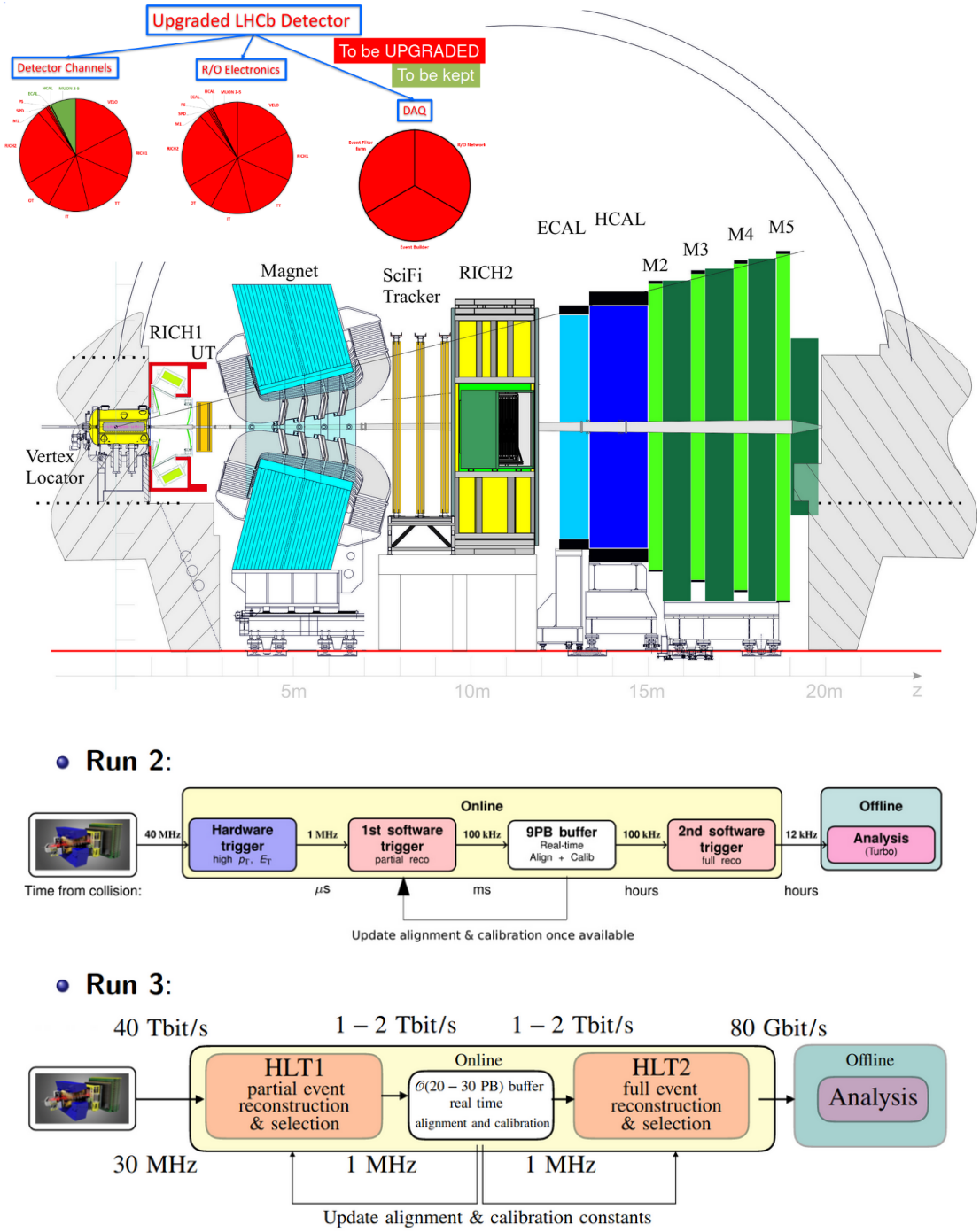


Figure IV.1.1: Figures of merit for the upgrade of the LHCb experiment taking place before the start of the LHC Run3 period of data-taking. The top plot presents the upgraded detector structure and the graphs in top-left show the fraction of detector (left), electronic (middle) and data acquisition (right) channels that are maintained (in green) or upgraded (in red). The bottom figure compares the Run2 (top) and Run3 (bottom) data-acquisition strategies, highlighting the removal of the hardware trigger level.

Table IV.1.1: Comparison between the Run2 and Run3 VELO characteristics.

Feature	Run2 VELO	Run3 VELO
Years of operation	2010-2019	2022-2030
Sensors	semicircular $\sim 0.2M$ R, ϕ strips	L-shaped $\sim 41M$ pixels
Distance from the beam	8.2 mm	5.1 mm
Radiation resistance	$4.3 \cdot 10^{14} n_{\rm eq}/{\rm cm}^2$	$8 \cdot 10^{15} n_{\rm eq}/{\rm cm}^2$
Peak data rate	$\sim$ 150 Gb/sec	$\sim$ 2.8 Tb/sec
Operational temperature	$\sim$ -8°	$\sim$ -25°

vacuum. The RF foils, illustrated in the bottom of Fig. IV.1.2, have a 150-200 μm thickness, about half than Run2.
The performance of the upgraded VELO (in red) is studied on simulation and compared to the Run2 detector (in black) when operating in the same conditions, *i.e.* a total rate of collisions $\nu = 7.6$ at a $\sqrt{s} = 14$ TeV energy. Figure IV.1.3 presents two examples [152]: the VELO tracking efficiency as a function of the pseudorapidity (left) and the resolution on the x coordinate for the IP measurement as a function of the inverse of the transverse momentum (right). In both cases, a significant improvement is achieved.

The Upstream Tracker (UT) detector [153] is installed between the VELO and the magnet with the same goals and structure of the Run2 TT stations. It provides some reference coordinates for charged particle trajectories before the bending due to the magnetic field and minimises the fraction of tracks reconstructed from spurious hits. Moreover, since the matching of the VELO track segments with energy deposits in the UT provides a rough estimation of the particle momentum, very-low $p_{\rm T}$ tracks can be removed and tight search windows are opened to extrapolate the track across the magnetic field. Both aspects result into a significant speed-up of the data processing. Also, the reconstruction efficiency for long-living particles such as Λ and $K^0_{\rm S}$, mostly decaying outside the VELO, is improved. As illustrated in Fig. IV.1.4, the UT instruments four stations of silicon microstrip detectors arranged according to the same *xuvx* geometry as the TT, but the detector granularity and radiation resistance is improved. The minimum distance from the beam is reduced by $\sim$ 7 cm, while the strip pitch is maintained to 187.5 μm in the external region (the green one in the figure) and reduced to 93.5 μm for the others.

The tracking system is completed by the Scintillating Fibre (SciFi) detector [153], replacing the Run2 IT and OT. Considering the upgrade running conditions, the detector occupancy in a gaseous detector such as the OT would give poor track reconstruction efficiencies and a different technology employing scintillating fibres has indeed been chosen. As shown in Fig. IV.1.5, the SciFi detector is composed of four layers whose fibres are arranged according to the same *xuvx* geometry as the UT. Each layer is equipped with three stations with 128 2.4 m-long and 250 μm-diameter fibres, for a total of 11000 km of fibres and 524000 Silicon PhotoMultiplier (SiPM) operating at -40° to reduce the noise and the radiation damage. A resolution ≤ 100 μm is achieved over a total active area of 340 m^2.

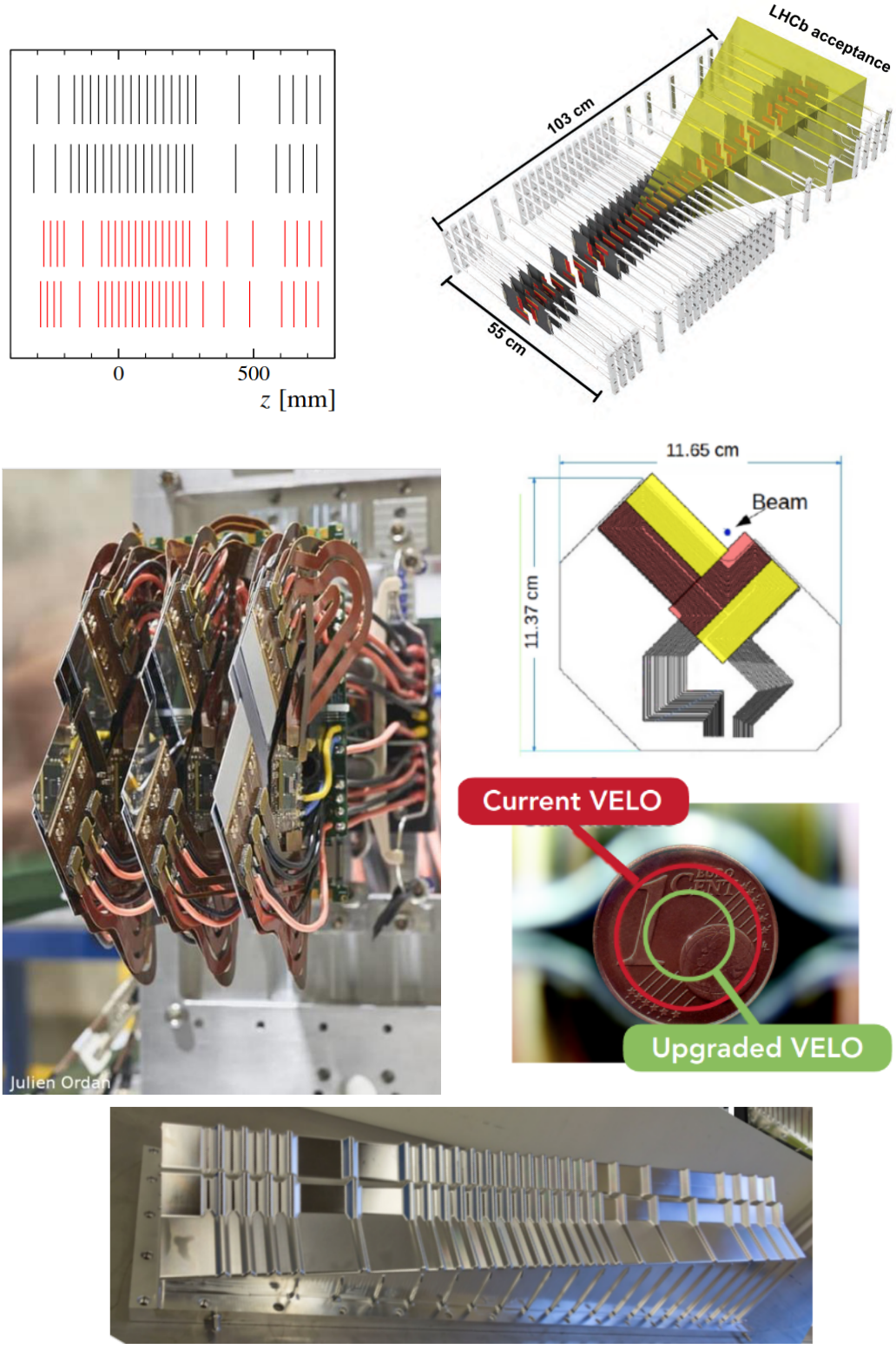


Figure IV.1.2: Upgrade of the VELO detector. The top left plot compares the Run3 (in red) and Run2 (in black) module z positions, the top right with the detector acceptance. The middle plots illustrate the module shapes and dimensions, the bottom is a picture for the upgraded RF foil.

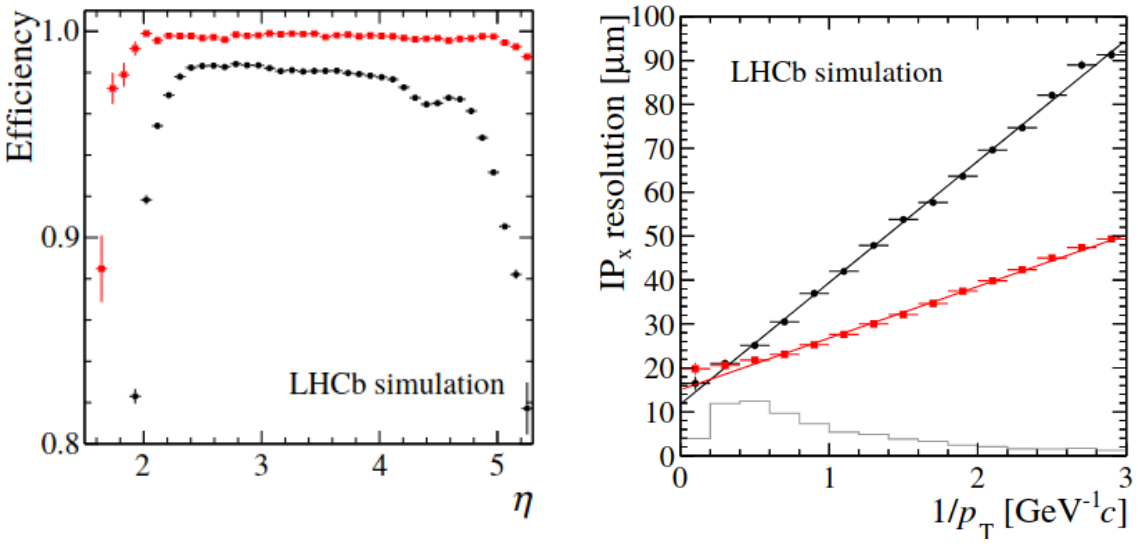


Figure IV.1.3: Two figures of merit for the performance of the upgraded VELO detector (in red) when compared to the Run2 one (in black) at the same conditions ($\nu = 7.6$, $\sqrt{s} = 14\,\text{TeV}$). The left plot presents the track reconstruction efficiency as a function of the pseudorapidity for particles leaving one hit in at least three VELO modules, the right figure the resolution on the impact parameter x coordinate for reconstructed VELO segments with $2 < \eta < 5$ as a function of the inverse of the transverse momentum. Figures from Ref. [152].

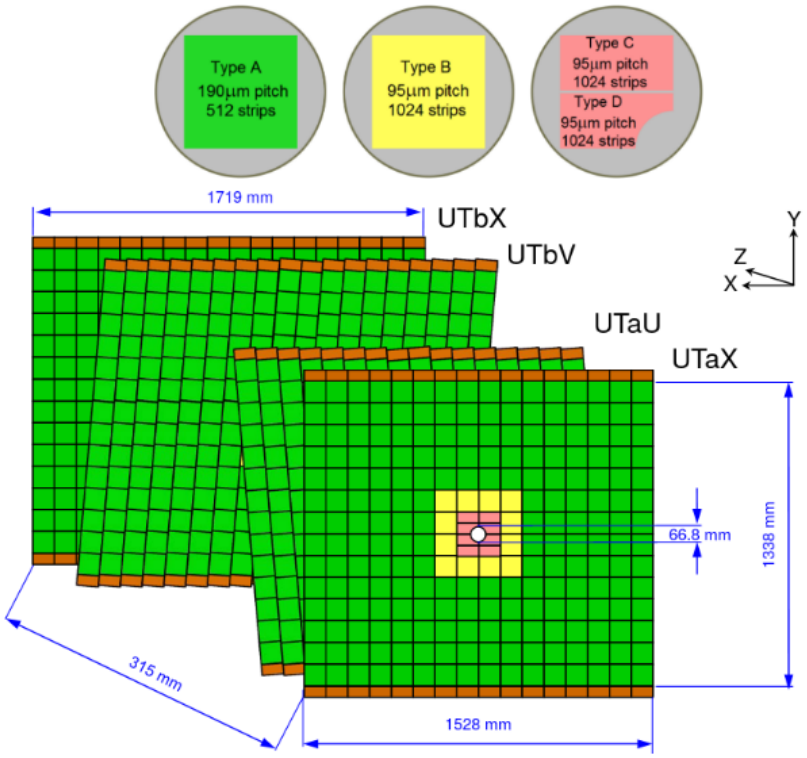


Figure IV.1.4: Structure of the UT detector. The orientations and the dimensions of the four stations are indicated. Figure from Ref. [153].

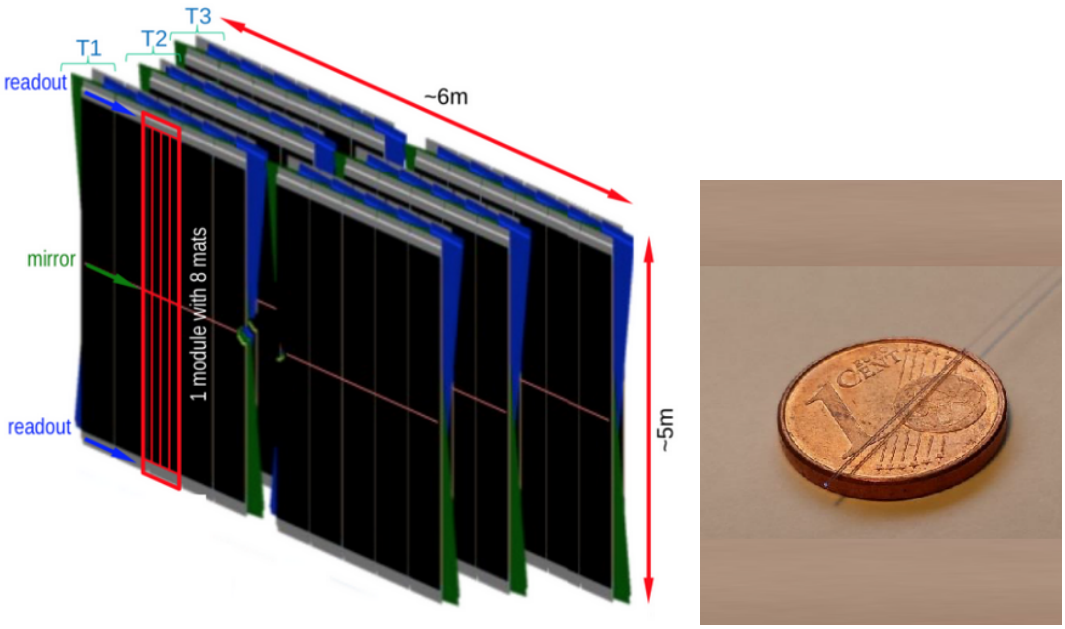


Figure IV.1.5: General structure (left) and example of a fibre (right) of the SciFi tracker of the LHCb upgrade detector.

A description of the algorithms decoding the tracking detectors and reconstructing tracks is included in Sec. IV.3. Two examples of the comparison for the performance of the Run3 (in light blue) and Run2 (in black) tracking system with the same conditions are shown in Fig. IV.1.6. The left plot shows the track reconstruction efficiency as a function of the pseudorapidity for only particles with $p > 5\,\text{GeV}/c$ in a simulated $B_s \to \phi\phi$ sample; the right figure presents the fraction of reconstructed tracks from spurious hits (ghosts) as a function of the number of reconstructed primary vertices in the event. In both cases, a significant improvement is achieved.

IV.1.2.2 The PID system

The overall structure of the PID system will remain unchanged in the upgrade, while major modifications are planned for future upgrades (see *e.g.* ref. [154]). Concerning the RICH system, the read-out HPDs, whose 1 MHz readout electronics is encapsulated within the tube, will be replaced by multi-anode photomultiplier tubes, capable to cope with the upgrade data rate. The RICH1 optical system has also been modified, in order to reduce the large LHCb hit occupancy in the central regions of the detector. By increasing the spherical mirror focal lengths from 2.7 m to 3.7 m, these are halved with respect to Run2. For the RICH2 detector, the optical system and the gas enclosure structures have been demonstrated to sustain the harsher LHCb Run3 running conditions and will retained. The only intervention for this detector is the installation of the new photon detector.
Also the layout of the calorimeter and muon system will be unchanged, other than the removal of the scintillators and the lead converter upstream ECAL and of the M1 muon station. The calorimeters read-out electronics will be rebuilt to work at the upgrade rate, while the gain of the photomultiplier tubes reading the calorimeter cells will be decreased

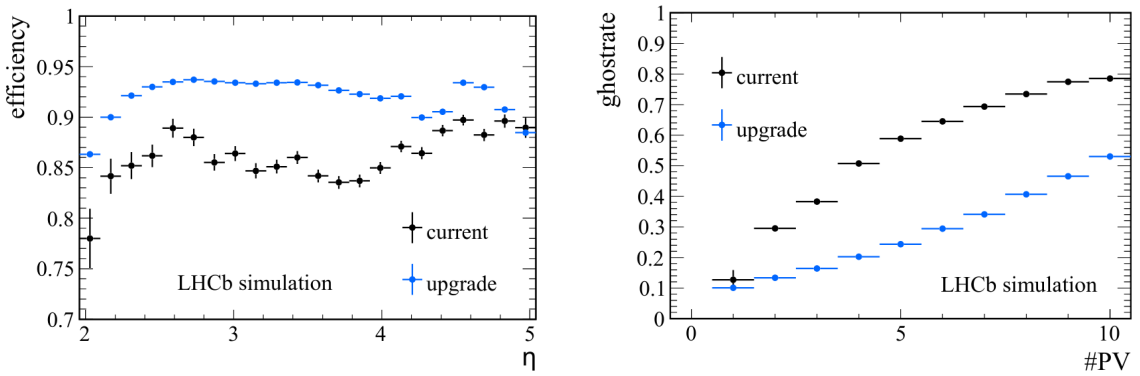


Figure IV.1.6: Two examples of the comparison between the performance of the Run2 (in black) and upgraded (in light blue) detector with the same conditions (simulated $B_s \to \phi\phi$ decays with $\nu = 7.6$ and $\sqrt{s} = 14\,\text{TeV}$). The left plot presents the tracking efficiency as a function of the pseudorapidity for tracks with $p > 5\,\text{GeV}/c$; the right one the fraction of tracks reconstructed from spurious hits as a function of the number of reconstructed primary vertices in the event. In both cases, a significant improvement is achieved. Figures from Ref. [153].

by a factor five to prevent their fast ageing. Regarding the muon system, the status of the MWPCs has been studied at the end of Run2 and no significant ageing effect has been observed [155]. Only the front-end electronics has been thus replaced to cope with the upgrade readout rate.
Since not crucial to the research activities performed during my *Ph.D.* programme, further details about the upgrade of the PID system are not discussed here and can be found in Ref. [156].

IV.1.3 The LHCb upgrade data acquisition strategy

Up to the end of the LHC Run2, the first stage of the LHCb trigger was implemented on dedicated FPGA cards and reduced the event rate to $\sim 1\,\text{MHz}$ based on minimum transverse momentum or energy deposition information from the calorimeter and muon systems. Such a strategy was highly efficient for dimuon channels, while typically half of the fully hadronic decays were removed. Assuming a constant output event rate, the yields that an hardware trigger level would present as a function of the instantaneous luminosities is illustrated in Fig. IV.1.7. Following the poor discriminative power of the simple hardware selections, the yield for hadronic channels saturates. For Run3, a purely-software trigger solution was thus chosen [157–159], as illustrated in Fig. IV.1.8. At the Run3 luminosity, the LHC collisions will correspond to a 4 Tb/s data volume, comparable to what the ATLAS and CMS triggers will have to process during High-Luminosity LHC running from 2027 onwards. An unprecedented rate of heavy flavour production will be reached: a b quark is produced with a frequency of 200 kHz, a c with 1 MHz.

The raw detector information is collected with the upgraded readout electronics and combined by 250 servers called Event Building (EB) farm. Assembled events are then sent to a separate Event Filter Farm (EFF), where, in the original proposal of

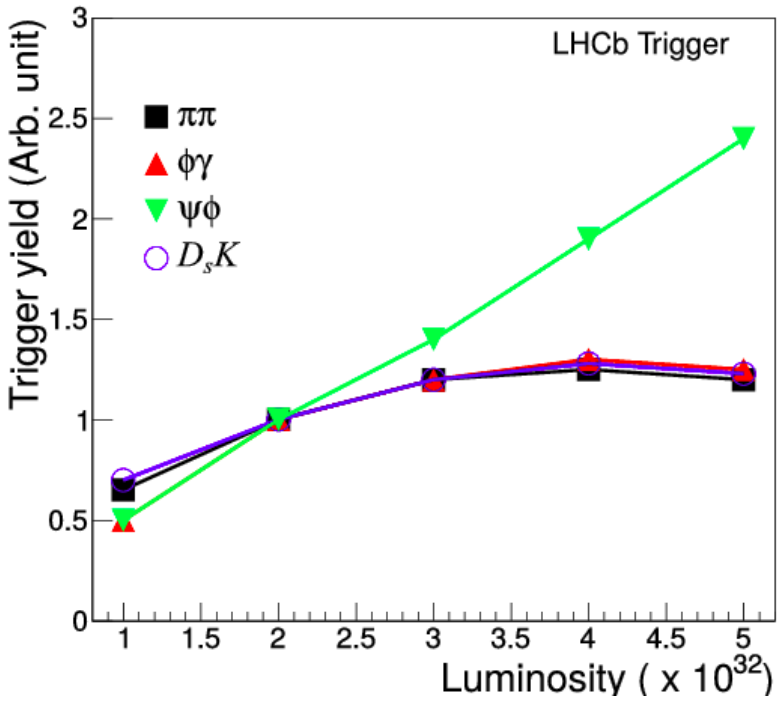


Figure IV.1.7: Hardware trigger yield for the indicated decays of B mesons as a function of the instantaneous luminosity assuming a constant output rate. Figure from Ref. [151].

the upgrade trigger, both software trigger levels run. The former performs a partial event reconstruction, including the decoding and pattern recognition tasks of all tracker detectors. At the output of the HLT1 trigger, as discussed in the next chapter, tracks, primary vertices and the muon PID information are available. The reconstructed event topology is then considered to define inclusive selection algorithms, accepting an event according to its general properties and not specifically optimised on a signal decay. For example, decisions on a single track can be taken according to its displacement with respect to any primary vertex or to its transverse momentum and reconstruction quality. Secondary vertices formed as two- or many-track combinations are also reconstructed in HLT1 and can fire a selection depending for example on their reconstruction quality or on their displacement with respect to the primary vertex. The set of HLT1 selections lowers the data rate to $\sim 1\,\text{MHz}$. The reduction is per-event, meaning that the full event information is persisted if there is at least one successful trigger decision and discarded elsewhere. Two different solutions for the HLT1 have been developed between 2018 and 2022: a purely-CPU-based, exploiting the most modern multi-core architectures, and an R&D activity exploring the possibility to run the full HLT1 sequence on Graphic Processing Unit (GPU) cards [160]. These were originally developed to accelerate graphics rendering, but their performance have been dramatically improved over the last ~ 20 years to parallel tasks. An algorithm, called kernel, is executed in parallel on thousands of threads running on the GPU sections named blocks. For HEP, this data processing paradigm, called Single Instruction Multiple Threads (SIMT), can be efficiently exploited to the acquisition of the LHC collisions. Each bunch crossing is independent and, within each event, tasks like the reconstruction of tracks and vertices can be conceived as parallel. Both the CPU and GPU options were demonstrated to meet the timing and physics

constraints [146] and the decision to use the latter was taken in 2020. GPU cards can indeed be hosted in the EB farm and execute the partial event reconstruction and HLT1 selections before the transmission to the EFF. All of this significantly reduces the data transfers between the two farms and the related costs. The project [160, 161] to run a complete high-throughput GPU reconstruction and trigger at LHCb is the first ever proposed among large HEP experiments.

Events selected by the HLT1 trigger stage feed an intermediate persistence step within a disk buffer. As already performed in Run2 [162], dedicated HLT1 samples are used to real-time alignment and calibration tasks of some detectors. For example, the beam-line position is needed by the primary vertex reconstruction algorithm, while the configuration of the RICH optical system and the relative displacement of the tracking modules affect the reconstruction performance. All these aspects are ruled by a set of calibration and alignment constants, which are regularly updated in real-time. Since the HLT2 trigger reads asynchronously the data from the buffer, it also allows its workflow to be distributed over time, including the LHC non-stable and machine development periods where data are not acquired. The HLT2 performs a full offline-quality event reconstruction and reduces the data rate to $\mathcal{O}$ (10 Gb/s). The HLT1 event reconstruction sequence is repeated with the updated alignment and calibration constants, a more precise fit of the track is performed and the PID information from the RICH and calorimeter system is processed. $\mathcal{O}$ (1000) selection algorithms, mostly exclusively reconstructing a specific signal decay chain, run in parallel. Following the approach developed in Run2 [113, 162], the data reduction is not only per-event, when no selection algorithm is fired and the full event is discarded, but can also be intra-event. In order to reach the 10 Gb/s output data rate, $\sim 70\%$ of the selection algorithms has to adopt the selective persistency, *i.e.* only a user-defined subset of the reconstructed event is saved permanently. This reduced in Run2 the event size by a half, while a reduction by up to a factor seven is expected for the upgrade [163]. On the opposite, selection algorithms requiring the persistency of the full bunch crossing information are called full stream.

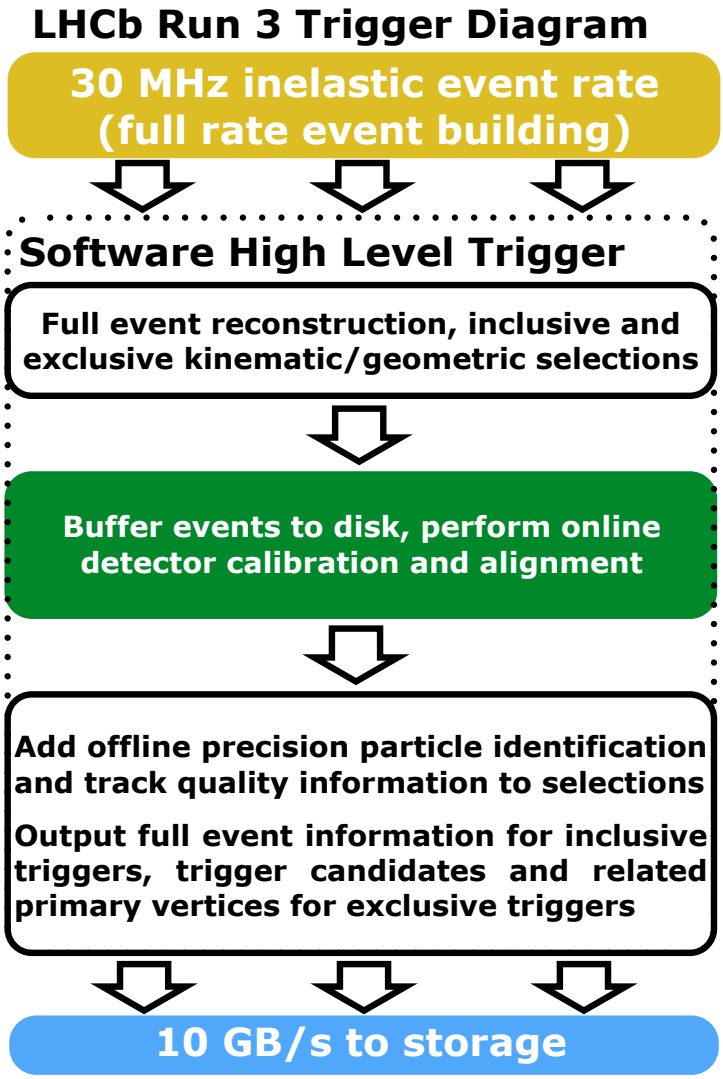


Figure IV.1.8: Scheme for the LHCb Run3 data acquisition strategy. No hardware trigger stage is used, while two software levels, HLT1 and HLT2, provide the detector decoding, calibration and alignment and the real-time event reconstruction and selection. Figure from Ref. [109].

IV.2 The SMOG2 project

In this chapter, the extension of the fixed-target programme exploiting the new SMOG2 target is presented. Section IV.2.1 summarises the limitations of the Run2 programme, motivating the installation of the confinement cell and of the new system for the gas injection described in Sec. IV.2.2. Section IV.2.3 discusses the physics prospects, particularly focussing on the completion of the cosmic-ray-physics motivated programme.

IV.2.1 Motivations for the upgrade

As discussed in part I, the injection of noble gases in the LHC beam-pipe, originally conceived to decrease the systematic uncertainty for the luminosity measurement, was exploited in Run2 to perform a rich physics programme. Anyhow, some technical and operational aspects limited the project:

- the spread of the injected gas in ± 20 m around the LHCb nominal interaction point. In order to keep the beam contamination low to the overall LHC operations, the statistics of the collected samples was limited, since:
  - only some noble gases at a maximum pressure $\mathcal{O}\left(10^{-7}\right)$ mbar could be injected;
  - the overlap between the beam-beam and beam-gas collision regions suggested a SMOG data acquisition mostly in short dedicated periods. For simultaneous data-taking with beam-beam, only the bunches entering the detector from the VELO side without a corresponding colliding bunch in the other beam, typically about 10% of the total, were used to record beam-gas collisions and a significant interference between the two types of collisions was observed;
- since the injection system could only host one gas bottle at time, a gas replacement required an intervention in the experimental cavern. This prevented to acquire data with different targets during the dedicated data-taking periods;
- SMOG was not equipped with gauges for the gas pressure that could be calibrated *in situ*. As discussed in Sec. I.4.2.2, the luminosity was determined through the analysis of proton-electron elastic scatterings, but the related uncertainty was the leading contribution affecting the cross-section measurements.

In order to overcome all these limitations, an upgrade of the SMOG system [164] is ongoing to install:

- an open-ended 20 cm-long confinement cell for the gas covering z values between 30 and 50 cm upstream the LHCb nominal interaction point;
- a more sophisticated Gas Feed System (GFS) with respect to Run2, equipped with four gas recipients and a precise determination and control of the injected gas flow and pressure.

As outside the cell the vacuum is maintained at the nominal level by the ion pumps serving the VELO volume, the beam-gas interaction region is better defined with respect to SMOG and the effective areal density of the target can be increased by up to two orders of magnitude with the same gas flow as SMOG. With the increased density, the background contamination from collisions with the LHC beam-induced residual gas is expected to be negligible. More gases, also non-noble species such as nitrogen, oxygen, hydrogen or deuterium and heavy noble gases like krypton and xenon could be injected. Studies are ongoing to determine the maximum integrated gas flow that is suitable for each gas species. Indeed, light gases like hydrogen or deuterium could induce embrittlement effects on the cell and beam-line elements, *i.e.* a degradation of their mechanical properties and, as for all active gases, could affect the integrity of the coating around the nominal interaction point. On the other side, heavy gases like krypton or xenon could be absorbed at the warm-to-cold transitions of the quadrupole magnets, increasing the local secondary electron yield. The cell displacement with respect to the nominal interaction point and the consequent separation between the beam-beam and beam-gas collision regions opens the possibility to simultaneously acquire the two types of collisions with all LHC bunches. The LHCb detector would be in that case the only LHC experiment simultaneously running in collider and fixed-target mode. Finally, the new calibrated GFS will allow the gas to be replaced in short time and the injected gas pressure, hence the beam-gas collision luminosity, to be precisely determined. The related systematic uncertainty on cross-section measurements is expected to be decreased by a factor between 2 and 3 with respect to Run2.

IV.2.2 The SMOG2 cell and the gas feed system

The installation of the gas confinement cell, representing the core idea of the SMOG2 upgrade, was accomplished in August 2020 and is shown in Fig. IV.2.1.

In order to maximise the injected target density seen by the LHC circulating particles, the cell is a tube coaxial with the beam. The gas is injected through a capillary at the centre of the cell and is extracted at its boundaries, resulting the injected gas density evolution as a function of z to approximately follow a triangular profile, as illustrated in Fig. IV.2.2. The volume density at the injection point, ρ_0, is a function of the injected particle flow I and of the cell conductance C_{tot}:

$$\rho_0 = \frac{I}{C_{\text{tot}}} = \frac{I}{3.81\sqrt{T/M}} \cdot \frac{L + 1.33\,D}{D^3} \tag{IV.2.1}$$

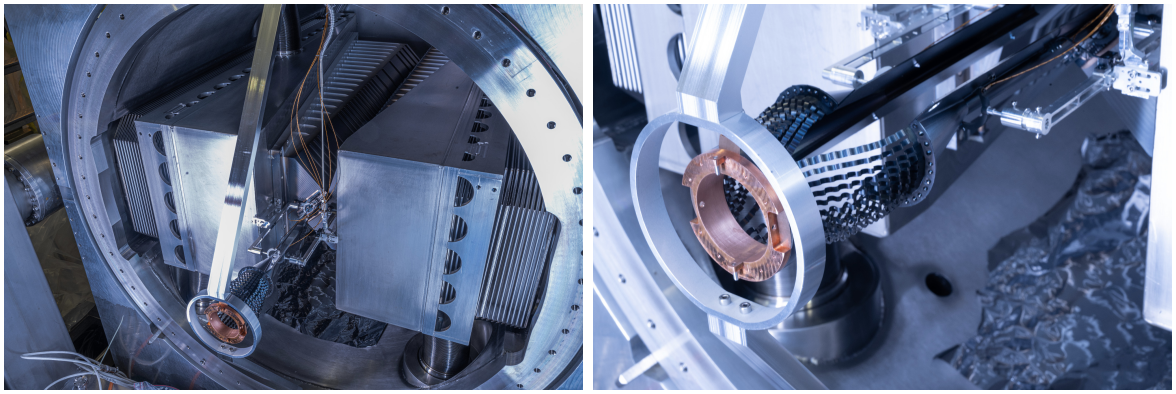


Figure IV.2.1: Pictures of the installed SMOG2 cell in its closed (left) and open (right) positions. Figures from Ref. [165].

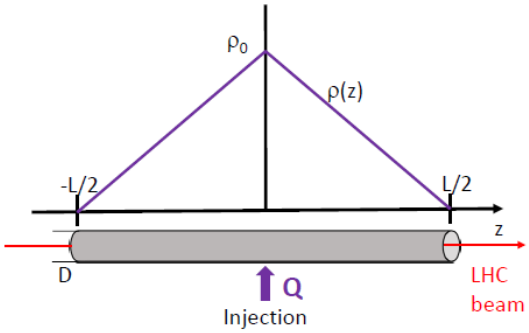


Figure IV.2.2: Sketch of the injected gas density as a function of z inside the SMOG2 cell. Figure from Ref. [164].

being L and D the cell length and diameter, T and M the gas temperature and molecular mass number. The target areal density θ_{target}, related to the fixed-target luminosity and to the beam flux ϕ_{beam} as $\mathcal{L} = \phi_{\text{beam}}\theta_{\text{target}}$, is given by

$$\theta_{\text{target}} = \frac{1}{2}\rho_0 L. \tag{IV.2.2}$$

The storage cell conception and implementation, exemplified by Fig. IV.2.3, is the result of the following requirements:

- the cell dimensions must fit into the available space within the VELO vessel, which limits its length to $L = 20\,\text{cm}$. The cell diameter is set to $D = 1\,\text{cm}$, chosen to optimise the luminosity. Being the average installation position 40 cm upstream the nominal interaction point, this value is validated against the radial aperture of the beam $r(z)$. In proximity to the interaction point, this evolves according to $r(z) = r_0\sqrt{1 + (z/\beta^*)^2}$, being r_0 the beam radius at $z = 0\,\text{mm}$ and β^* the distance from the interaction point where $r(\beta^*) = 2r_0$, respectively. For different beam scenarios, the $r(z)$ function is illustrated in Fig. IV.2.4 and the edges of the VELO detector (in black) and of the SMOG2 cell (in red) are indicated. For all scenarios, the cell radius is safely above the limits;

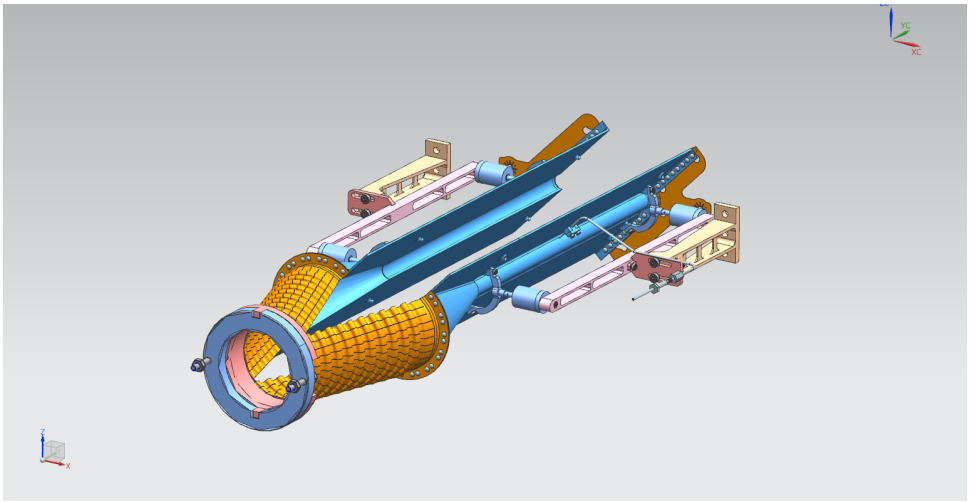


Figure IV.2.3: SMOG2 cell overview. Figure from Ref. [164].

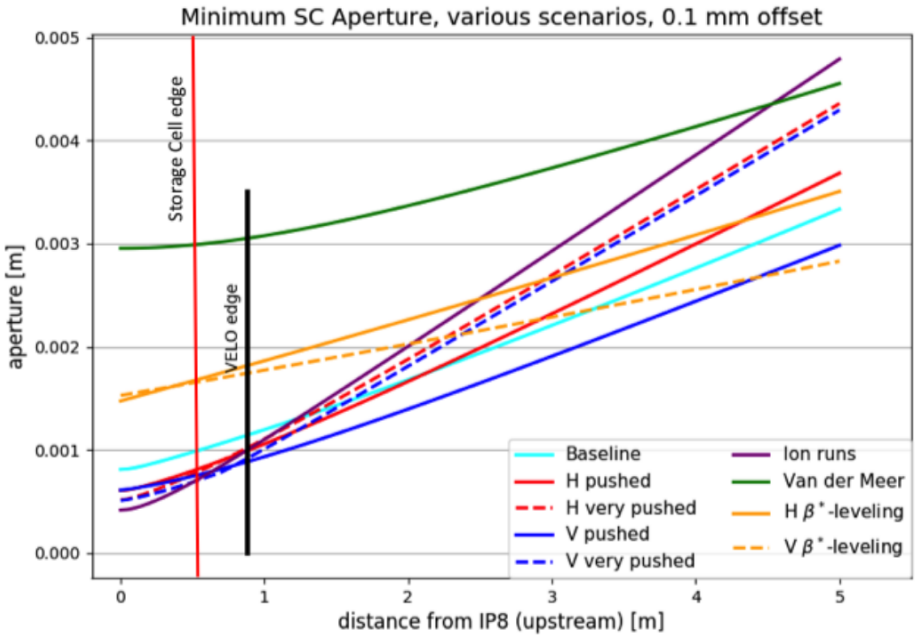


Figure IV.2.4: Beam radial aperture as a function of the distance from the LHCb nominal interaction point in different beam configurations. For all scenarios, the storage cell radius, whose position is indicated by a red line, is safely above the limits. Figure from Ref. [166], with vertical lines added by the authors of Ref. [164].

- the cell is made of two halves which, as the rest of the VELO detector, can be retracted from their operating position while the LHC beams are being injected and tuned, and closed when stable beams are declared. In order to account for the variations of the exact VELO closed position, only one of the two SMOG2 cell halves is rigidly connected to the VELO, while the other is equipped with a spring system to ensure enough movement flexibility. The closed (left) and open (right) positions of the cell are shown in Fig. IV.2.1;

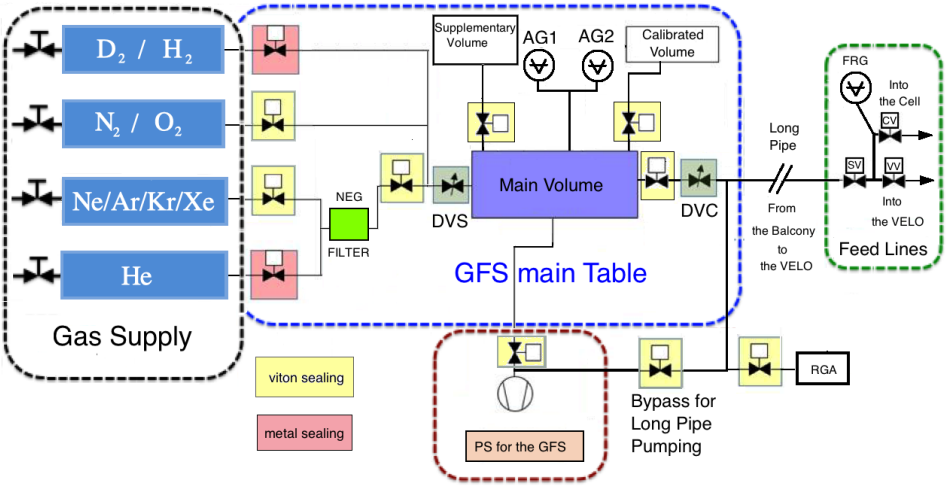


Figure IV.2.5: Sketch of the SMOG2 gas feed system. Figure from Ref. [164].

- the cell internal surfaces must be shielded against the radio-frequency fields generated by the bunched beam structure. This is achieved by the cell internal smoothness, excluding cavity-like or resonance structures and by a smooth connection with the VELO detector. Dedicated studies to the cell impedance and dissipated power and to the beam stability variations have been carried out in Ref. [164], with no degradation effects observed;

- the cell ensures electrical continuity through a CuBe Wake Field Suppressor (WFS), a cylindrical structure with a 56 mm diameter.

- the cell surfaces are coated with amorphous carbon to avoid Electron Clouds (ECs) formation, a phenomenon where electrons produced by different ionization effects are trapped around a positive-particles beam and cause avalanche effects leading to beam instabilities or losses. The coating with amorphous carbon, chemically less prone to ionization processes, limits the effect;

- the cell material, even with no gas injection, must have a low enough material budget to give a negligible additional Machine Induced Background (MIB). This was studied on simulation and, when properly compared to *pp* collisions, was demonstrated [164]. With the gas injection, the additional beam-loss effect of the proton-gas collisions is studied and measured in a relative loss below 0.5% (2%) in 10 hours of data-taking of proton (lead) circulating beams;

- To precisely measure the beam-gas luminosity through the areal density, the cell is equipped with four accurate temperature sensors.

The installation of the gas injection system, sketched in Fig. IV.2.5, is scheduled before the start of the Run3 LHC operations. As mentioned in the previous section, it is equipped with four gas reservoirs and a valve system able to replace the gas in a short time. A Residual Gas Analiser (RGA) employs mass spectrometer techniques to verify the absence

Table IV.2.1: Expected yields of reconstructed events for selected processes using the largest fixed-target data sample acquired with SMOG during the LHC Run 2 and with SMOG2 during Run3 in a scenario where one third of the beam time is shared with the beam-beam acquisition. Table from Ref. [121].

	SMOG largest sample pNe @69 GeV	SMOG2 example pAr @ 115 GeV
Integrated luminosity	$\sim 100\,\text{nb}^{-1}$	$\sim 45\,\text{pb}^{-1}$
J/ψ yield	15k	15M
D^0 yield	100k	150M
Λ_c^+ yield	1k	1.5M
$\psi(2S)$ yield	150	150k
Υ (1S) yield	4	7k
Low-mass Drell-Yan yield	5	9k

of gas contaminations. The gas pressure is precisely determined by two absolute gauges (AG1 and AG2), determining the injected flow comparing in Eq. IV.2.1. According to preliminary studies on simulation, a ∼2% precision on the luminosity measurement can be achieved.

IV.2.3 The SMOG2 physics opportunities

In this section, the physics opportunities [121] opened by the increased accuracy of the upgrade LHCb detector, the improvements in the fixed-target data acquisition and the possibility to inject new gases are discussed. The injection of hydrogen will offer reference runs for the study of nuclear effects with the other targets and will allow the high-x PDFs to be studied, sheding light on the possible nucleon intrinsic charm content. By comparing with injected deuterium, possible isospin-violation in the observed channels will be studied and the exploitation of heavy gases will allow nuclear effects to be precisely addressed [121]. Finally, with circulating lead ions, the improved LHCb upgrade reconstruction performance will allow central events to be reconstructed in PbA collisions. The prospects are mainly intended to:

- advance our understanding of the Quark-Gluon Plasma (QGP) providing key experimental results for the interpretation of heavy-ion data with different collision systems and at an unexplored energy between the SpS ($\sqrt{s_{\text{NN}}} \in [10, 30]\,\text{GeV}$) and the RHIC ($\sqrt{s_{\text{NN}}} \sim 200 - 500\,\text{GeV}$) ones;
- complete the LHCb programme of cosmic-ray-physics interest;
- contribute to describe the high-x gluon and quark nucleon content.

The feasibility for the QGP-motivated studies has already been proved by the Run2 measurements of the J/ψ and D^0 production cross-sections in pHe and pAr collisions [122], with the precision limited by the luminosity measurement and the available statistics ($\mathcal{O}$ (100) J/ψ and $\mathcal{O}$ (1000) D^0 candidates). For Run3, assuming a data-taking

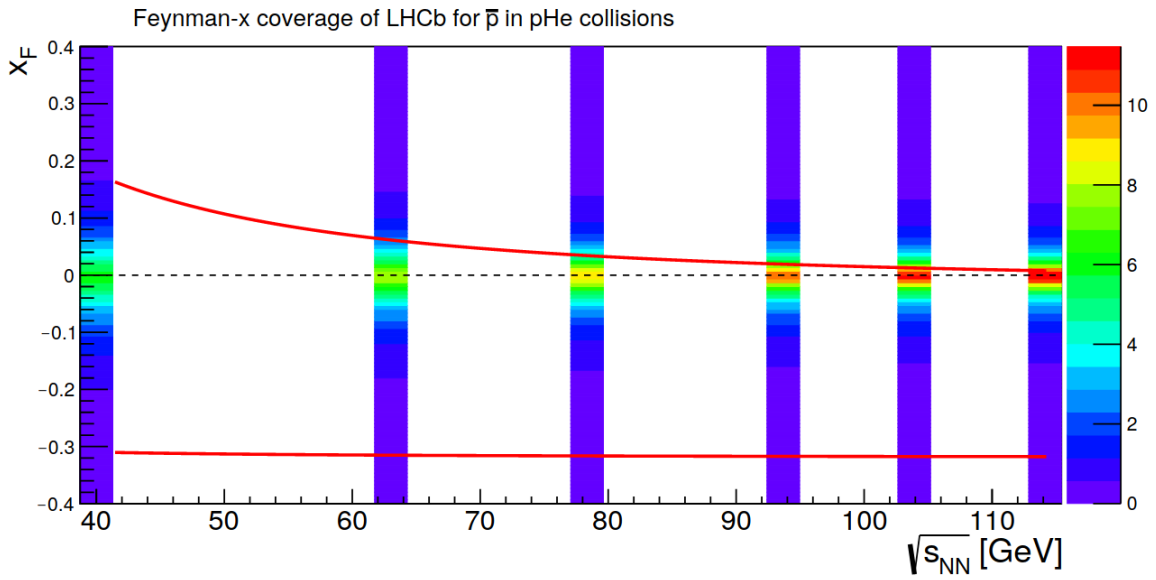


Figure IV.2.6: Expected Feynman-x, x_F, probability density function for antiprotons with momentum between 12 and 110 GeV, produced in pHe collisions at different centre-of-mass energies, according to the QGSJETII04m model [11]. The two red thick lines indicate the x_F minimum and maximum values accessible to LHCb. Figure from Ref. [121].

scenario with one third of the beam time shared between beam-beam and beam-gas collisions, the expected yields for some c- and b- hadrons are listed in Tab. IV.2.1. When comparing to the highest-statistics sample collected in Run2, larger values by up to three orders of magnitude are found. These production measurements are expected to provide key inputs to the interpretation of QGP-related effects, as the sequential suppression of excited charmonia states, observed at higher energies. Also, fixed-target charm meson and baryon states will give access to the high-x region and are expected to clarify the extent of the intrinsic heavy quark content in the nucleon, whose interpretation is still debated. Drell-Yan, whose expected rate is also listed in Tab. IV.2.1, and inclusive quarkonia production measurements will contribute to shed light on the nucleon structure, addressing the transverse momentum dependence of the quark and gluon PDFs. Other measurements of heavy-ion interest comprise the potential to study the elliptic flow over three fully-instrumented units of pseudorapidity and the possible Λ polarization induced by vorticity in heavy-ion collisions [167].

SMOG2 will also provide a unique kinematic regime for photon-induced processes, Ultra-Peripheral Collisions (UPC), with lower γ-hadron or γ-γ centre-of-mass energies with respect to pp. Dimuon decays such as the $J/\psi \to \mu^+\mu^-$ and two-body processes like $\phi \to K^-K^+$ or the $\rho \to \pi^+\pi^-$ will be extensively produced in SMOG2. Especially for high-Z targets, also the $\eta_c \to p\bar{p}$ production could be observed.

Measurements for hadronic production and spectra in different collision systems will also be performed with SMOG2, providing important inputs to the modelling of the underlying event and to a better understanding of the collisions in hadronic machines. Among these, some are also needed to the completion of the LHCb programme of

cosmic-ray interest, notably aiming to:

- extend towards lower collision energy the antiproton production cross-section measurement. While with the analysis at $\sqrt{s_{\rm NN}}$ =110 GeV, discussed in Sec. I.4.2.2, the LHCb acceptance corresponds to $x_F \in [-0.24, 0]$, the positive Feynman-x range can be accessed at lower energy values, as illustrated in Fig. IV.2.6;

- increase the measurement precision thanks to the lower luminosity uncertainty and comparing cross-sections with different injected gases collected with the same experimental conditions such as $\sigma(p\mathrm{He} \to \overline{p}X)/\sigma(p\mathrm{H} \to \overline{p}X)$, which is facilitated by the fast gas replacement;

- constrain the antiproton production in antineutron decays and investigate possible isospin violation effects through the measurement of $\sigma(p\mathrm{D} \to \overline{p}X)/\sigma(p\mathrm{H} \to \overline{p}X)$. As already discussed for bottom Fig. I.2.12, the available data for the antineutron-to-antiproton prompt production Δ_{IS} are scarce and a LHCb-SMOG2 result would have a strong impact on the theoretical models;

- explore the possibility to study the production of light anti-nuclei such as anti-deuterium or anti-helium;

- reproduce the collisions between secondary CRs and the atmosphere with the injection of nitrogen and oxygen. Measurements in these channels would provide key experimental inputs to the modelling of the background due to charmed-particle decays for neutrino astronomy in ground-based observatories. The deviations of the muon density in atmospheric showers from the predictions, the so-called muon puzzle [12], are also expected to be better understood profiting from these measurements. In case the proposal to perform a Run with circulated oxygen beams will be accepted [145], the injection in SMOG2 of hydrogen would open the very forward region (up to $\eta \simeq 7.6$) in the projectile rest frame.

IV.3 The data processing

In this chapter, the LHCb upgrade data reconstruction sequence is discussed. Particularly, the activities I was responsible for during my Ph.D. programme to test and optimise the reconstruction performance on simulated beam-gas collisions and to demonstrate the feasibility of a simultaneous data-taking with the beam-beam ones are described. Section IV.3.1 introduces the arguments in support and the requirements to the simultaneous data-taking; Sec. IV.3.2 presents the results for the reconstruction algorithms that are most dependent on the collision system and Sec. IV.3.3 gives some examples of the selection algorithms that I implemented to fulfil the SMOG2 physics programme.

IV.3.1 The beam-gas data-taking strategy

In order to maximise, for a given gas flow determined by the LHC safety constraints, the acquired luminosity of fixed-target data, a simultaneous LHCb exploitation of all LHC bunches to produce beam-beam and beam-gas collisions would be desirable. Within the SMOG2 project, this is supported by:

- the distinction between the two interaction regions, resulting from the beam-gas collisions confinement within the SMOG2 cell;
- the expected small interference of the injected gas on the beam-beam programme. Indeed, the beam-gas nucleon-nucleon centre-of-mass energy, ranging between 29 and 110 GeV, is much lower than the beam-beam one and the expected rate of collisions, to be defined depending on the injected gas species and pressure, is between 0.1 and 0.3 per-event. For pp data, these must be compared with an average of ~ 5 pp collisions at $\sqrt{s_{\mathrm{NN}}} = 14\,\mathrm{TeV}$. The detector multiplicity originating from beam-gas events is thus expected to be a fraction of $\mathcal{O}$ (1%) with respect to beam-beam.

One of the goals of my *Ph.D.* activities in preparation to the SMOG2 event acquisition is to prove the actual feasibility of this simultaneous data-taking, which requires to:

- demonstrate that the reconstruction algorithms provide at the same time an efficient reconstruction of beam-gas and beam-beam data, despite the different physics and geometry characteristics;
- prove that the interference between the two collision systems is minimal;

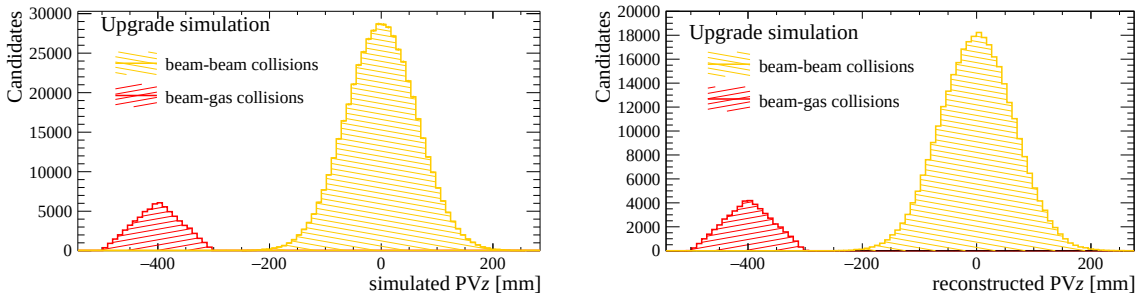


Figure IV.3.1: Simulated (left) and reconstructed (right) primary vertex z coordinate distributions for overlapped beam-beam (in orange) and beam-gas (red) collisions. For the former, the relative proportion is fixed considering the Run3 beam-beam μ and one fixed per-event beam-gas interaction. A clear distinction between the two interaction regions is found.

- measure the time required to process beam-gas data, which needs to fit into the tight constraints posed by the hardware trigger removal;
- implement a selection strategy maximising the efficiency on the signal of interest and the background rejection and also being able to avoid the beam-beam contamination on beam-gas data (and vice versa).

In case one of the conditions above is not met, mitigation strategies are available. If the separation in the selection algorithms between beam-beam and beam-gas data is not efficient, a compromise data-taking strategy could be that adopted for the 2017 pNe data sample, with beam-gas collisions only acquired for LHC bunches not colliding in LHCb, as discussed in Sec. I.4.2. This would simplify the trigger strategy, decoupling the two data types, and, differently than the 2017 pNe data, the interaction regions separation would exclude the contamination from spurious beam-beam collisions. On the other hand, only $\sim 10\%$ of the LHC bunches would be exploited in the fixed-target programme. In case the loss in beam-beam performance caused by the gas injection is also not negligible, limited and dedicated data-taking periods as it was done in 2016 will be the only viable option.

In order to perform these studies, simulated samples for the different data-taking scenarios were produced employing the Pythia [79] (EPOS-LHC [10]) generator for beam-beam (beam-gas) collisions. Following the expected gas pressure profile within the cell, illustrated in Fig. IV.2.2, a smearing of the simulated beam-gas collisions is applied requiring:

- a triangular distribution in $z \in [-500, -300]$ mm;
- a Gaussian distribution in x and y, given by the convolution of the uniform gas pressure and the beam transverse profile distributions.

In order to compare the different data-taking strategies, samples with stand-alone beam-gas and overlapped beam-gas and beam-beam collisions were produced considering helium and argon as two examples of light and heavy gases already injected in Run2. Moreover, a sample of beam-beam collisions with the same configuration as the previous ones but without the gas was considered as a reference. For the samples with overlapped beam-beam and beam-gas collisions, the Run3 μ condition is assumed for the former, one fixed

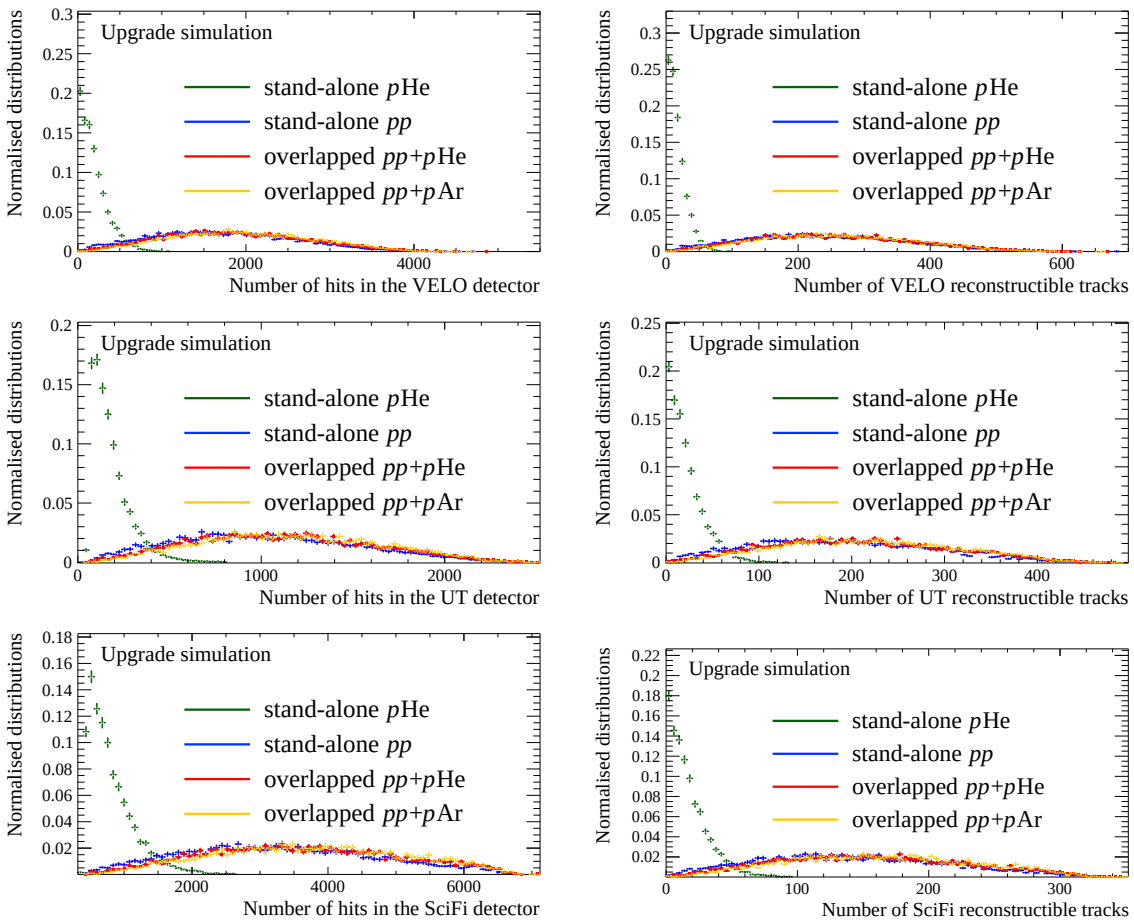


Figure IV.3.2: Comparison between the normalised distributions for the number of energy deposit clusters (left) and reconstructible tracks (right) in the VELO (top), UT (middle) and SciFi (bottom) detectors measured on simulated samples with stand-alone pHe (in green), pp (in blue) and overlapped $pp + p$He and $pp + p$Ar collisions. The gas addition to the pp collisions induces a negligible shift for all quantities.

per-event interaction for the latter. This results into an overestimation of the beam-gas rate of collisions, expected to be about 0.2, but the assumption is easy to scale to other injected pressures.

Figs. IV.3.1 and IV.3.2 confirm on these simulated samples the arguments supporting the simultaneous data-taking discussed above. The former shows the simulated (left) and reconstructed (right) distributions of the longitudinal coordinate of the collision vertices in the sample with overlapped beam-beam (in orange) and beam-gas (in red) collisions and a clear separation between the two components is found in both cases. The latter compares the normalised distributions for the occupancy in the VELO (top), UT (middle) and SciFi (bottom) tracking detectors for stand-alone pHe (in green), pp (in blue) and overlapped $pp + p$He (in red) and $pp + p$Ar collisions (in orange). The left histograms show the number of energy deposit clusters (hits), the right one the number of reconstructible tracks in the detector. For the VELO, UT and SciFi detectors, these are defined as the tracks presenting at least [168]:

- a hit in three VELO modules;

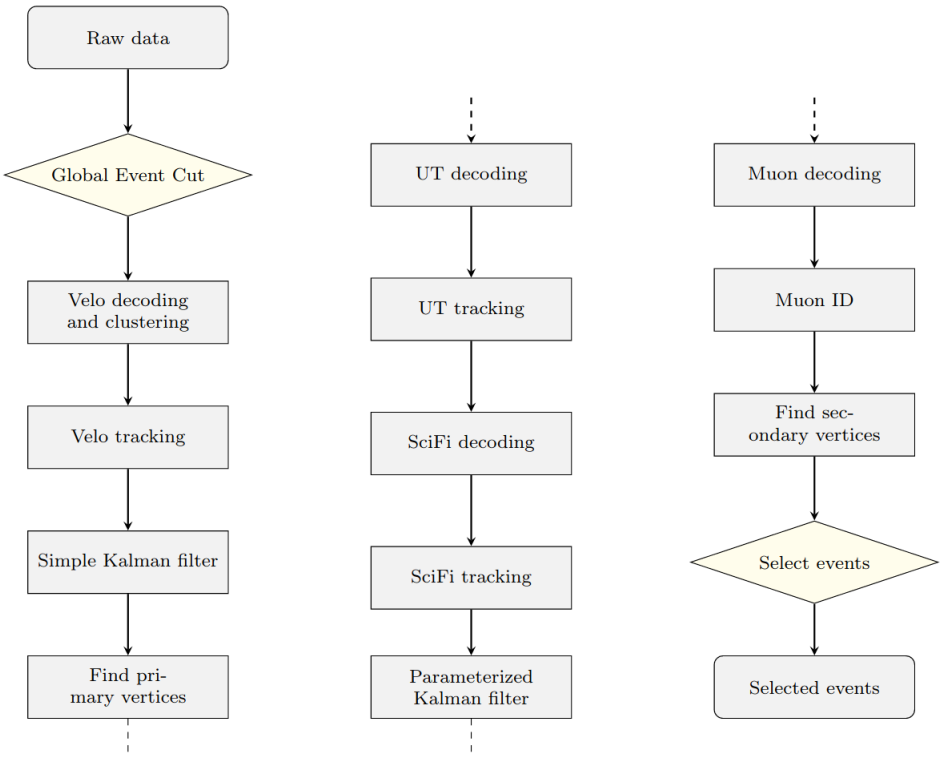


Figure IV.3.3: HLT1 reconstruction sequence. Figure from Ref. [161].

- one hit between the first two UT and one between the third and the fourth modules;
- one x and one u or v hit in each SciFi module,

respectively. For all plots, the increase induced by the gas injection in parallel to the beam-beam data is negligible.

IV.3.2 The event reconstruction sequence

As discussed in Sec. IV.1.3, the removal of the hardware trigger level poses strict timing constraints for the processing of the acquired data. The first trigger level, HLT1, has to sustain the 30 MHz visible collision rate and lowers it by a factor between 30 and 60. As discussed in the previous chapter, it performs a partial event reconstruction and runs selection algorithms to discard events not containing signatures for the physics channels of interest. The reconstruction sequence, summarised in Fig. IV.3.3, comprises:

- the decoding, clustering and reconstruction of VELO tracks;
- the determination of the position and composition of the collision vertices;
- the decoding and clustering of the UT detector;
- the VELO segments extrapolation to match energy deposits in the UT detector. A rough estimation of the particle momentum thanks to the small bending in the magnetic field is evaluated;

- the decoding of the SciFi detector energy deposits;
- the application of a Global Event Cut (GEC) on the sum of the number of energy deposits in the UT and SciFi detectors, which is a good indicator for the detector occupancy [158]. With this cut, the $\sim 10\%$ busiest events, which would require a not sustainable processing time to be reconstructed, are discarded;
- the upstream track extrapolation across the magnetic field to be matched with corresponding hits in the SciFi and to form the so-called long tracks;
- a simplified Kalman filter rejecting fake tracks;
- the extrapolation of long tracks to the muon detector. This is the only particle identification algorithm allowed by the timing constraints;
- the reconstruction of decay vertices.

As mentioned in Sec. IV.1.3, all algorithms have been developed in parallel employing multi-core modern CPU architectures and GPU cards. For both implementations, my *Ph.D.* activities in preparation to the SMOG2 data-taking contributed to test and adapt the performance on beam-gas collision events (see *e.g.* [169]). In particular, the different event geometry and energy with respect to beam-beam is expected to mostly affect the reconstruction of the VELO tracks and, as a consequence, of the primary vertices, as described in Sec. IV.3.2.1 and IV.3.2.2, respectively. Indeed, being these algorithms the first in the reconstruction sequence, they are designed to early reduce the data rate with the application of tight thresholds optimised on *pp* data. The extrapolations of the VELO segments to the UT and SciFi detectors, described in Sec. IV.3.2.3, are not expected to significantly depend on the z and thus to differ between beam-beam and beam-gas data, provided the kinematic properties are the same. Finally, Sec. IV.3.2.4 deals with the timing aspects of the sequence.

At the second stage of the event reconstruction, HLT2, a full offline-quality event reconstruction is performed. The included algorithms are intended to:

- fully reconstruct the tracks, exploiting all the information from the tracking system and performing pattern recognition tasks that do not fit in the HLT1 tight timing constraints. A more precise Kalman filter is performed to determine the track parameters and the particle momentum and redundant track reconstruction algorithms are added to reconstruct the particles that are kicked out from the LHCb detector region;
- reconstruct neutral particles with the information of the calorimeters;
- discriminate among the charged particle species thanks to the RICH system.

With respect to HLT1, all these algorithms are expected to be less dependent on the particle OV, differing between beam-beam and beam-gas data, and their study has been thus less central during my *Ph.D.* programme.

IV.3.2.1 The VELO track reconstruction

The first step of the HLT1 reconstruction is the decoding, clustering and tracking of the VELO detector and the evolution of the design of the algorithm in preparation to Run3 is described in the following.

A first algorithm to reconstruct the VELO tracks, named *PrPixelTracking*, was implemented in 2012 [170]. The raw energy deposits in the VELO are represented as pixels with a binary outcome and grouped in clusters when connected and active. Starting form the downstream-most VELO module ($z \simeq 700\,\text{mm}$), track seeds are defined as subsequent pairs of clusters and extrapolated backwards searching for corresponding hits. Only the tracks that would be compatible with the LHCb detector acceptance are considered and the extrapolation is interrupted in case hits in three following stations are missing. A straight-line model is assumed, considering that the magnetic field effect is negligible at the VELO module positions. Spurious reconstructed tracks are rejected by considering the fit χ^2/ndf for a straight-line fit firstly and then for a parametrization accounting for the particle trajectory deviations due to a scattering with the detector material. The physics performance of this algorithm already met the requirements for the upgrade of the VELO detector [152], but the algorithm vectorization, which could result into the required speed-up, was not considered convenient because of the time required for the data preparation.

In 2018, the algorithm was improved and renamed *PrPixelTrackingFast*. The clusters were ordered according to their ϕ and forward and backward tracks were processed separately. To reduce the algorithm output rate, flags and ϕ-windows for the track extrapolations were introduced to early kill segments in case of missing hits. As an example, by default, the search for forward track seeds was only performed for $z > -100\,\text{mm}$ and only extrapolated segment with $dr/dz > 0$ were considered. Even with these assumptions, the algorithm still lead the overall HLT1 processing time.

In 2019, the Single Instruction Multiple Data (SIMD) paradigm, *i.e.* the replication of a same instruction over different bunches of data, was adopted to profit from the highly-parallel modern CPU architectures and to vectorise the VELO pattern recognition algorithm. As mentioned above, the algorithm speed-up is eased by a coherent data structure, which indeed was migrated to the Structure Of Array (SOA) model, *i.e.* to the organization as contiguous storage of all the same structure data members, such as for example the x coordinates of all track hits. The strategy of this reconstruction algorithm, named *VeloClusterTrackingSIMD* [171], replaced the previous ϕ tolerance windows with a nearest-in-ϕ approach so that also displaced tracks, which are important for the LHCb physics programme, could be efficiently reconstructed without significant effects on the algorithm output rate [171]. Its timing and physics performance were discussed in the CPU-GPU performance comparison document [146].

For the GPU implementation of the VELO detector processing, the parallelization and the SOA structures were kept but adapted to the GPU architectures and to their Single Instruction Multiple Threads (SIMT) paradigm, according to which a single instruction is executed on data with multiple GPU threads. Subsequent collision events,

physically independent, are reconstructed in parallel. The VELO track reconstruction by the new algorithm, called *Search by triplet* [172], is also parallelised and executed in five logical steps:

- the VELO hits are prepared as SOA structures and sorted in ϕ;
- tolerance ϕ windows are opened among subsequent VELO modules;
- for each hit in a module with a preceding and a subsequent one, triplets are searched for in the opened ϕ windows. If more than a triplet is found, that with the best χ^2 for a straight-line fit is kept as track seed;
- track segments are forwarded among modules, allowing a configurable number of hits to be ignored in order to account for hit inefficiency;
- a maximum χ^2 threshold is set to reject fake tracks.

A further speed-up of the VELO processing can be obtained by the implementation of the clustering phase on dedicated FPGA cards [173, 174]. This was demonstrated to not degrade the performance on both beam-beam and beam-gas data, but is considered out of scope and not discussed in this thesis.

The performance for the VELO track reconstruction is studied considering [168]:

$$\begin{aligned} \text{efficiency} &= \frac{N_{\text{rec.ed}}}{N_{\text{rec.ible}}} \\ \text{fake rate} &= \frac{N_{\text{fake}}}{N_{\text{fake}} + N_{\text{rec.ed}}} \end{aligned} \tag{IV.3.1}$$

being:

- $N_{\text{rec.ed}}$ the number of reconstructed tracks with at least 70% of the hits belonging to one simulated particle;
- $N_{\text{rec.ible}}$ the number of particles that are reconstructible in the VELO detector, according to the definition above;
- N_{fake} the number of reconstructed tracks associated to more than one particle.

The ideal algorithm should have the largest efficiency and the possible lowest fake rate, together with the fastest processing. While the performance and the timing for all the described algorithms are discussed in the cited references, the comparisons between beam-beam and beam-gas simulated data, object of my studies, are presented here. The differences between the track topologies in the two samples are firstly studied. Those expected to mostly affect the VELO reconstruction performance, the simulated number of track VELO hits and the track pseudorapidity are compared for reconstructible VELO tracks from beam-beam and beam-gas collisions in Fig. IV.3.4. The top and middle plots show the bidimensional distributions for stand-alone pp and pHe collisions, respectively, the bottom ones the normalised projections onto the number of VELO hits (left) and pseudorapidity (right) axes. As expected from the installed position of the cell upstream the nominal interaction point, an average larger number of track VELO hits

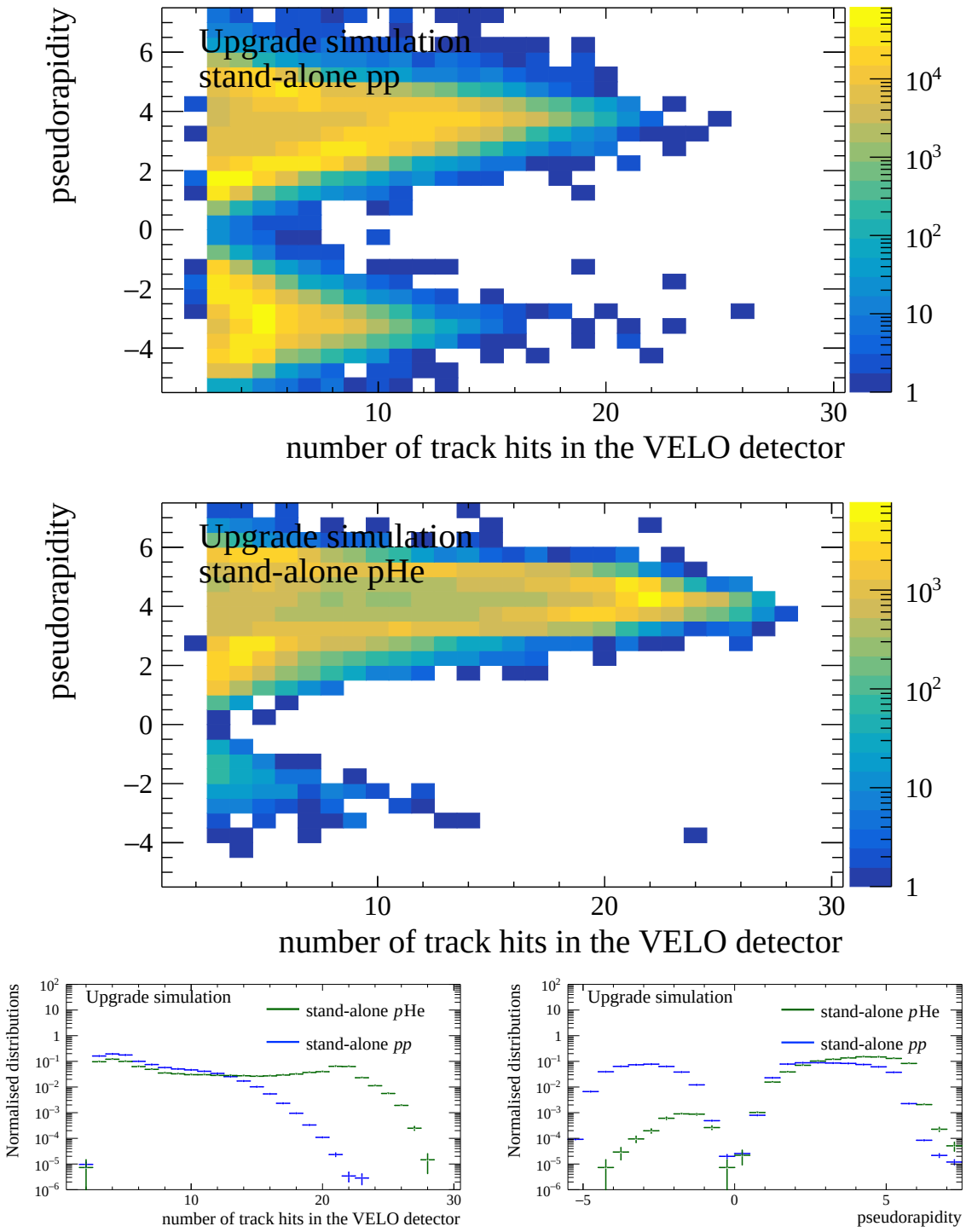


Figure IV.3.4: Comparison between the distributions in simulated stand-alone pp and pHe collisions for the track number of hits in the VELO detector and the pseudorapidity. The top and middle plots present the bidimensional distributions for the two samples, the bottom ones the normalised projections superimposed.

Table IV.3.1: *PrPixelTrackingFast* algorithm internal thresholds values optimising beam-beam or beam-gas performance.

Requirements	beam-beam	beam-gas
Minimum VELO sensor z in seed searching [mm]	-100	-300
Seed extrapolation with $dr/dz < 0$	disabled	enabled

is found for beam-gas than for beam-beam tracks. For the same geometric effect, the LHCb acceptance range is shifted towards larger pseudorapidity values for beam-gas events.

When my *Ph.D.* activities in preparation to the SMOG2 data-taking started, the *PrPixelTrackingFast* algorithm was the baseline option. Its tracking efficiency and inefficiency with the default implementation are shown as the red curves of Fig. IV.3.5. The top plot shows the performance as a function of the pseudorapidity for stand-alone pHe particles reconstructible in the VELO detector, the middle (bottom) for $pp + p$He particles reconstructible as VELO (long) as a function of the z coordinate of their primary vertex, so that the two types of collisions can be clearly distinguished. An efficiency drop at low pseudorapidities and an averaged 10% lower efficiency with respect to beam-beam is found for the beam-gas collisions. Such a worse performance was demonstrated not to be related to an intrinsic limitation of the algorithm, but to some of the thresholds applied to speed up the reconstruction. The thresholds mentioned above for the search of forward track seeds disfavoured the reconstruction of beam-gas-produced particles. Indeed, considering the SMOG2 cell position, these also cross the most-upstream sensors of the VELO detector and can also present $dr/dz < 0$. The tolerance ϕ windows were also studied as possible sources of inefficiency but turned out not to significantly contribute. By loosening these thresholds, as summarised in Tab. IV.3.1, most of the inefficiency for beam-gas events especially at low pseudorapidity values is recovered, as illustrated in the orange curves of Fig. IV.3.5. By comparing the default and tuned implementations of the *PrPixelTrackingFast* algorithm, an improvement of the efficiency is also observed at the boundaries of the beam-beam interaction region, which has a non-negligible impact on the output data and on the tracks fake rates. This breaks the requirement of the SMOG2 project of a minimal interference with the beam-beam physics and is an undesired effect. The solution was granted with the superseding of the *PrPixelTrackingFast* algorithm by *VeloClusterTrackingSIMD*. The beam-gas use-case had already been proposed and the repeated reconstruction efficiency studies, the blue curves of Fig. IV.3.5, indicated a better performance for beam-gas than for beam-beam events as a consequence of the larger number of track VELO hits, as discussed in Fig. IV.3.4. The introduction of the nearest-in-ϕ approach, conceived to boost the efficiency of displaced tracks, was also optimal for particles originating in SMOG2.
When the GPU reconstruction sequence was adopted, a comparison between the CPU and GPU implementations was performed and an increase in the VELO tracking efficiency below 1% was found, as presented in Fig. IV.3.6 for particles reconstructible as VELO (top) and long (bottom) tracks in stand-alone pHe collisions. The *Search by triplet* tracking performance is illustrated in Fig. IV.3.7 compared between beam-beam and beam-gas data. The top plot presents the VELO reconstruction efficiency as a function of

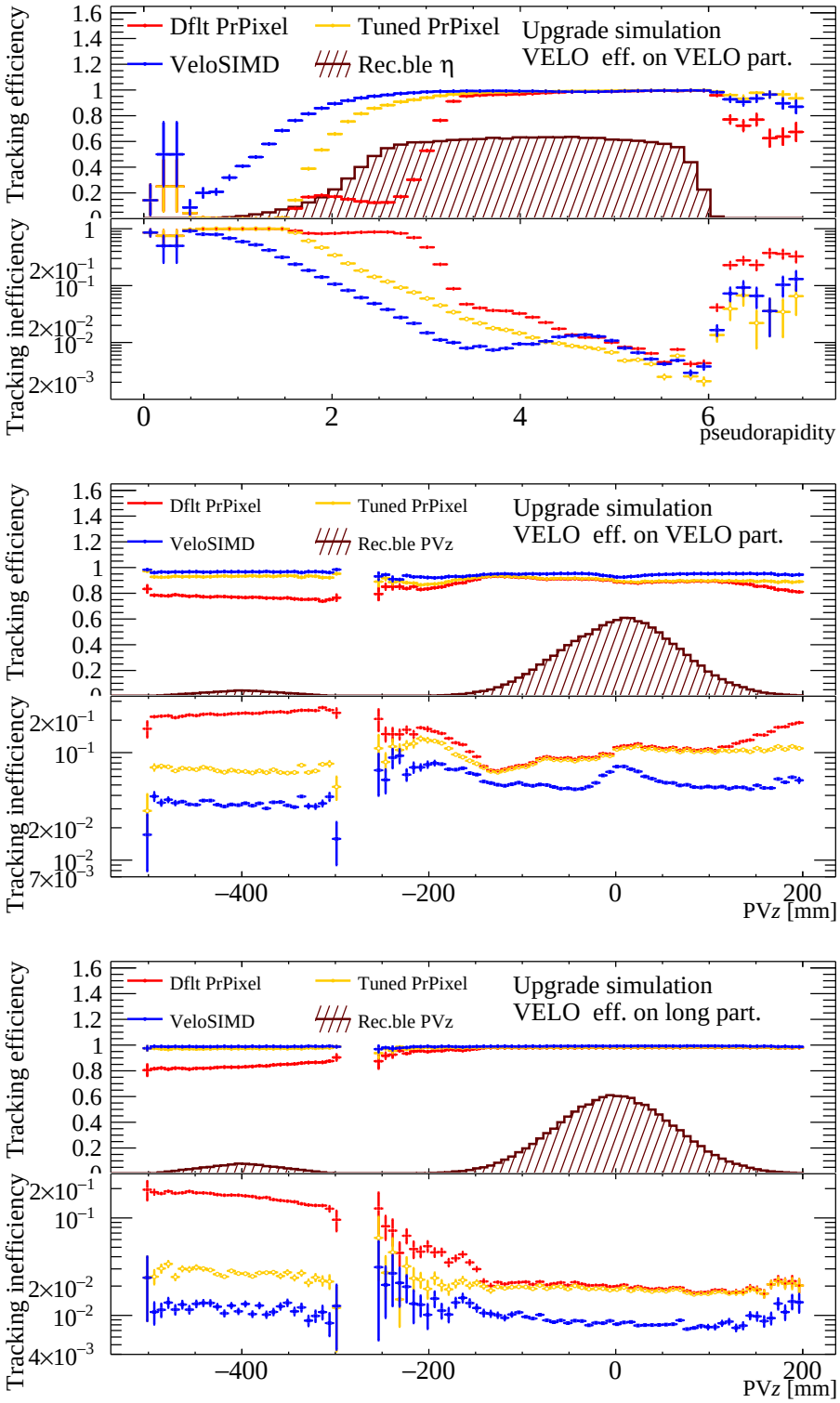


Figure IV.3.5: Tracking efficiency and inefficiency with the default (in red) and tuned (in orange) implementations of the *PrPixelTrackingFast* and with the *VeloClusterTrackingSIMD* (in blue) algorithms. The top plot shows the efficiencies for particles reconstructible as VELO tracks in stand-alone pHe collisions as a function of their pseudorapidity, the middle (bottom) one for particles reconstructible as VELO (long) tracks in $pp + p$He collisions as a function of the z coordinate of their primary vertex.

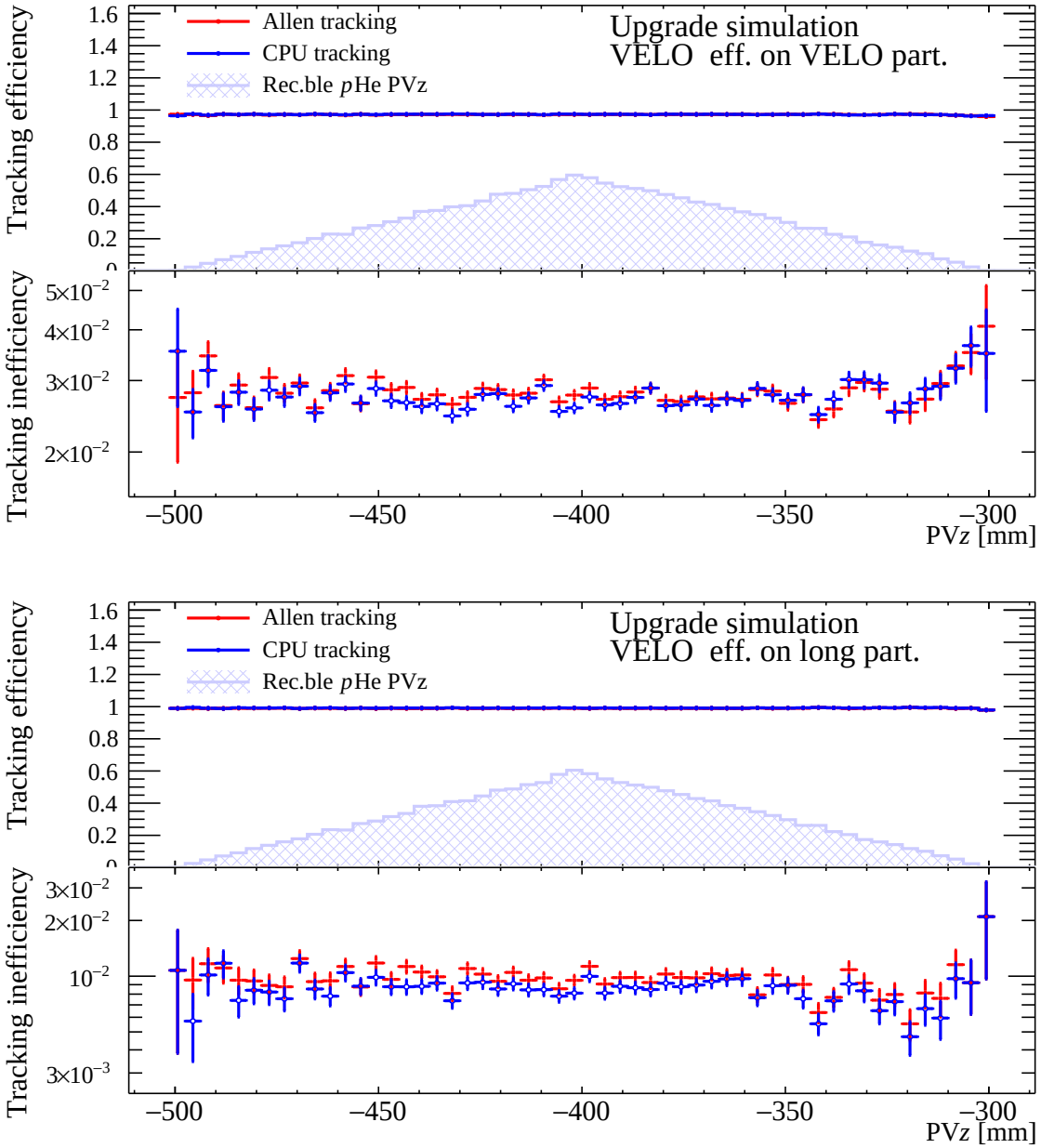


Figure IV.3.6: Tracking efficiency and inefficiency for the VELO reconstruction with the *VeloTrackingSIMD* (in blue) and the *Search by triplet* (in red) algorithms for particles reconstructible as VELO (top) and long (bottom) tracks in stand-alone pHe collisions.

the pseudorapidity for particles reconstructible as VELO tracks in stand-alone pHe (in green) and pp (in blue) collisions, together with the distributions arbitrarily scaled. A different performance can only be seen for low-pseudorapidity beam-gas events, which is expected considering the position of the cell and, as a consequence, the averaged lower number of track VELO hits for low-η tracks. Middle (bottom) plot compares the efficiencies for reconstructible VELO (long) tracks coming from stand-alone pHe (in green), pp (in blue) and overlapped $pp + p$He (in red) or $pp + p$Ar (in orange) collisions as a function of the z coordinate of the primary vertex. The same performance is obtained

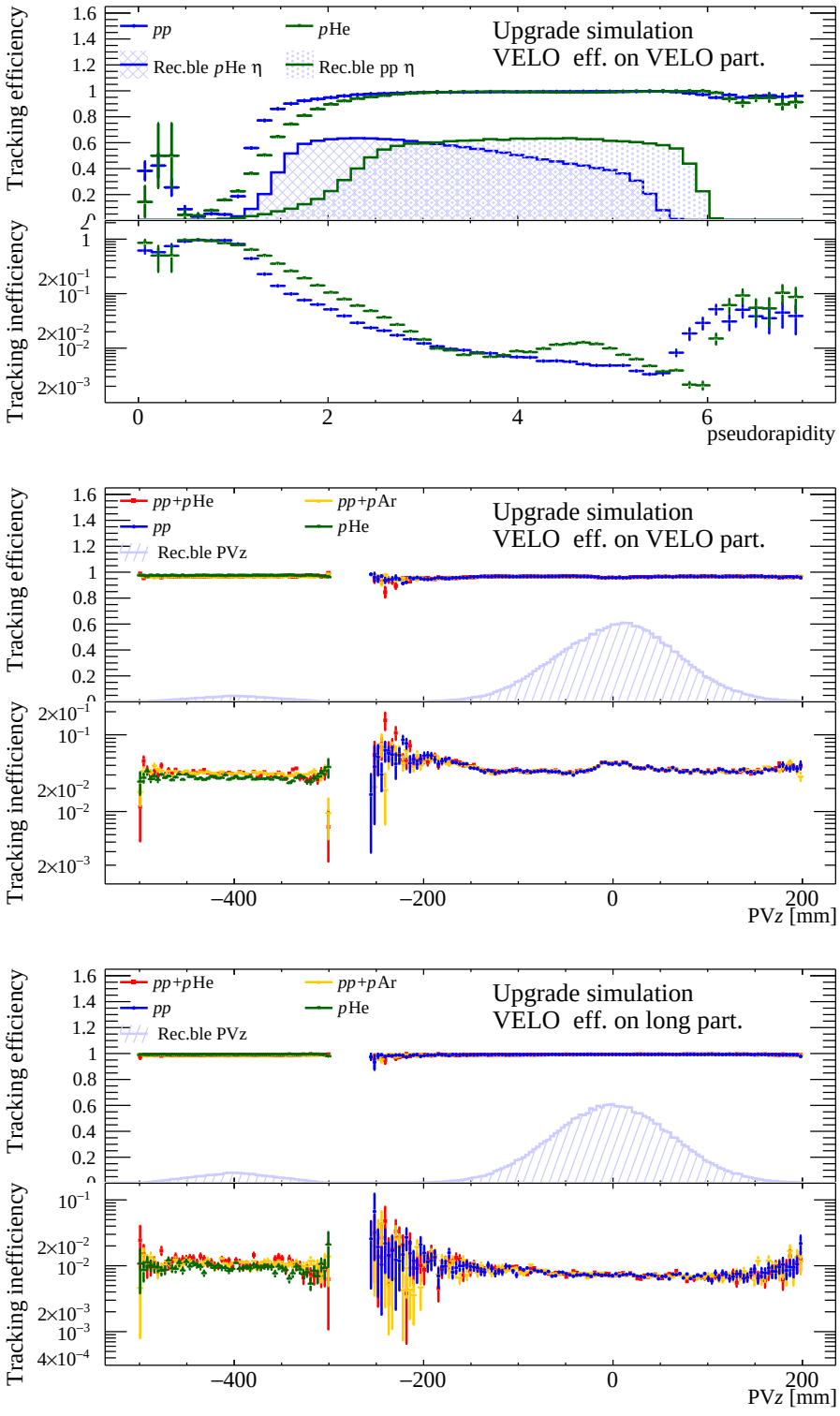


Figure IV.3.7: Tracking efficiency and inefficiency for the VELO track reconstruction with the *SearchByTriplet* algorithm. Top and middle plots are for all particles reconstructible as VELO tracks as a function of the pseudorapidity and z coordinate of the primary vertex, the bottom one for particles reconstructible as long.

for beam-gas and beam-beam events. Moreover, by comparing the *pp* component in the stand-alone and overlapped data-taking scenarios, the consistency between the two curves proves the gas injection to not disturb the beam-beam reconstruction. The comparison between the beam-beam and beam-gas reconstruction efficiencies and ghost rates are listed in the first row of Tab. IV.3.3 and are found to be comparable.

IV.3.2.2 The primary vertex reconstruction

The reconstructed VELO tracks feed the algorithm finding the positions and compositions of the visible collision vertices. The PV reconstruction is performed in two steps:

- the seeding, where VELO tracks with a low Distance Of Closest Approach (DOCA) are grouped to form vertex seeds. This is typically conceived to be fully efficient, but a significant fraction of fake seeds can also be accepted;
- the fit, where the PV position is determined with an iterative minimization procedure. At LHCb, already for the Run2 PV reconstruction, Tukey weights [175] (w_i) are employed to define a vertex χ^2 as:

$$\text{Vertex } \chi^2 = \sum_{i=1}^{nTracks} (\chi^2_{\text{IP}})_i \cdot w_i \tag{IV.3.2}$$

$$w_i = \begin{cases} [1 - (\chi^2_{\text{IP}})_i / C_T^2]^2 & (\chi^2_{\text{IP}})_i < C_T, \\ 0 & (\chi^2_{\text{IP}})_i \leq C_T \end{cases} \tag{IV.3.3}$$

being $(\chi^2_{\text{IP}})_i$ the i_{th} track χ^2_{IP} and C_T a configurable parameter. With this method, contributions from secondary or spurious reconstructed VELO tracks, that would shift the reconstructed PV position towards a Secondary Vertex (SV), are minimised.

The development of the PV reconstruction algorithm, called *PatPV3D*, already started in Run2 with the transition to the real-time analysis paradigm. Previously, PVs were reconstructed online with VELO tracks and refitted offline considering long tracks, this making the efficiency measurements challenging. By introducing a simplified Kalman fit of VELO tracks, the refitting was demonstrated to not significantly improve the performance and discarded [176]. The seeding phase of the algorithm consisted in searching for groups of at least four VELO tracks with a DOCA below 1 mm. Seed coordinates were defined as an averaged Position Of Closest Approach (POCA) to all the considered tracks, marked as used. This was iteratively performed in descending multiplicity order, so that, by firstly determining the largest-multiplicity PV positions and only considering non-used tracks, the rate of false reconstructed PVs was minimised. In the fitting phase, a PV proto-position is determined, tracks are assigned to one seed and the vertex χ^2 is obtained according to Eq. IV.3.2.

An intermediate step in the speed-up of the algorithm was called *PatPV3DFuture* [177]. The seeding phase, based on a three-dimensional track extrapolation, was replaced by a simpler and faster cluster search in the z coordinate. This has no impact on the reconstruction performance, considering that the beam-beam PVs are spread in $z \sim 40\,\text{cm}$ but only in $\sim 0.2\,\text{mm}$ for the x and y coordinates. All VELO tracks are firstly extrapolated to the beam axis, assumed to be parallel to z. Their coordinates

of the POCA to the beam and the related uncertainty, proportional to the pseudorapidity, are calculated. Isolate clusters form the PV seeds and are passed to the fitting phase.

To profit from the intrinsically-parallelised SOA event model introduced for VELO tracks, a new algorithm improving the one-dimensional seeding approach and further simplifying the fitting phase, *TrackBeamLineVertexFinder* (TBLV), was designed. The largest difference with respect to *PatPV3D* is in the seeding, following a histogram peak-search approach:

- the POCA coordinates between the beam axis and the VELO tracks are determined;
- a histogram with the tracks POCA z coordinate is filled;
- PV seeds are defined as peak structures in the histogram;
- tracks are assigned to PV seeds.

The TBLV seeding approach avoids combinatorial operations on track pairs, achieving a significant speed-up with respect to the previous algorithm. On the other side, for vertices that are far from the beam, a less efficient extrapolation of the VELO tracks is expected. In the PV reconstruction, some thresholds optimised on pp data are set. Some of them have been found to be relevant with respect to the reconstruction of SMOG2 data and are discussed here in more details. The track extrapolation and the related uncertainty are calculated assuming a straight line model

$$z_{POCA} = z_{\text{trk}} + \frac{tx_{\text{trk}}(x_b - x_{\text{trk}}) + ty_{\text{trk}}(y_b - y_{\text{trk}})}{tx^2_{\text{trk}} + ty^2_{\text{trk}}} \tag{IV.3.4}$$

$$\sigma_{z_{POCA}} = \left[\begin{pmatrix} tx_{\text{trk}} \\ ty_{\text{trk}} \end{pmatrix}^T \begin{pmatrix} \sigma^2_{x,\text{trk}} & \sigma_{xy,\text{trk}} \\ \sigma_{xy,\text{trk}} & \sigma^2_{y,\text{trk}} \end{pmatrix}^{-1} \begin{pmatrix} tx_{\text{trk}} \\ ty_{\text{trk}} \end{pmatrix} \right]^{-\frac{1}{2}},$$

being x_b, y_b the beam axis position and $\{x, y, z, tx, ty, \sigma_{x,y,xy}\}_{\text{trk}}$ the track coordinates, slopes with respect to z and covariance matrix elements. A larger $\sigma_{z_{POCA}}$ is found for tracks with small slopes with respect to z or, equivalently, with large pseudorapidities. By default, only tracks with $\sigma_{z_{POCA}} < 1.5\,\text{mm}$ are accepted to fill the histogram. Instead of adopting a binary histogram filling, each track can contribute to more than one bin. The filling weight is the integral over the bin region of a normal distribution centred in z_{POCA} and with a $\sigma_{z_{POCA}}$ width. Subsequent bins with a larger content than 2.5 are assumed as proto-clusters, separated as clusters in case a significant dip is found between two following proto-clusters edges. In the vertex fit, tracks are associated to vertices according to a one-to-one relation. The largest speed-up in this phase is due to the fact that the track covariance matrix is calculated and inverted only in the extrapolation phase and then cached. The PV position is iteratively determined minimising Eq. IV.3.2 with $C_T = 12$. If the updated PV proto-position differ from the previous one by less than 500 nm and its radial distance from the beam is below a threshold, the PV is accepted. The algorithm, considered in the choice between the CPU and GPU options [146], met the physics and timing constraints.

The algorithm was then adapted to the GPU reconstruction, profiting from the

offered highly intra-event parallelization. The most important difference with respect to the CPU version is that the one-to-one track-vertex association is discarded, since it would break the parallel processing. Rather, a multi-vertex fitter approach [178] is adopted to allow in principle each track to contribute to each vertex. In case one track takes part to two vertices, a lower weight for both is considered.

For all algorithms, the primary vertex reconstruction is monitored according to the following parameters:

- the PV reconstruction efficiency, defined as the ratio between reconstructed and reconstructible vertices. The former are defined as those that are matched to a simulated vertex within 5 times the reconstruction uncertainty in z, the latter as those presenting a minimum number of charged particles, four by default;
- the absolute PV reconstruction efficiency, defined as the previous one but considering at the denominator all simulated vertices;
- the fake PV rate, defined as the fraction of reconstructed vertices that is unmatched to a simulated vertex;
- the PV resolution, defined as the width of the distribution for the distance between the simulated and reconstructed vertex positions;
- the PV fit pulls, obtained as the ratio between the PV resolution and the reconstruction uncertainty. If there is no bias in the PV reconstruction, a normal distribution around zero with a width equal to one is expected.

The differences between the beam-beam and beam-gas vertices are studied on simulation. Those that affect most the PV reconstruction are presented in Figs. IV.3.8 and IV.3.9 for reconstructed VELO tracks and simulated primary vertices in pHe (in green) and pp (in blue) stand-alone collisions, respectively. The former compares the distributions that are considered in the extrapolation of the VELO tracks to the beam. The top and middle left plots present the tx and ty slopes and the pseudorapidity and, as expected, tracks originating from fixed-target collisions have a lower radial aperture with respect to beam-beam. Such a difference makes the primary vertex reconstruction more challenging, considering that the extrapolation to the beam axis is affected by larger uncertainties with the increasing of the pseudorapidity. The comparison between the VELO track POCA z to the beam is presented integrated and in pseudorapidity bins in the middle right and bottom plots, respectively. Most of the reconstructed tracks originates in the respective interaction regions, but tails are also observed. For tracks originating in pp collisions, these are mostly of high pseudorapidities and are hence expected to only minimally contribute to the reconstruction of primary vertices in SMOG2, considering the uncertainty in their extrapolations. Low-pseudorapidity SMOG2 VELO tracks could instead contribute to the beam-beam PV reconstruction. Nevertheless, when considering the numbers of reconstructed tracks in beam-beam and beam-gas collisions, the effect is expected to be minor. Figure IV.3.9 shows the PV features affecting the reconstruction. Top plots show the normalised distributions for the vertex multiplicity (left) and radial distance with respect to the z axis (right) and, as expected by the lower collision energy and the cell positions, different distributions are found between stand-alone pp (in blue)

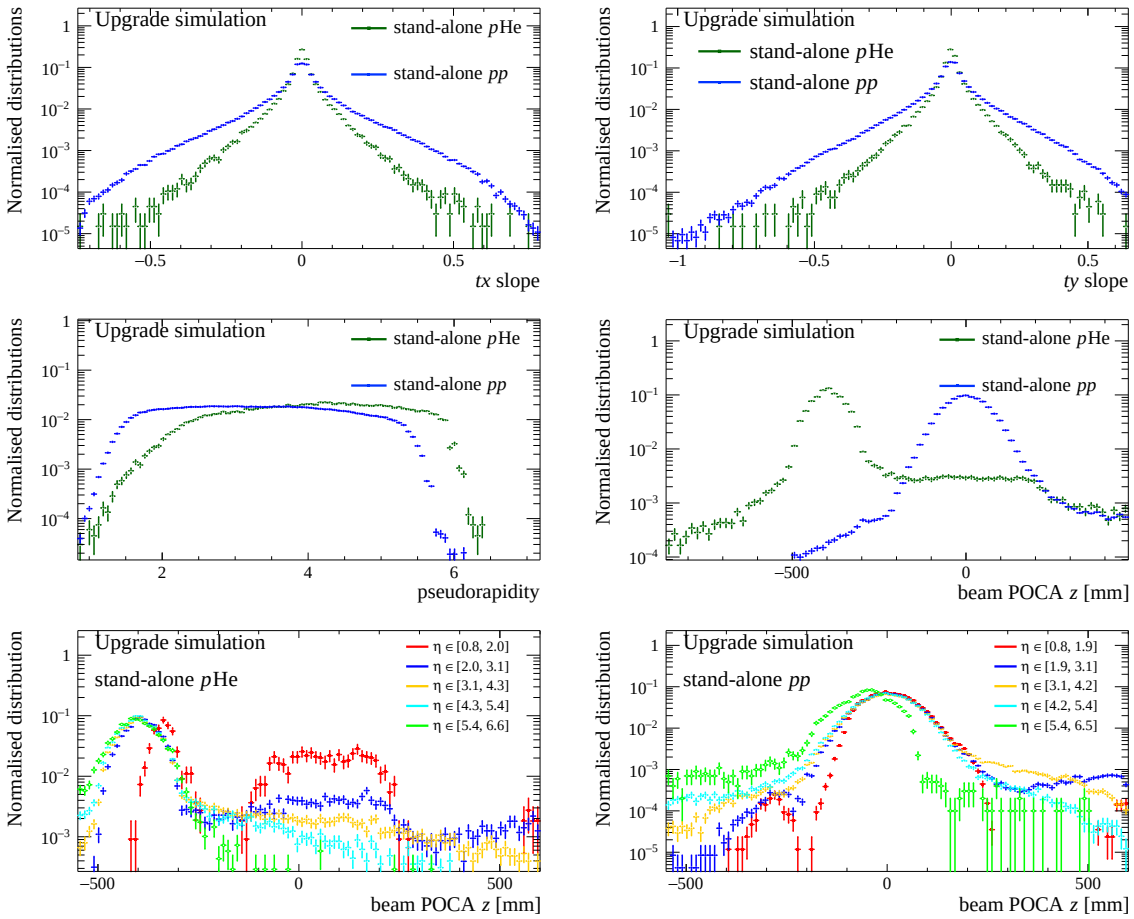


Figure IV.3.8: Comparison between the tx (top left) and ty (top right) slopes, the pseudorapidity (middle left) and z coordinate of the POCA to the beam (middle right) for reconstructed VELO tracks in stand-alone pHe (in green) and pp (in blue) collisions. The bottom plots present the POCA z dependence on the pseudorapidity for the pHe (bottom left) and the pp (bottom right) simulated sample.

Table IV.3.2: TBLV thresholds optimising beam-beam and beam-gas efficiencies.

Requirements	beam-beam	beam-gas
histogram z range [mm]	[-300, 300]	[-500,-300]
minimum number of tracks in the vertex	4	3
minimum track integral contributing to the cluster	2.5	1.75
maximum $\sigma_{z_{POCA}}$ to add a track to the histogram [mm]	1.5	10

and pHe (in green) collisions. The effect of the beam tilting is studied as a possible source of bias in the SMOG2 vertex reconstruction, considering that a maximum PV distance from the beam is imposed. By comparing the PV radial distance distributions in equally-populated z bins for $z \in [-500, -300]$ mm, only a minor effect is observed.

When my *Ph.D.* activities in preparation to the SMOG2 data-taking started, I contributed to test the *PatPV3DFuture* algorithm, with the efficiency shown as

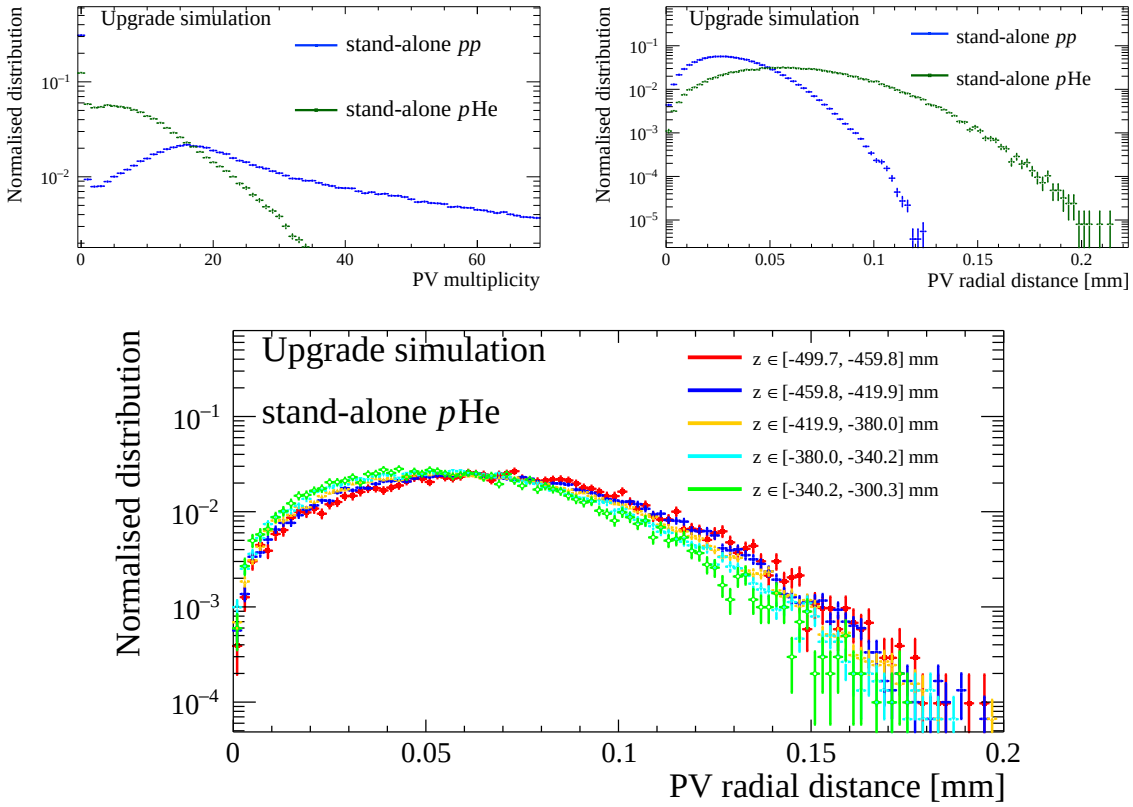


Figure IV.3.9: Comparison for the distributions of the PV multiplicity (top left) and radial distance with respect to the z axis (top right) for simulated stand-alone pp (in blue) and pHe (in green) collisions. The bottom plots shows the correlation between the pHe radial and longitudinal coordinates and demonstrates the beam inclination far from the nominal interaction point to only induce a minor effect.

the red curves in Figs. IV.3.10 and IV.3.11 for simulated stand-alone pHe and overlapped $pp + p$He collisions, respectively. Both figures present the PV reconstruction efficiency considering at the denominator the reconstructible (left) or all (right) simulated vertices. A flat evolution is observed with respect to the longitudinal and radial coordinates while, as expected, a decreasing one for low-multiplicity vertices. Considering the comparisons presented in Fig. IV.3.9, the low-multiplicity worse performance explains the averaged $\sim$ 10% lower efficiency for beam-gas events with respect to beam-beam. The PV reconstruction is repeated requiring a minimum of three tracks instead of four to accept a beam-gas vertex, as it was for Run2 data. The results are presented as the blue curves of Figs. IV.3.10 and IV.3.11. The absolute reconstruction efficiency for pHe collisions increases, that only considering reconstructible vertices decreases because of a larger number of low-multiplicity vertices. The algorithm was then replaced by TBLV, offering a higher output data rate. Its performance with the default values of the parameters (except the minimum z coordinate, which is lowered to -500 mm to also include the SMOG2 region) is illustrated as the orange curves of Figs. IV.3.10 and IV.3.11. In contrast to *PatPV3DFuture*, a clear dependence on the PV_z coordinate is found and the difference in reconstruction performance between beam-beam and beam-gas data is increased. Considering the comparisons presented in Fig. IV.3.8, the VELO tracks extrapolation to

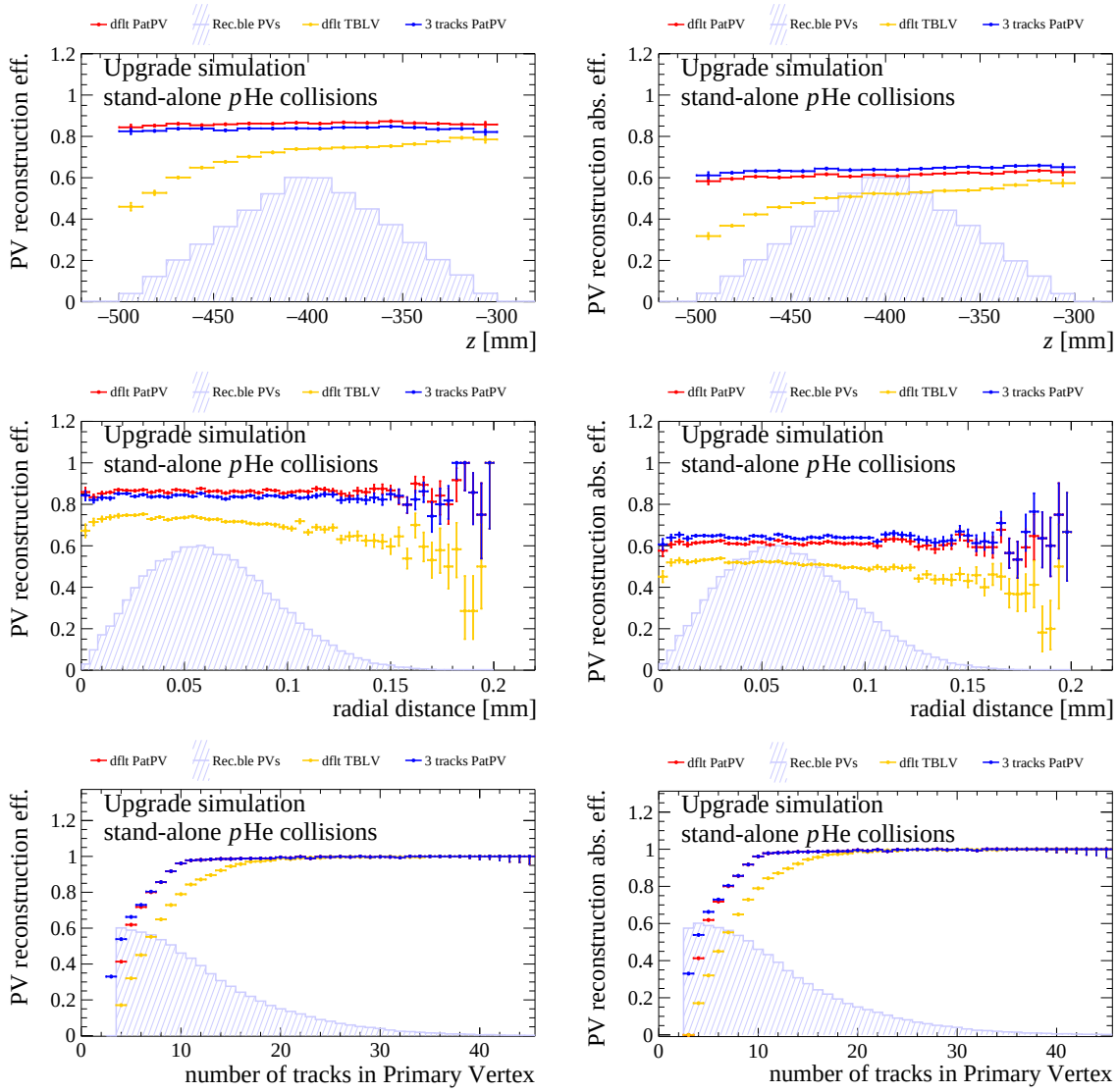


Figure IV.3.10: Vertex reconstruction efficiency as a function of the PV longitudinal (top) and radial coordinates (middle) and of the vertex multiplicity (bottom) for simulated stand-alone pHe collisions with the default implementation of the *PatPV3DFuture* (in red) and of the *TrackBeamLineVertexFinder* (in orange) algorithms and with *PatPV3DFuture* but requiring minimum three tracks instead of four for the reconstruction (in blue). All left plots present the efficiencies with respect to reconstructible vertices only, the right ones with respect to all simulated vertices.

the beam was firstly studied to recover the beam-gas vertex reconstruction inefficiency. Figure IV.3.12 shows the superimposition of the correlation between the VELO tracks POCA z coordinate and $\sigma_{z_{POCA}}$ in beam-gas (in green) and beam-beam (in blue) data. The two populations can be clearly distinguished and, as shown as the red lines in the figures, $\sim 70\%$ ($\sim 40\%$) of the beam-beam (beam-gas) tracks are selected by the default selection cut $\sigma_{z_{POCA}} < 1.5\,\text{mm}$. The different efficiency, combined with the averaged lower beam-gas PV multiplicity, explains most of the beam-gas reconstruction inefficiency observed in Fig. IV.3.11 and its decreasing evolution upstream the nominal interaction point. A looser threshold, indicated in Tab. IV.3.2 and shown in violet in the top plot

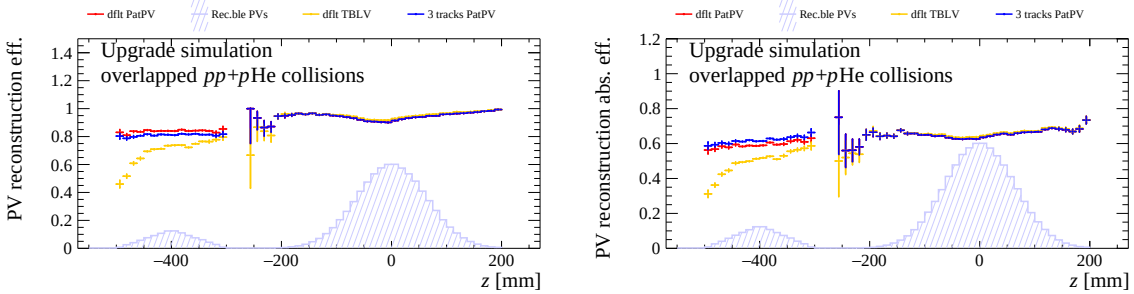


Figure IV.3.11: Vertex reconstruction efficiency as a function of z for simulated pp+pHe collisions with the default implementation of the *PatPV3DFuture* (in red) and of the *TrackBeamLineVertexFinder* (in orange) algorithms and with *PatPV3DFuture*, but requiring a minimum of three tracks instead of four for the reconstruction (in blue). All left plots present the efficiencies with respect to reconstructible vertices only, the right ones with respect to all simulated vertices.

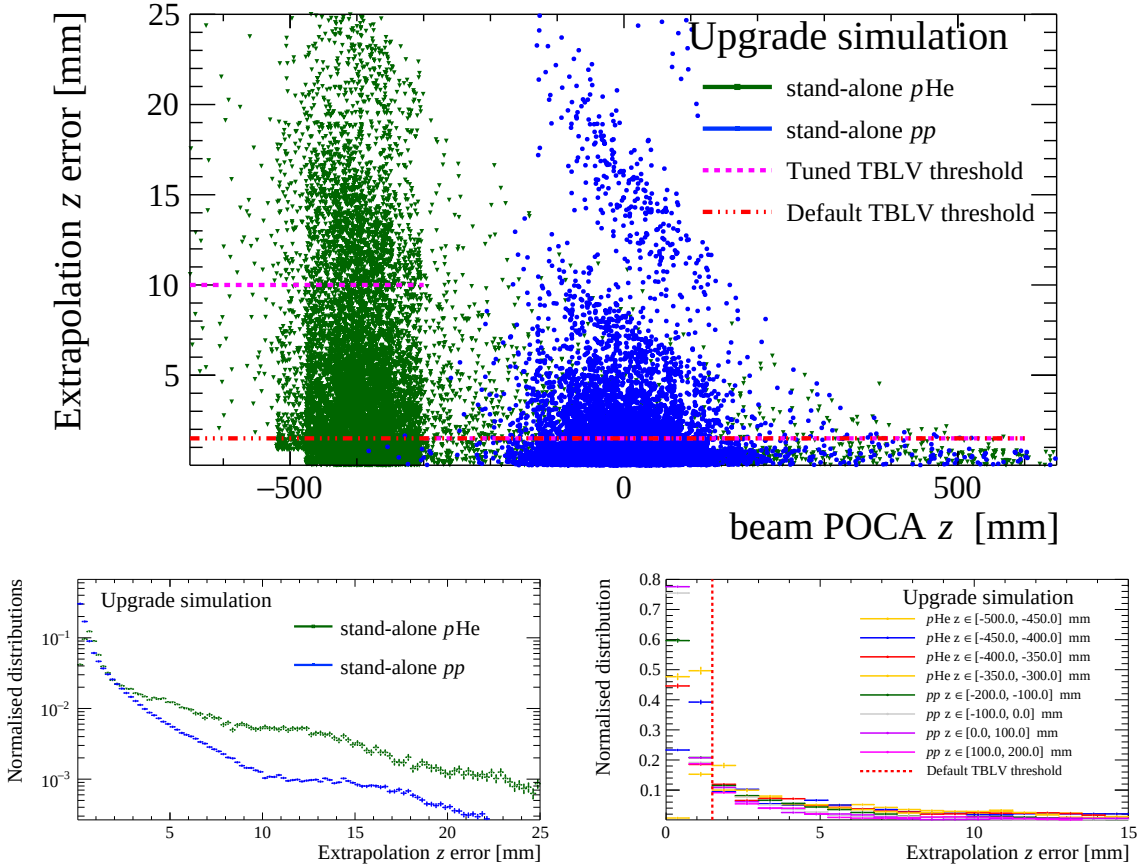


Figure IV.3.12: Uncertainty on the POCA to the beam z coordinate for the PV reconstruction with the *TrackBeamLineVertexFinder* algorithm. Top plot presents the correlation with the z coordinate of the track POCA to the beam and the beam-gas (in green) and beam-beam (in blue) populations can be clearly distinguished. The bottom plots compare it integrating (left) and in z bins (right).

of Fig. IV.3.12, is hence considered for beam-gas data. The threshold value is chosen imposing a similar selection efficiency as beam-beam vertices. The histogram filling and the clustering phases were then considered and, following the lower PV multiplicity of beam-gas events, looser requirements on the minimum cluster size and on the fit χ^2/ndf have been set, as listed in Tab. IV.3.2. All tunings are only applied for $z < -300\,\text{mm}$, in order to respect the criterion of the minimum interference with the beam-beam programme. The vertex fit χ^2/ndf threshold was also considered as a possible source of inefficiency, but the default value has finally been kept to avoid a too large increase of the PV fake rate.

After the tuning of the TBLV thresholds, the resulting reconstruction efficiency is shown as a function of the PV multiplicity (top) and of the radial coordinate (middle) for the pHe (in green) and pp (in blue) simulations and as a function of the z (bottom) for all samples in Fig. IV.3.13. Almost the full inefficiency on beam-gas events is recovered and the same performance as beam-beam is achieved. The plot also shows the efficiency on the overlapped simulated samples and, as for the reconstruction of VELO tracks, the comparison between the results for pp and $pp + p$He or $pp + p$Ar collisions demonstrates there is no efficiency loss because of the gas injection. For the SMOG2 component, better results are observed for injected Ar than for He as a consequence of the averaged larger PV multiplicity when considering a heavier gas.

Figure IV.3.14 presents the achieved results for the primary vertex resolution. The top plot compares the integrated resolution for stand-alone pHe and pp collisions, the bottom one the evolution with the z coordinate for all simulated samples. A worse resolution for beam-gas collisions is observed, but, differently than for the efficiency, this is a geometric effect related to the cell position and can not be solved with a tuning. Nevertheless, considering the expected rate of collisions in the cell, the vertex resolution is not a critical requirement for the SMOG2 physics programme. As for the efficiency, the comparison between He and Ar injection gives a better resolution for the latter, which also is a consequence of the larger averaged occupancy. Figure IV.3.15 presents the achieved results for the primary vertex reconstruction fake rate, monitored together with the absolute reconstruction efficiency for different threshold configurations. The bottom plot presents the comparison among the fake rates on all the simulated samples and, despite the loosening of the vertex reconstruction, the obtained beam-gas rate is tolerable. The combined effect of the gas addition and of the algorithm tuning is studied by comparing the default reconstruction of the stand-alone beam-beam and the tuned one of beam-beam vertices in the $pp + p$He simulated sample. As shown in Fig. IV.3.16, no significant effects are found on the efficiency (top left), the resolution (top right), the pull (bottom left) and the fake rate (bottom right).

With the described VELO track and primary vertex reconstructions, the resolution on the IP, which is used to discriminate between primary and secondary particles, is studied and shown in Fig. IV.3.17 as a function of the inverse of the transverse momentum for tracks in the stand-alone pHe (in green) and pp (in blue) sample. As a consequence of the resolution on the PV position, also that for the IP is worse for pHe than for pp tracks, but the evolution with the kinematics is found to be the same.

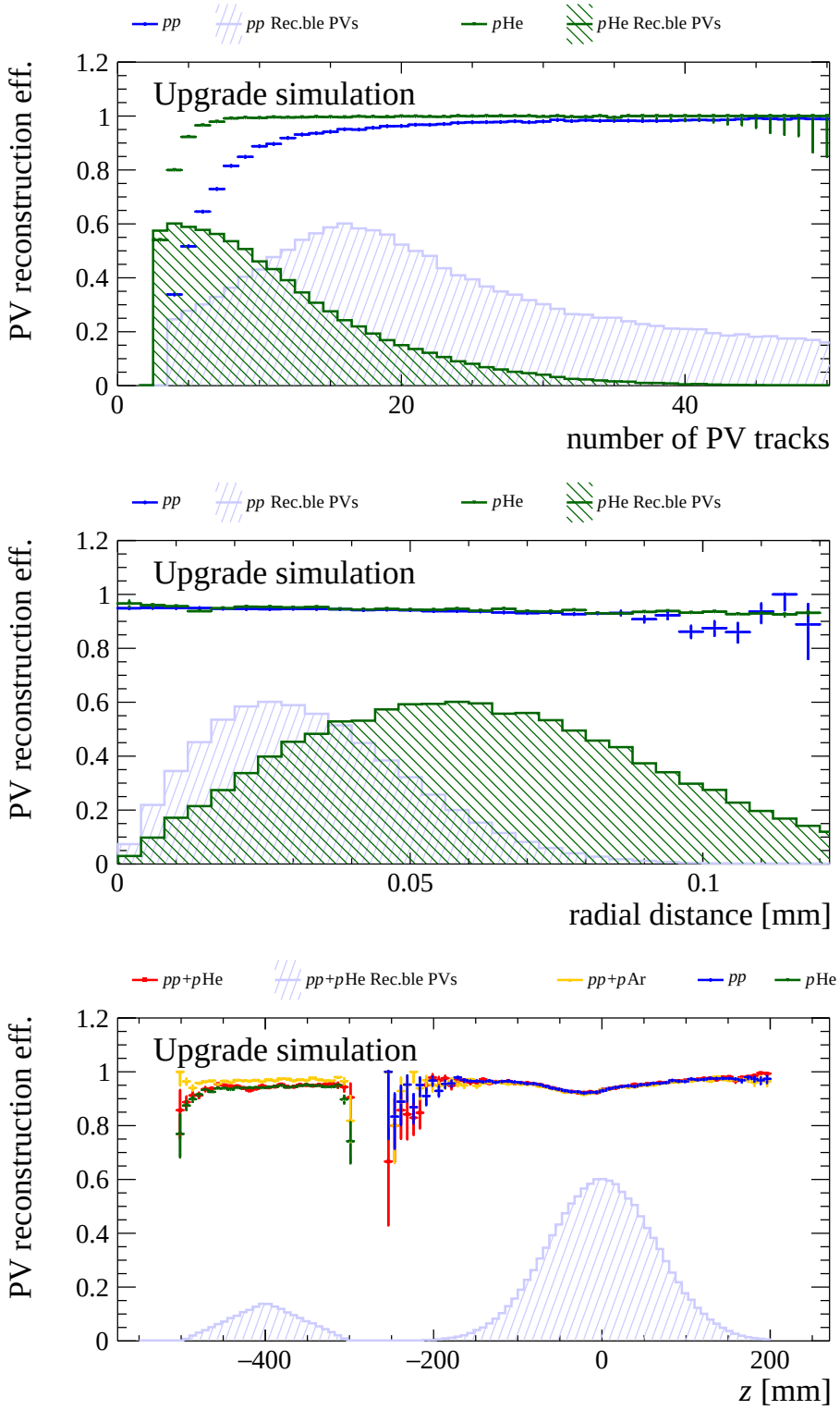


Figure IV.3.13: Efficiency for the pHe (in green), pp (in blue), $pp + p$He (in red) and $pp + p$Ar (in orange) primary vertex reconstruction with the *TrackBeamLineVertexFinder* algorithm after the tuning of some internal thresholds as a function of the vertex multiplicity (top) and of the radial (middle) and z coordinate (bottom).

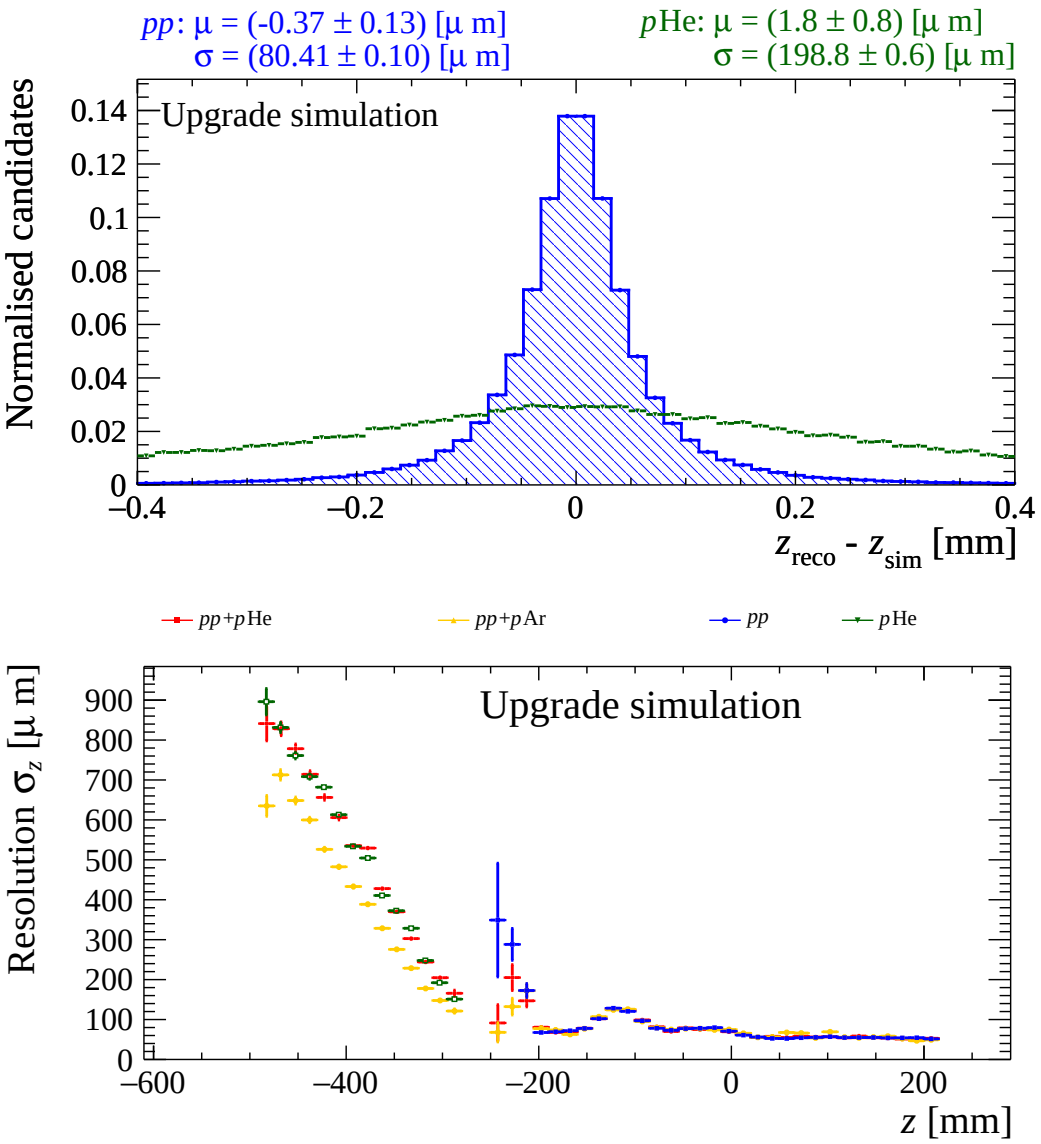


Figure IV.3.14: Total (top) and evolution with the z (bottom) of the resolution for the pHe (in green), pp (in blue), $pp + p$He (in red) and $pp + p$Ar (in orange) primary vertex reconstruction with the *TrackBeamLineVertexFinder* algorithm after the tuning of some internal thresholds.

IV.3.2.3 The long track reconstruction

The VELO segments are extrapolated forward looking for corresponding energy deposits in the UT and SciFi detectors. For both steps, the dependence on the origin z is expected to be less important than for the VELO and the PV reconstruction, but a worse performance could arise from the different beam-beam and beam-gas pseudorapidity distributions. Thus, also the evolution of the algorithms dedicated to the reconstruction of long tracks has been tested on SMOG2 data.

As for the VELO reconstruction, the tracking in the UT and SciFi detectors comprises a pattern recognition and a fitting task to determine the track candidate and to precisely measure the particle momentum. Considering that the UT tracker position in

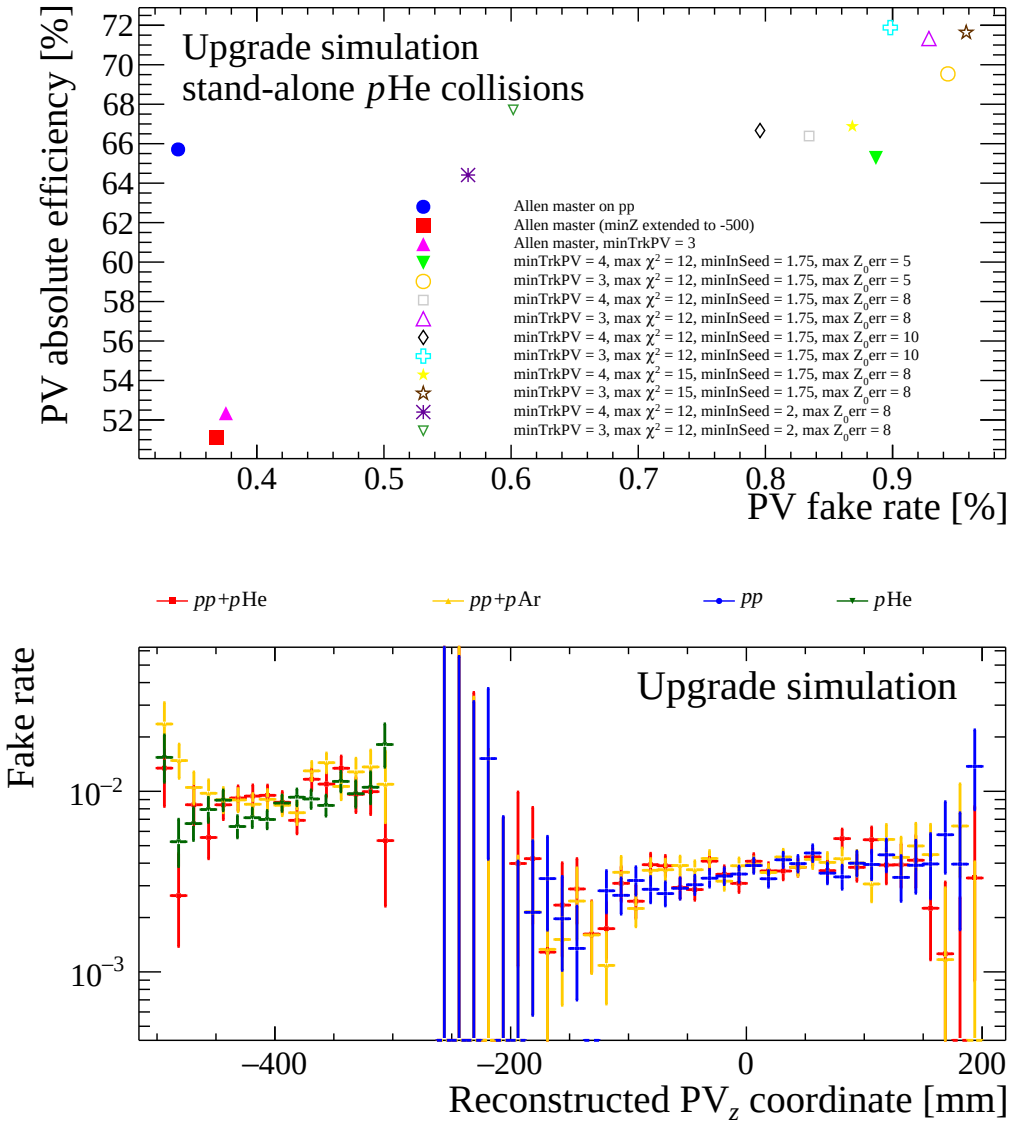


Figure IV.3.15: PV fake rate and absolute reconstruction efficiency for the stand-alone pHe sample *TrackBeamLineVertexFinder* vertex reconstruction with different configurations of the algorithm parameters and comparison among the resulting pHe (in green), pp (in blue), $pp+p$He (in red) and $pp+p$Ar (in orange) samples at the considered working point.

the detector has not changed with respect to the TT stations, the Run2 algorithm was firstly considered and adapted to cope with the new geometry and running conditions (*PrVeloUT* [153]). VELO tracks are linearly extrapolated and paired to the energy deposits in the UT detector according to tolerance windows. Multiplets are formed taking into account possible missing hits due to the sensors inefficiency. In case more than one track candidate is found, a straight line fit is performed to estimate a χ^2/ndf. To clean from spurious tracks, the candidate with the largest number of hits firstly and with the lower χ^2/ndf then is kept. For each VELO-UT track candidate, a determination of the particle momentum with a 15-20% resolution is available. Considering that the magnetic field is not intense at the UT sensors, about 0.15 Tm, this is obtained by approximating

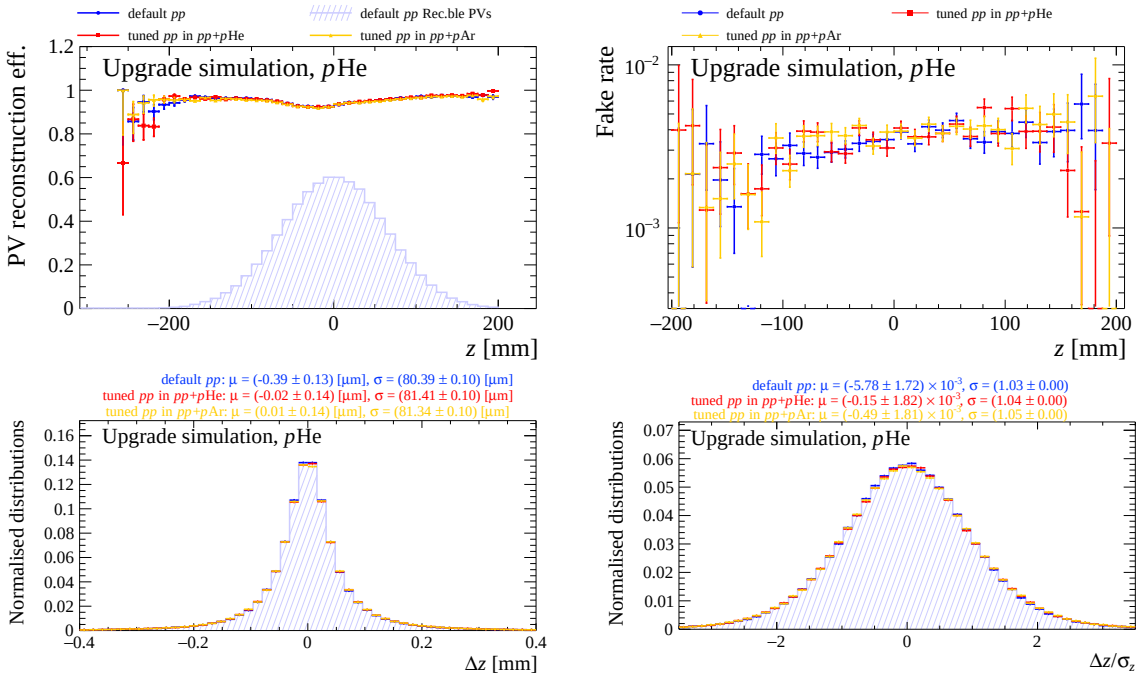


Figure IV.3.16: Combined effect of the tuning of the algorithm and of the gas addition to the beam-beam vertex reconstruction efficiency (top left), resolution (top right), pull (bottom left) and fake rate (bottom right). For both the injection of helium (in red) and argon (in orange), no significant variations with respect to the default reconstruction on stand-alone beam-beam events (in blue) are found.

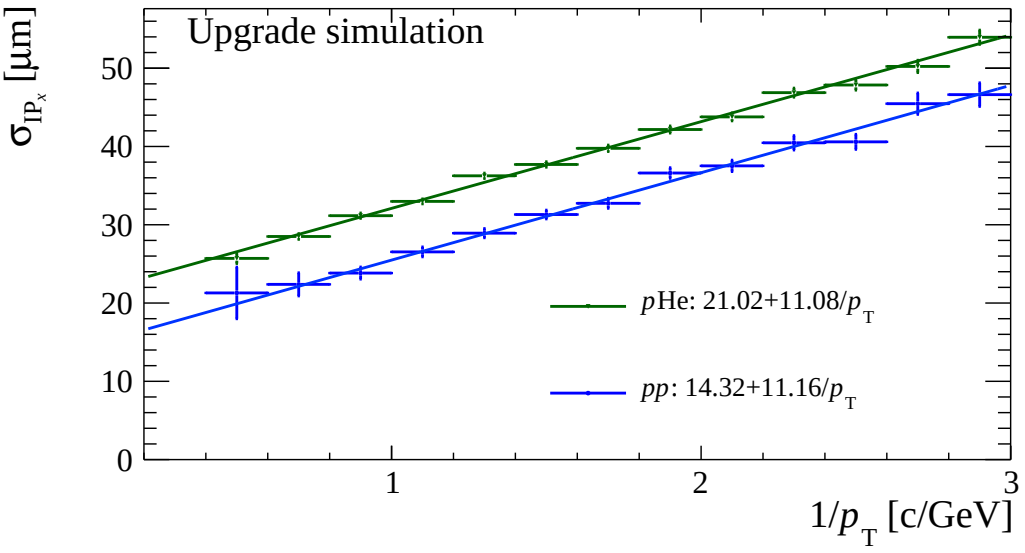


Figure IV.3.17: Resolution on the x coordinate of the impact parameter as a function of the inverse of the transverse momentum for stand-alone pHe (in green) and pp (in blue) collisions.

the bending of the trajectory with a kink and the measured slopes at the VELO and UT z positions are used to calculate the momentum [179]. Similar optimization and parallelization techniques as the VELO tracking algorithm have been applied, without changing the described core idea. The algorithm adaption to the GPU processing, called *UTCompass* [180], adopts the SIMT execution model and all VELO tracks within the UT acceptance are processed in parallel. The binary search is firstly performed in the first and last UT station and, in case a doublet is formed, a tolerance window is opened to look for the intermediate hits. The multiplet is then tested by performing a χ^2 fit, allowing the candidate with the same number of hits to be ordered.

The extrapolation to the SciFi detector follows. In order to reduce the number of processed inputs and to decrease the ghost rate, only VELO-UT tracks with a minimum $3\,\text{GeV}/c$ momentum are processed by default. Considering the momentum estimation provided by the VELO-UT matching, narrow tolerance windows are opened to search for corresponding hits in the SciFi stations, speeding-up the process. The measurement of the track parameters and the rejection of fake tracks is then performed by a Kalman fit [181].

The figures of merit for the long track reconstruction performance are the same as Eq. IV.3.1, requiring reconstructible tracks to present hits in all tracker detectors, according to the definitions listed in Sec. IV.3.1. As in the previous sections, the distributions of the quantities affecting most the track reconstruction are compared between beam-beam and beam-gas data. Figures IV.3.18 and IV.3.19 present the results for the UT and SciFi detectors, respectively, for the number of hits and the pseudorapidity in a bidimensional binning scheme and projected onto the two axes. By comparing with VELO tracks, smaller differences are found and, as a consequence, also a similar performance is expected provided that the kinematic is the same.
The UT and SciFi reconstruction efficiencies with the GPU processing are listed in Tab. IV.3.3 and presented for all samples in Figs. IV.3.20 and IV.3.21, respectively. For both figures, the top plot shows the results for all reconstructible long tracks, the bottom one for the only fraction presenting $p > 3\,\text{GeV}/c$ and $p_\text{T} > 500\,\text{MeV}/c$. As expected, while a $\sim 1\%$ lower efficiency is observed for beam-gas data in the former, the same performance is achieved when considering more similar kinematic intervals. By comparing the results with the employed simulated samples, a per-mill efficiency loss is found for beam-beam data, a $\sim 1\%$ for beam-gas. Both are expected to have a negligible impact on the acquisition of data.

Fig. IV.3.22 shows the ghost rate of the long track reconstruction as a function of the momentum (top) and of the pseudorapidity (bottom) for all simulated samples. As expected, it increases at low and high momentum, where multiple scatterings or the ineffective bending of the magnetic field can occur, and at extreme pseudorapidity values, where tracks are characterised by a lower number of hits. When comparing the different samples, the ghost rate increase due to the gas addition to beam-beam collisions is minor and is not expected to significantly affect the data-taking.
The reconstruction performance for electrons is also briefly studied. As discussed in Sec. IV.2.3, this is not a key requirement for the SMOG2 physics programme, but the measurement of elastic scattering between the LHC beams and the atomic electrons of the injected gases, already demonstrated to be a viable option in Run2, could be employed also in Run3 to independently measure the luminosity of fixed-target samples. As shown

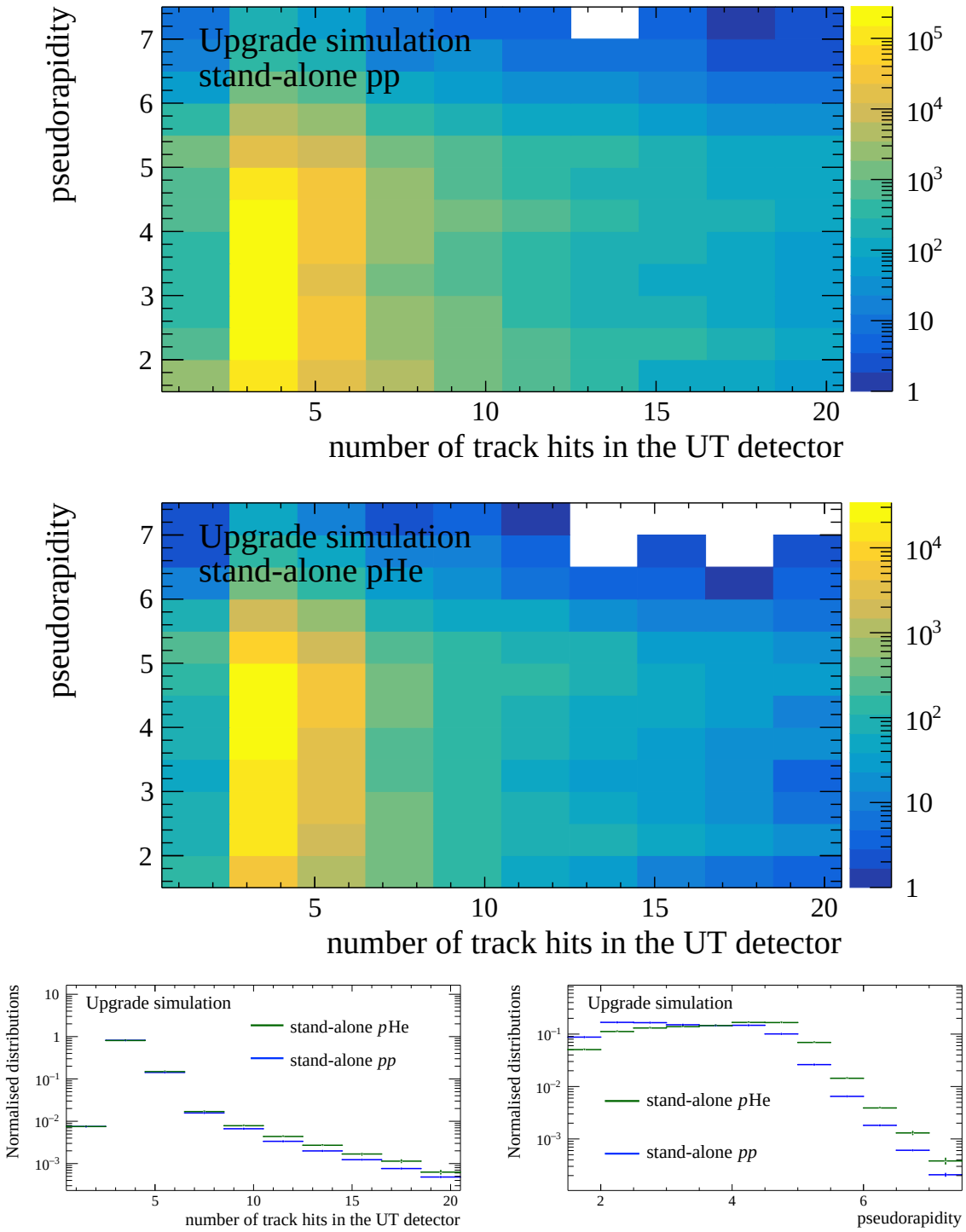


Figure IV.3.18: Comparison between the distributions in simulated stand-alone pp and pHe collisions for the track number of hits in the UT detector and the pseudorapidity. The top and middle plots present the bidimensional distributions for the two samples, the bottom ones the normalised projections superimposed.

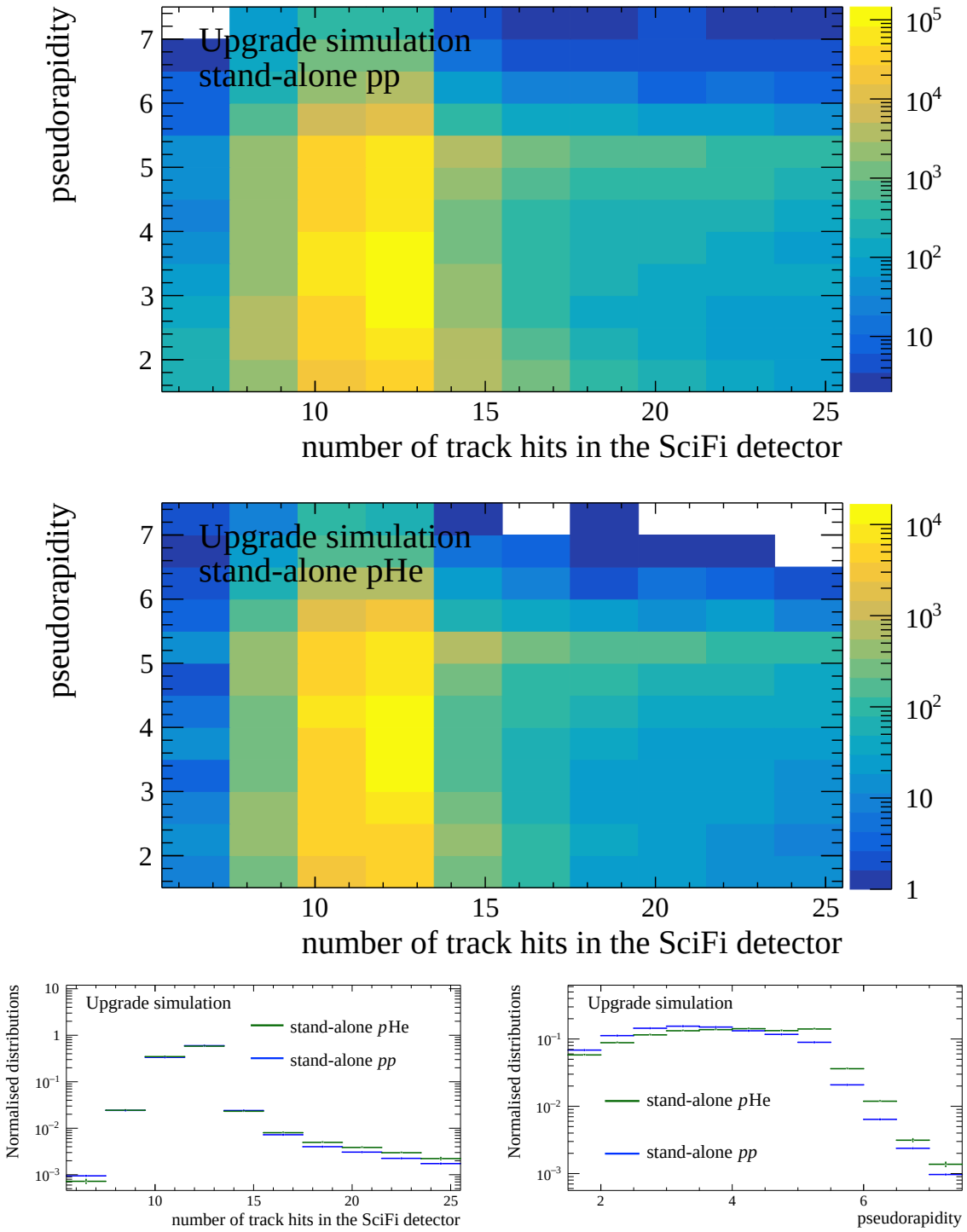


Figure IV.3.19: Comparison between the distributions in simulated stand-alone pp and pHe collisions for the track number of hits in the SciFi detector and the pseudorapidity. The top and middle plots present the bidimensional distributions for the two samples, the bottom ones the normalised projections superimposed.

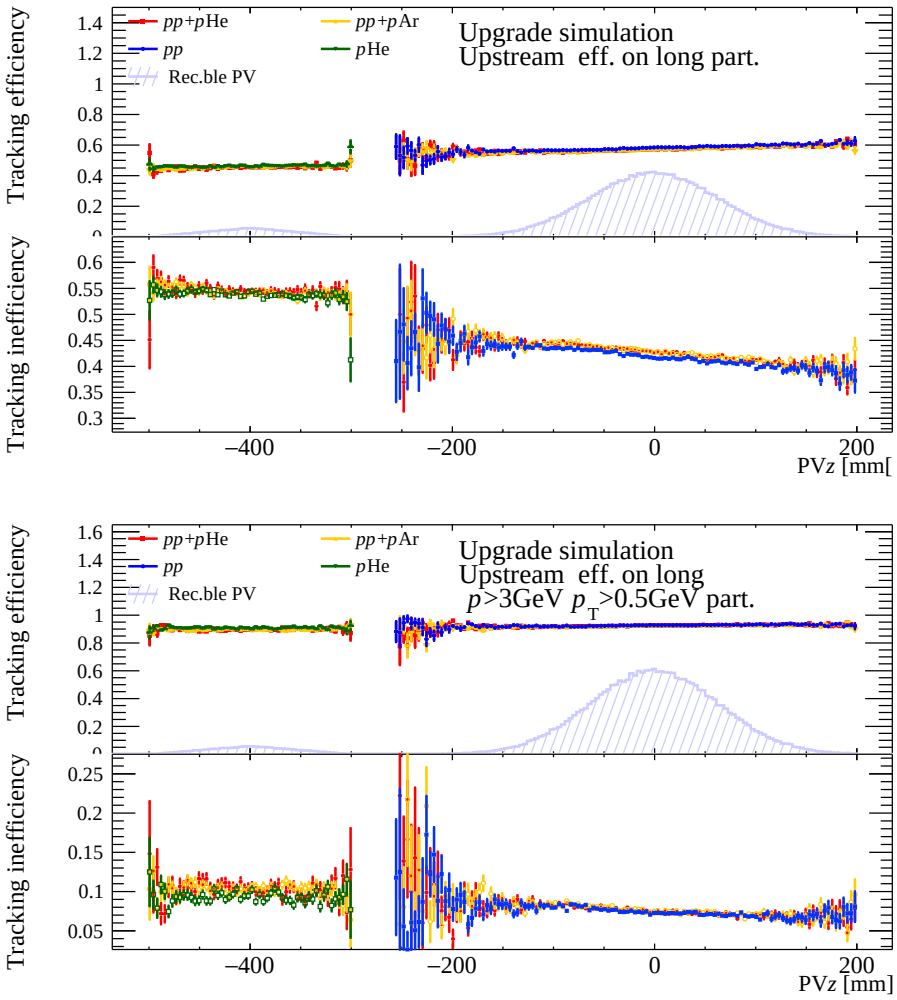


Figure IV.3.20: Tracking efficiency and inefficiency for the UT track reconstruction with the *CompassUT* algorithm as a function of the z coordinate of the primary vertex in the stand-alone pHe (in green) and pp (in blue) and in the overlapped $pp + p$He (in red) and $pp + p$Ar simulated samples. The top plot is for all reconstructible tracks, the bottom for the only ones with $p > 3\,\text{GeV}/c$ and $p_\text{T} > 0.5\,\text{GeV}/c$.

in Fig. IV.3.23, the beam-gas reconstruction efficiency for electrons is found to be lower with respect to beam-beam. Anyhow, being its evolution with the pseudorapidity similar, this is expected to mostly come from the reconstructible distribution. A more detailed study of these differences is left for future development.
Another figure of merit which is studied and compared between stand-alone pHe (in green) and pp (in blue) simulated samples in Fig. IV.3.24 is the resolution on the momentum, presented as a function of the momentum (top) and of the pseudorapidity (bottom). A difference in the track resolution at the per-mill level is found between the two types of collisions, as a consequence of the different track topology and fraction of crossed material, but also this is not expected to significantly affect the SMOG2 data-taking.

Similar studies have been performed for the track reconstruction at the HLT2 level, where,

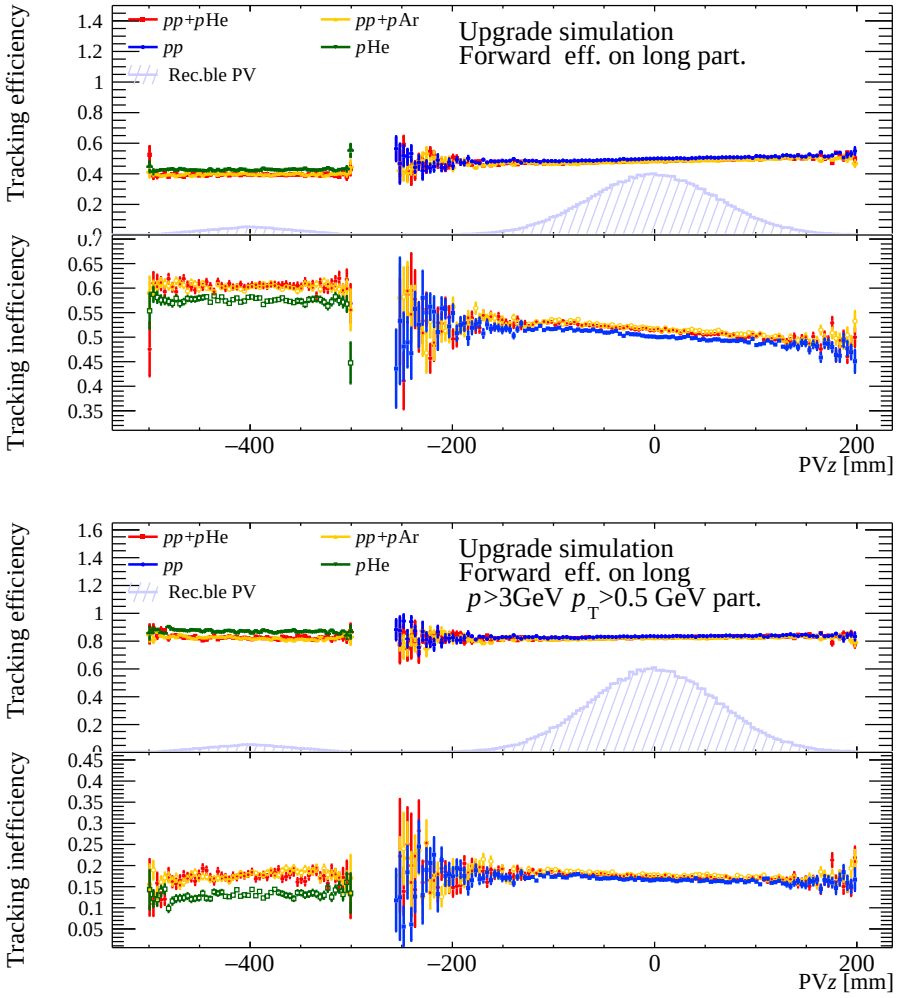


Figure IV.3.21: Tracking efficiency and inefficiency for long tracks with the GPU reconstruction algorithm as a function of the z coordinate of the primary vertex in the stand-alone pHe (in green) and pp (in blue) and in the overlapped $pp + p$He (in red) and $pp + p$Ar simulated samples. The top plot is for all reconstructible tracks, the bottom for the only ones with $p > 3\,\mathrm{GeV}/c$ and $p_\mathrm{T} > 0.5\,\mathrm{GeV}/c$.

Table IV.3.3: Efficiency and ghost rate for the GPU VELO, UT and SciFi tracking for stand-alone pHe and pp collisions.

		pp stand-alone sample [%]	pHe stand-alone sample [%]
VELO	long part. eff.	99.26	99.07
	ghost rate	0.80	0.73
UT	long part. eff.	58.11	46.12
	ghost rate	5.76	1.77
SciFi	long part. eff.	49.71	42.42
	ghost rate	4.70	1.56

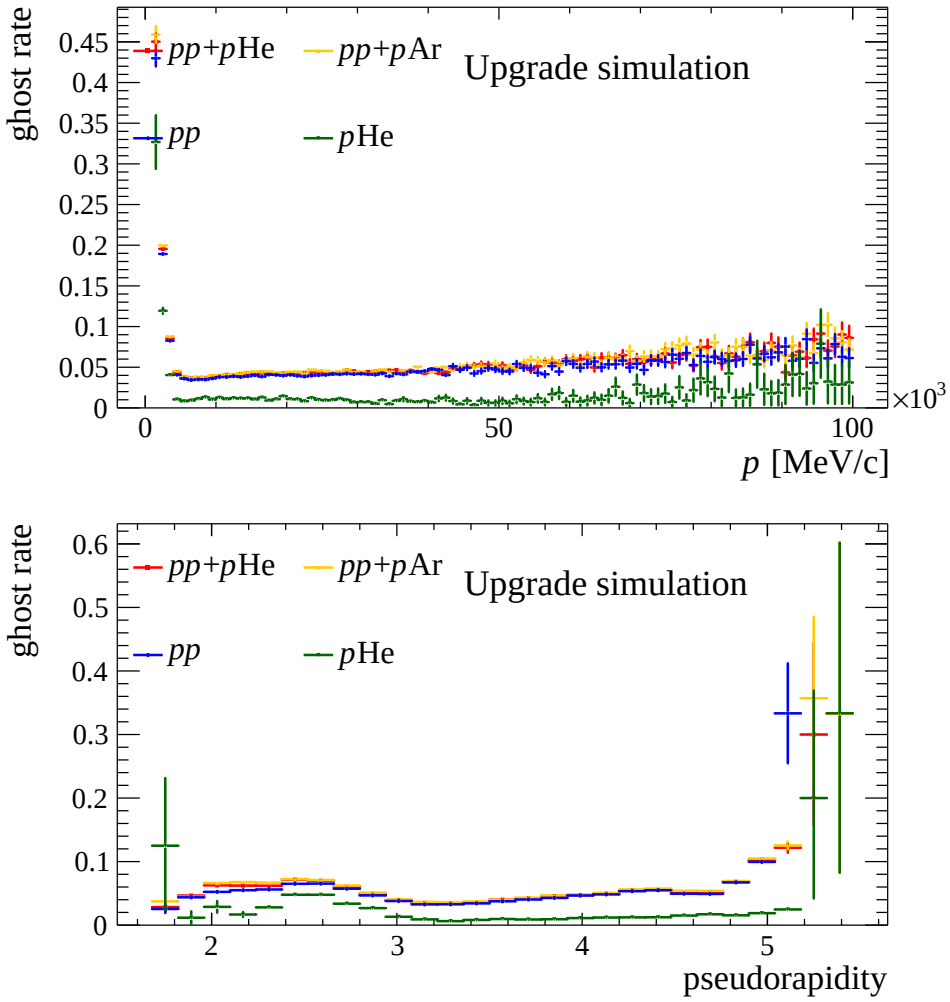


Figure IV.3.22: Ghost rate for the long track reconstruction on GPUs as a function of the momentum (top) and of the pseudorapidity (bottom) in the stand-alone pHe (in green) and pp (in blue) and in the overlapped $pp + p$He (in red) and $pp + p$Ar simulated samples.

as discussed above, a more precise fit of the track is performed. An example of these is presented in Fig. IV.3.25, comparing the tracking performance as a function of the pseudorapidity (top) for the stand-alone samples and for all the considered ones as a function of the z coordinate of the primary vertex (middle and bottom). In all cases, tracks that are employed in physics analyses are efficiently reconstructed independently on their origin in a beam-beam or a beam-gas collision.

IV.3.2.4 The HLT1 reconstruction sequence timing

As mentioned in Sec. IV.3.1, other than an efficient reconstruction performance, a critical requirement to the SMOG2 project and in particular to the possible simultaneous data-taking is that the additional required processing time is tolerable. To investigate this, the full HLT1 reconstruction sequence has been measured on dedicated GeForce RTX 2080 Ti GPU cards, one of the models under consideration for the HLT1 deployment, included in a test-bench that reproduces the proposed layout of the EB farm. The results for the

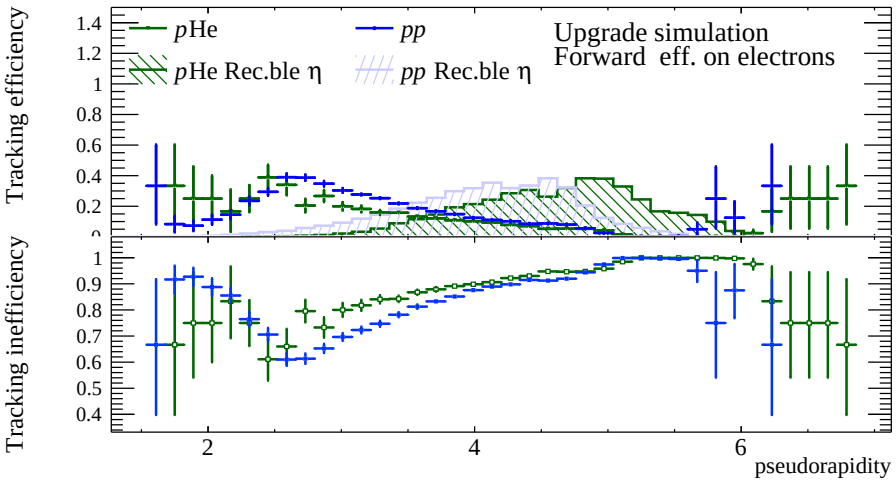


Figure IV.3.23: Reconstruction efficiency and inefficiency with the GPU algorithms as a function of the pseudorapidity for electrons from stand-alone pHe (in green) and pp (in blue) collisions.

stand-alone pp simulation reconstructed with the default implementation of all algorithms and for overlapped $pp + p$He and $pp + p$Ar samples after the described optimizations are compared. Two configurations, both being explored for Run3 pending the physics requirements and the available data output rate, have been considered:

- *HLT1_pp_default*, where only particles with $p > 3\,\mathrm{GeV}/c$ are extrapolated as long tracks;
- *HLT1_no_ut_cut*, with no applied thresholds.

To account for the fluctuations, each measurement is repeated 30 times considering a configuration with 16 threads and 1000 repetition of the same 500 events for each thread. The measured event processing rate on the stand-alone pp and on the overlapped $pp+p$He and $pp+p$Ar samples for both sequences are listed in Tab. IV.3.4. The measured difference is divided by a factor five, to obtain the expected performance for a reference value $\mu = 0.2$, while each simulated event contains one beam-gas collision. The preliminary results show that the gas injection in parallel to the beam-beam acquisition only induces a $\sim 1.5-3.5\%$ decrease to the event rate. The processing rate decrease, if needed, can be mitigated by decreasing the gas injected pressure and, consequently, the beam-gas μ.
A second measurement of the data rate is performed with the nvproof [182] tool in order to understand which algorithms in the reconstruction sequence are the most time-consuming. The results for the default GPU reconstruction sequence running on stand-alone pp (top) and pHe (bottom) data are presented in Fig. IV.3.26. Beside the memory transfers between the GPU and CPU devices, that are dominant in both cases, pattern recognition is the most time-consuming task for pp data. For pHe, instead, the occupancy in the detector is lower, and combinatorial algorithms are faster.

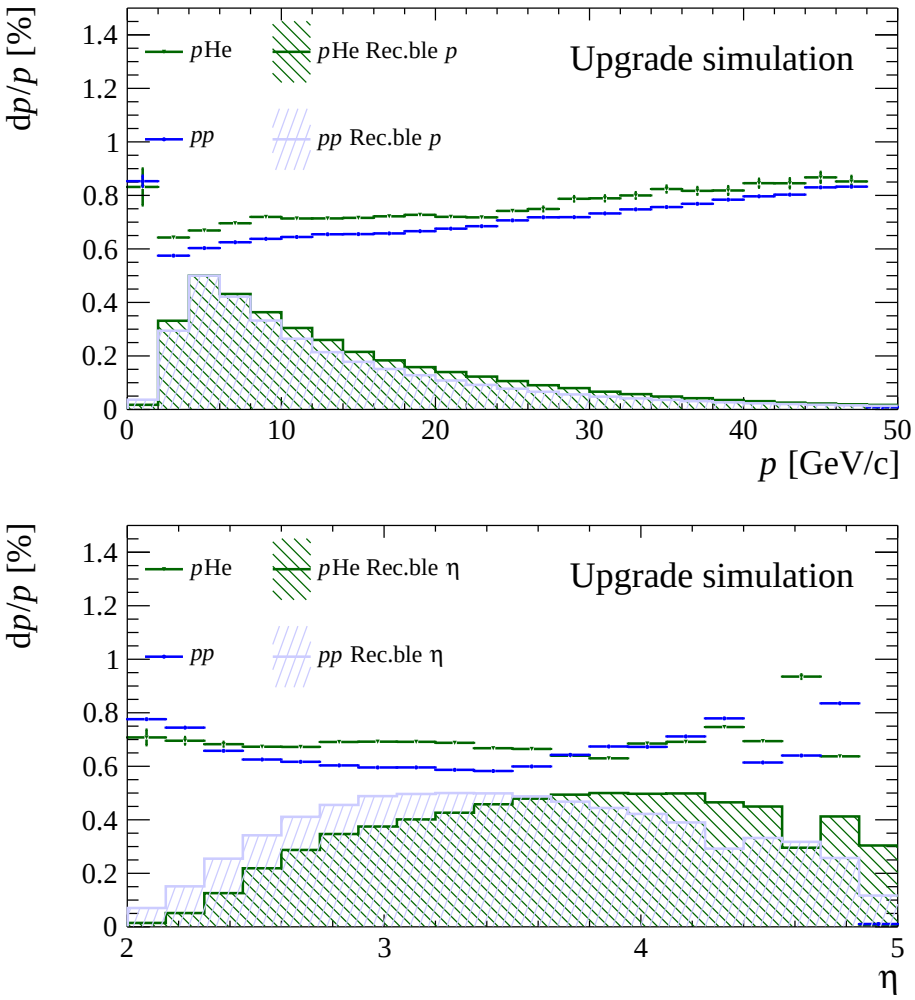


Figure IV.3.24: Percent resolution on the momentum as a function of the momentum (top) and of the pseudorapidity (bottom), with the distributions also shown arbitrarily scaled, for the stand-alone pHe (in green) and pp (in blue) simulated samples.

IV.3.3 The SMOG2 selection algorithms

The challenges and two examples of the implementation of the selection algorithms that are required to fulfil the physics prospects described in Sec. IV.2.3 are discussed in this section. For the beam-beam acquisition, inclusive selections are mostly implemented at the HLT1 level. The large number of selection lines that are required to cover all the LHCb physics programme, $\mathcal{O}$ (1000), would require indeed an unsustainable time to be processed and the information available at the HLT1 is not discriminative enough. Instead, being the beam-gas collision rate and the event complexity lower than beam-beam, also specific two- or three-body decays can be selected with requirements on the invariant mass. As a starting point, the algorithms included in the 2016 and 2017 data-taking, listed in Tab. IV.3.5 and offering a large efficiency on signal candidates and background rejection, have been considered to be ported to the new selection framework. Additional requirements must be set for Run3 to ensure, in case of a simultaneous data-taking, the minimal interference between the acquisition of the two types of collisions.

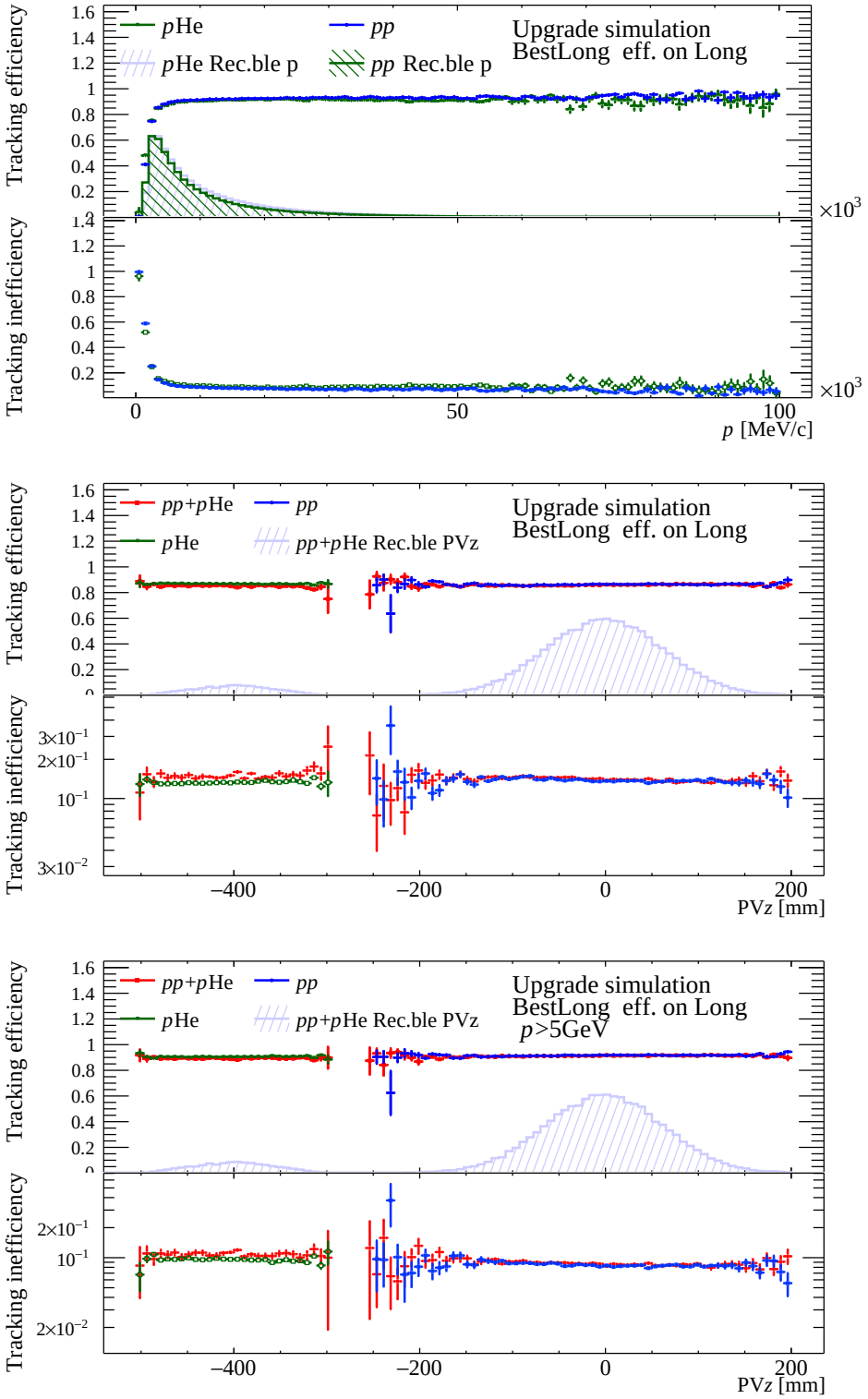


Figure IV.3.25: Offline-quality (HLT2-level) tracking efficiency and inefficiency for all particles reconstructible as long tracks (top and middle) or for their fraction with a larger momentum than 5 GeV/c (bottom) as a function of the momentum (top) for the stand-alone pHe (in green) and pp (in blue) and of the z coordinate of the primary vertex (middle and bottom) for all simulated samples.

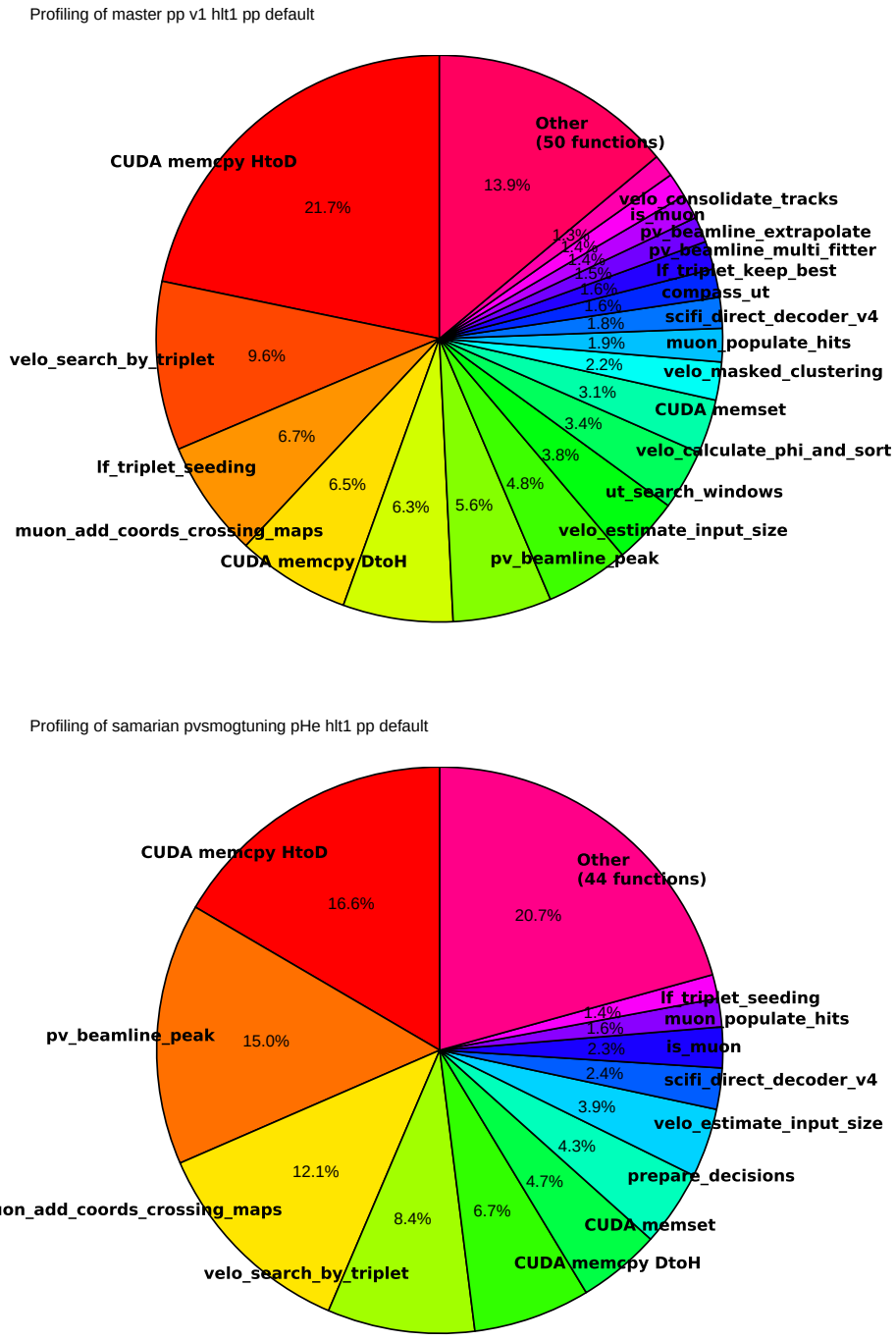


Figure IV.3.26: Comparison between the profiling of the default GPU reconstruction sequence implementation on stand-alone *pp* (top) and with the tuning on *p*He simulated sample (bottom).

Table IV.3.4: Measured output data rate for the *HLT1_pp_default* (top) and the *HLT1_no_ut_cut* (bottom) sequences for the default implementation of the algorithms ran on stand-alone pp and after all tunings on the $pp + p$He and $pp + p$Ar samples. The measured excess is divided by five to scale the simulated one fixed per-event beam-gas interaction to a realistic Run3 beam-gas rate of collision $\mu \simeq 0.2$.

Measurement	Data rate [Hz]	Diff. [Hz]	Diff. ($\mu = 0.2$, [%])
HLT1_pp_default			
pp, master	171385.4 ± 496.5	—	—
$pp + p$He, tuned	158693.4 ± 280.6	12692	-1.48
$pp + p$Ar, tuned	144613.3 ± 334.3	26772	-3.12
HLT1_no_ut_cut			
pp, master	158675.2 ± 347.8	—	—
$pp + p$He, tuned	144593 ± 172	14081.7	-1.77
$pp + p$Ar, tuned	131254 ± 251	27421.2	-3.68

Selection algorithms can be grouped in four categories, according to the object they take decision on:

- *EventLine*, firing on event-level quantities;
- *OneTrackLine*, requiring a Kalman-fitted long track to satisfy some conditions;
- *TwoTrackLine*, considering fitted secondary vertices;
- *CustomLine*, running on a user-defined type of object.

For example, a selection algorithm requiring a maximum number of reconstructed tracks in the VELO will be an *EventLine*, one optimised on $J/\psi \to \mu^+\mu^-$ decays a *TwoTrackLine*. Another key aspect of the selection framework is its multi-event scheduler feature [183], allowing each selection algorithm to run a custom reconstruction sequence. For example, the acquisition of a selection algorithms for the only events satisfying a condition can be imposed. For the beam-gas data-taking, I implemented such filters requiring:

- the presence of a reconstructed primary vertex within a z range;
- a maximum number of VELO reconstructed tracks in the event;
- a beam-empty LHC configuration.

The first filter is implemented twice, with the two ranges $z \in [-500, -300]$ mm and $z \in [-200, 200]$ mm. The former prevents beam-gas lines to acquire data when there is no reconstructed primary vertex in the SMOG2 region. Being the baseline beam-gas rate of collision $\mu \sim 0.2$, this mitigates fake triggers on beam-beam objects. In the same way, beam-beam lines reject all events with no reconstructed vertex around the nominal interaction point, this being unlikely for pp collisions but possible *e.g.* for ion runs. The second filter is used to select photon-induced processes, which are characterised by a

low activity in the detector and, when combined with the third one, for the luminosity measurement through pe^- elastic scatterings. The third filter will be used in case fixed-target data will have to be acquired for the bunches non colliding in LHCb.
A similar strategy has been followed for the HLT2 trigger level, with the main difference that the full event reconstruction information is available. Exclusive selection lines are being implemented for different c- and b- composite particle decays, always requiring the final-state particles to satisfy PID selections and their combination to give an invariant mass compatible with the considered resonance.

IV.3.3.0.1 Minimum-bias selection

As described in Sec. IV.2.3, light hadron production measurements are foreseen with SMOG2. These do not present a specific signature in the LHCb detector and need to be acquired with a selection introducing the minimum possible bias. For example, as indicated in Tab. IV.3.5 and discussed in Sec. I.4.2.2, the 2016 pHe data employed to the antiproton production measurement were acquired only requiring the presence of a reconstructed track in the VELO detector. For Run3, assuming the simultaneous data-taking scenario, such a minimal selection would be dominated by beam-beam contamination and unsustainable. In order to implement a more discriminative selection, the z separation between the beam-beam and beam-gas interaction regions is exploited. Reconstructed VELO tracks are assumed to originate from SMOG2 when presenting an extrapolated position to the beam line with $z \in [-500, -300]$ mm, from a beam-beam collision elsewhere. As shown in the middle right and bottom plots of Fig. IV.3.8, this removes the largest fraction of the contamination from beam-beam events. Anyhow, when taking into account the differences in the track multiplicity for beam-gas and beam-beam collisions, even the residual small contamination from large-pseudorapidity tracks would dominate the rate. As an additional criterion, the different track geometry is thus considered by setting a threshold, by default to 12, on the minimum number of energy deposits in the VELO used to reconstruct the particle. Crossing a larger fraction of the VELO detector, this is largely efficient on beam-gas events, while reducing most of the beam-beam contamination.

IV.3.3.0.2 Multi-body selections

As listed in Tab. IV.3.5, other physics channels can be acquired with exclusive selections. Particles reconstructed combining the information from all the LHCb tracking detectors are extrapolated to the beam and selected only when the intersection point has $z \in [-500, -300]$ mm, hint for their origin in a beam-gas collision. All two- or three-track combinations are required to give a vertex reconstructed with good quality. For each of these, the invariant mass is calculated according to Eq. I.3.5 and an interval compatible with the decaying particle mass is defined. At the second trigger stage, when also PID information is available, a tighter interval is imposed. In addition to the Run2 requirements, the combination vertex is asked to be reconstructed in $z \in [-500, -300]$ mm, in order to remove the contamination from beam-beam events. In the same way, two- or three-body vertices presenting $z > -300$ mm are only accepted by beam-beam trigger selections. With this filter, all beam-beam HLT1 selections show compatible accepted event rates when running on the pp and the $pp + p$He simulated samples, this demonstrating that the interference from beam-gas-produced event is negligible.

Table IV.3.5: HLT1 selection algorithms implemented in the 2016 and 2017 data acquisition with the SMOG system.

Line name	**Run2 implementation**
	Minimum-bias lines
MicroBias	$\mathrm{N}(VELO) \leq 10$
MicroBiasLowMult	$2016:$ 1< N(VELO) < 10 & N(VELO, $\eta < 0$) = 0 $2017:$ 1< N(VELO) < 7.5 & N(VELO, $\eta < 0$) = 0 N(VELO, $\eta \in [3.5, 7]$) > 0
	Single hadron lines
SingleTrack	$2016: p_{\mathrm{T}} > 800\,\mathrm{MeV}/c$ & $p > 3\,\mathrm{GeV}/c$ & trk $\chi^2/\mathrm{ndf} < 4$ $2017: p_{\mathrm{T}} > 1\,\mathrm{GeV}/c$ & $GhostProb < 0.2$
	Muon lines
DiMuonHighMass	isMuon & $p > 3\,\mathrm{GeV}/c$ & $p_{\mathrm{T}} > 500\,\mathrm{MeV}/c$ & trk $\chi^2/\mathrm{ndf} < 3$ $\mathrm{M}(\mu^+, \mu^-) > 2500\,\mathrm{MeV}/c^2$ & $q_1 \cdot q_2 == -1$
	Two body lines
SMOGGeneric	$p > 3\,\mathrm{GeV}/c$ & $p_{\mathrm{T}} > 400\,\mathrm{MeV}/c$ & trk $\chi^2/\mathrm{ndf} < 4$ $DOCA < 1\,\mathrm{mm}$ & $max_{child}(p_{\mathrm{T}}) > 800\,\mathrm{MeV}/c$ & DV $\chi^2/\mathrm{ndf} < 25$
SMOGKpi	$p > 3\,\mathrm{GeV}/c$ & $p_{\mathrm{T}} > 400\,\mathrm{MeV}/c$ & trk $\chi^2/\mathrm{ndf} < 4$ $DOCA < 1\,\mathrm{mm}$ & $max_{child}(p_{\mathrm{T}}) > 400\,\mathrm{MeV}/c$ DV $\chi^2/\mathrm{ndf} < 25$ & $\lvert M - M(D^0)\rvert < 150\,\mathrm{MeV}/c^2$
SMOGppbar	$p > 3\,\mathrm{GeV}/c$ & $p_{\mathrm{T}} > 400\,\mathrm{MeV}/c$ & trk $\chi^2/\mathrm{ndf} < 4$ $DOCA < 1\,\mathrm{mm}$ & $max_{child}(p_{\mathrm{T}}) > 400\,\mathrm{MeV}/c$ DV $\chi^2/\mathrm{ndf} < 25$ & $\lvert M - M(\eta_c(1S))\rvert < 150\,\mathrm{MeV}/c^2$
	Three body lines
SMOGKKpi	$p > 3\,\mathrm{GeV}/c$ & $p_{\mathrm{T}} > 400\,\mathrm{MeV}/c$ & trk $\chi^2/\mathrm{ndf} < 4$ $DOCA < 1.0\,\mathrm{mm}$ & $max_{\mathrm{child}}(p_{\mathrm{T}}) > 400$ $\lvert M - M(D_s^+)\rvert < 150$& DV $\chi^2/\mathrm{ndf} < 25$
SMOGKpipi	$p > 3\,\mathrm{GeV}/c$ & $p_{\mathrm{T}} > 400\,\mathrm{MeV}/c$ & trk $\chi^2/\mathrm{ndf} < 4$ $DOCA < 1\,\mathrm{mm}$ & $max_{child}(p_{\mathrm{T}}) > 400\,\mathrm{MeV}/c$ $\lvert M - M(D^+)\rvert < 150\,\mathrm{MeV}/c^2$ & DV $\chi^2/\mathrm{ndf} < 25$
SMOGpKpi	$p > 3\,\mathrm{GeV}/c$ & $p_{\mathrm{T}} > 400\,\mathrm{MeV}/c$ & trk $\chi^2/\mathrm{ndf} < 4$ $DOCA < 1\,\mathrm{mm}$ & $max_{child}(p_{\mathrm{T}}) > 400\,\mathrm{MeV}/c$ $\lvert M - M(\Lambda_c^+)\rvert < 150\,\mathrm{MeV}/c^2$& DV $\chi^2/\mathrm{ndf} < 25$

Part V

Conclusions

The description of all observed phenomena within a unique theoretical framework is nowadays missing an explanation for a relevant fraction of the mass of the Universe, the Dark Matter. Experiments such as AMS-02 are searching a signature for its possible annihilation or decay processes through an excess of the antiproton abundance in cosmic rays with respect to its expected production from collisions between primary cosmic rays and the interstellar medium, mostly composed of hydrogen and helium. The interpretation of the data recently collected by the AMS-02 experiment for the antiproton flux relies on the knowledge of the antiproton production in collision systems involving protons and helium nuclei. For beam energies at the TeV scale, its experimental measurements are needed to improve the accuracy of the theoretical models.
Among the experiments operating at the CERN LHC accelerator, whose collisions match the needed energy scale, LHCb instruments a single-arm spectrometer in the forward region with excellent performance in track reconstruction and particle identification. It was originally designed to study particles containing b or c quarks, but its physics reach has been progressively extended, also embracing from 2015 the study of collisions between the LHC beams and noble gases injected into the accelerator beam-pipe. The first result of this novel physics programme, the first exploiting the energy and intensity of the LHC beams on fixed targets, was the determination of the prompt antiproton production in pHe collisions at a nucleon-nucleon centre-of-mass energy of $\sqrt{s_{\rm NN}} = 110$ GeV. For the first time, the cross-section on a helium target could be directly measured and the 100 GeV energy scale was reached in a fixed-target experiment.

In this thesis, my activities within the LHCb fixed-target physics programme are described. I extended the antiproton production measurement addressing the non-prompt component from $\overline{\Lambda} \to \overline{p}\pi^+$ (with a promptly- or a $\overline{\Xi}$-decay-produced $\overline{\Lambda}$ particle) and $\overline{\Sigma}^- \to \overline{p}\pi^0$ processes. These were expected to account for 25-40% of the prompt antiproton production, but were poorly constrained from previous experiments. Two approaches are followed and compared. In the first inclusive approach, the relative abundances of prompt and non-prompt (detached) antiprotons is determined with a template fit to the impact parameter distribution of selected antiproton candidates in the pHe data within the kinematic phase-space accessible to the LHCb detector

$$p > 12\,\text{GeV}/c,\ p_{\rm T} > 0.4\,\text{GeV}/c,\ \eta < 5.$$

The resulting ratio:

$$R = \frac{\sigma(p\text{He} \to (\overline{\Lambda} \to \overline{p}\pi^+)\,X) + \sigma(p\text{He} \to (\bar{\Sigma} \to \overline{p}\pi^0)\,X)}{\sigma(p\text{He} \to \overline{p}_{\rm prompt}\,X)}$$

varies between 0.3 and 0.52, slightly decreasing with the antiproton momentum and transverse momentum and is found to be consistent with the interpolation of previous measurements at lower or higher energies. The most common Monte-Carlo generators used in cosmic ray physics correctly predict the kinematic dependence, but significantly underestimate the R value at the 100 GeV energy scale.
The cross-section for the leading contribution among all antiprotons from anti-hyperons:

$$\sigma(p\text{He} \to (\overline{\Lambda}_{\rm prompt} \to \overline{p}\pi^+)X,\ \sqrt{s_{\rm NN}} = 110\,\text{GeV})$$

is independently measured. This second, exclusive, approach is highly complementary to the inclusive one, which exploited the antiproton identification capabilities offered

by the LHCb RICH detectors. In this case, the decays are selected relying on the reconstruction of the secondary vertex and on the decay kinematics, without any particle identification requirement. The obtained results as a function of the antiproton momentum and transverse momentum are divided by those for the prompt antiproton production to give:

$$R_{\overline{\Lambda}} = \frac{\sigma(p\mathrm{He} \to (\overline{\Lambda}_{\mathrm{prompt}} \to \overline{p}\pi^+)X)}{\sigma(p\mathrm{He} \to \overline{p}_{\mathrm{prompt}}X)},$$

where most of the systematics cancels out. The $R_{\overline{\Lambda}}/R$ ratio, representing the fraction of detached antiprotons originating in promptly-produced $\overline{\Lambda}$ decays, is found to be consistent with the predictions of the EPOS-LHC generator. Since these rely on isospin symmetry arguments and are thus expected to be more robust than for the detached-to-prompt antiproton ratio, the found consistency validates the results obtained with the two approaches.

During my *Ph.D.* programme, I also developed a data- and machine-learning-based approach to particle identification analyses of fixed-target samples, where the size of the collected calibration channels is insufficient to achieve the same precision as *pp* data. The $\overline{\Lambda} \to \overline{p}\pi^+$, $K^0_S \to \pi^-\pi^+$ and $\phi \to K^-K^+$ decays are reconstructed and selected in the most abundant fixed-target sample, obtained from *p*Ne collisions at $\sqrt{s_{\mathrm{NN}}} = 68$ GeV. The particle identification classifiers for antiprotons, pions and kaons are described for all calibration channels through a set of neural networks fed with the variables that are most relevant to the response of the particle identification detectors. This allows the particle identification classifiers distribution to be predicted on lower-size samples, also acquired with different energies and collision systems. A proof of the model is provided in several practical use-cases. The machine-learning-based model is found to describe the *p*He data sample used in the prompt antiproton production analysis better than the detailed LHCb simulation. It is used to scrutinise the systematic uncertainty from a possible difference in particle identification efficiency between prompt and detached antiprotons. In *pp* data, the model is found to be able to explain a long-standing tension between two calibration samples used to model the particle identification response to kaon mesons.

An upgrade of the fixed-target system has been implemented by the LHCb collaboration and is expected to operate from 2022. This mainly consists in the installation of a 20-cm-long confinement cell for the gas upstream the nominal LHCb collision point. A well-defined interaction region for the beam-gas collisions will allow the effective target areal density to be increased by up to two orders of magnitude with the same gas flow as in Run2. Studies to inject, other than helium, argon and neon already exploited in Run2, heavy noble gases like krypton or xenon and non-noble species such as oxygen, hydrogen and nitrogen are ongoing. The prospects for the related physics reach are unique at the LHC and cover several fields, such as probes for quark-gluon plasma studies, the nucleon structure and the extension of the programme of cosmic ray interest.
During my *Ph.D.* programme, I have been responsible for the software activities needed to include the upgraded fixed-target system in the challenging data acquisition and real-time reconstruction framework being deployed for the upgraded LHCb. I tested and optimised the performance of the event reconstruction sequence, used for both beam-beam and beam-gas collisions, achieving similar performance in both cases. By comparing samples with only beam-beam or beam-gas collisions with the possible

simultaneous data-taking scenario, I demonstrated there is no significant interference on the the beam-beam programme when injecting the gas and that the fixed-target event reconstruction performance is also not significantly affected by the simultaneous beam-beam collisions. The obtained results are paving the way to simultaneously operate LHCb as a fixed-target and a collider experiment. With the increase of the sizes of the collected samples and of the injectable gas species, the LHCb fixed-target programme will, according to the 2020 European Strategy for Particle Physics Update, "greatly extend the physics reach of the LHC complex at a limited cost", providing a unique laboratory for QCD studies.

Acknowledgements

What follows has to be in my native language, my apologies.

Durante questo mio dottorato, nelle brevi pause dalla frenetica danza delle scadenze, ho spesso riflettuto su quali aspetti rendessero validi tutti gli sforzi. Ho spesso pensato alla preziosità degli scambi che ho avuto con le persone più svariate, professionali e soprattutto personali, ai luoghi che ho potuto visitare e a tutto ciò per cui vale la pena continuare a impegnarsi, a lottare e a investire ogni energia.
In primo luogo, voglio ringraziare il gruppo di LHCb Firenze. Giacomo è stato il relatore di questa tesi e il tutor del mio lavoro, ma il modo in cui si è preso cura della mia formazione (e di me) si è spinto molto oltre, al ripetere di "benvenuto nella miniera di sale" e "il dottorando è il miglior amico dell'uomo e non paga mai il caffè". Sicuramente, di caffè ne ho scroccati parecchi. A Giovanni devo l'inizio di questa avventura, con l'offerta non rifiutabile di una strana tesi magistrale sullo studio di collisioni protoni-elio. Grazie per questo, e per il costante supporto e interesse al mio lavoro nei corridoi del CERN e di Firenze. Quantomeno, so di essermi in parte sdebitato passandoti sottobanco svariate ricette dei miei risotti. A Lucio devo poi tutto il resto, dalla pazienza agli aiuti e ai consigli. Sei stato e sei uno straordinario esempio, per le capacità, per la tenacia, per la dedizione e per l'attenzione. Non posso non ringraziare anche Antonio, per quanto continui a escludermi dalle paste per colazione. Compagno di numerose discussioni, per la maggior parte politiche, quasi tutte polemiche.
All'interno della collaborazione LHCb ho trovato tanti disponibili colleghi e più di un amico. I gruppi di lavoro con cui ho collaborato, *Ions and Fixed Target* e *Real Time Analysis*, mi hanno fatto crescere immensamente e ringrazio per le opportunità e il sostegno Vava, Sascha, Alex, Dorothea, Daniel, Agnieszka, Benjamin. All'interno della collaborazione, gli italiani sono più vivi e accoglienti che mai. Non posso fare tutti i nomi, ma ringrazio Pasquale Di Nezza per l'infinito entusiasmo e ottimismo, Luciano Pappalardo per la continua collaborazione, Barbara Sciascia e Matteo Palutan per i piacevoli momenti. Fin dall'inizio, mi sono sentito nel posto giusto.
Con il supporto dell'Istituto Nazionale di Fisica Nucleare e del mio direttore Oscar Adriani, verso il quale ho profonda riconoscenza e ammirazione, ho potuto trascorrere un anno al CERN. Molti dei progetti che avevo per tale anno sono purtroppo stati cancellati dal COVID, e non nascondo che un lockdown lontano da ogni affetto è stato una dura prova. Devo la mia sanità mentate alle ottime cene e vini sardi di Marianna e Dorothea e ai Cinefellow tutti, in primis a Tommaso e alle infinite passeggiate, discussioni, speranze e difficoltà, a Dario (e al suo pulled pork) e poi a Federico, Alberto, Luca, Chiara, Nico, Anita, Sabrina, Federica. Il simil-fellow più sfigato della storia, probabile, ma tutto sommato ci siamo decisamente divertiti.
Torino è stata (ed è ad oggi) un'altra casa, insieme alla radiosa cittadina di Saint-Genis Poully. Che arrivassi dopo ore di Flixbus e direttamente al Saint-Martin o che mi fermassi per la settimanale partita di beach, grazie davvero a Alessia, Chiara, Peppe, Fabio, Silvia e tutti. Un grazie speciale ai conquilini Irene e Stefano (rimani Ganella, comunque).
A Firenze ci sono poi le persone di una vita. Con Elena abbiamo di gran lunga superato i venti anni, tra i tuoi e i miei mille cambi di rotta. Grazie anche a Lorenzo, per come vi prendete cura l'uno delll'altra. Allo stesso modo devo ringraziare la mia Nana, che c'è sempre stata per qualsiasi cosa di cui avessi bisogno negli ultimi ormai tredici anni (!) e

ad Alessio. Grazie per ogni momento, per ogni cenino, per ogni birra. Che dire del mio gruppo di disturbati Smaug, che poi è una famiglia. Si vince e si cade insieme, per tutto, ormai da quasi dieci anni. Ringrazio infinitamente Giuls e Pulo, perchè so di potervi trovare sempre; Luca Monaco, perchè a forza di infinite (ma infinite davvero) discussioni sento più te di chiunque altro; Mailo per le infinite (davvero anche stavolta) polemiche e N***o (non ho saputo resistere ad abbondare con gli asterischi), Simo, Polpo, Binda, Virgi per tutte le parole, le risate e le birre, tante birre, tanti Negroni. Un grazie anche a tutti i fisici, Risto e Petruc per i pranzi, Viola, Carlona, Virginia e tutti.
Esprimere poi a parole un ringraziamento a quelli che sono i miei pilastri, a cui devo tutto, è ancora più difficile. Non posso che partire dalla mia famiglia, chiassosa, bizzarra, ma più unita che mai. Mia madre continua a inviarmi articoli di fisica chiedendomi se vi ho preso parte, nel tentativo di farsi una vaga idea di che cosa faccia nella vita. Forse è scontato ringraziarti per tutto quello che hai da sempre rappresentato, più probabilmente non lo è. Lo stesso vale per mio padre, sempre disponibile a starmi accanto. Da mia sorella è arrivato il principale vento di rivoluzione e la famiglia che hai creato con Marco e i miei adorati mimmi è sicuramente uno dei motivi che mi riporta ogni volta a casa. Vedere Matti e Lapo crescere, e così velocemente, è sicuramente tra le mie più grandi gioie. Ringrazio tutto il resto della mia enorme famiglia e dedico un pensiero in particolare a mio nonno, che, per quanto lo avrebbero reso infinitamente fiero, non è purtroppo riuscito a leggere queste righe. Descrivere a parole quello che è stato il contributo di Maddalena a questa tesi e alla mia vita in questi anni è quasi impossibile. Elencare i momenti e le sensazioni trascorse, in diverse parti di Italia e del mondo nei nostri folli viaggi, richiederebbe un testo a parte. La vita con te è un'avventura continua, mai facile e mai banale. Mi hai sostenuto in ogni occasione, distratto quando la tensione e il nervosismo erano troppi e pazientemente sopportato la mia continua vista vincolato a un computer. Non so mai quale sarà con te la meta successiva, purchè ci arriviamo insieme.

È opinione abbastanza comune che il dottorato sia in buona parte una sofferenza, per lo stress, per il basso stipendio e la mancanza di prospettive, per i continui impegni. Ho girato come una trottola, ho conosciuto persone da tutte al mondo, ho messo tutte le mie risorse lavorative e personali investendo su me stesso in questi tre anni. Vedere terminato questo documento, sintesi di tutti i miei sforzi, ripaga ogni serata o nottata dove ho acceso il computer per un'idea da implementare al piu' presto (per poi il più delle volte vederla fallire la mattina dopo), ripaga ogni testata sulla tastiera perchè non torna nulla, ripaga ogni inciampo, ripaga ogni volta in cui ho dovuto rimandare tutto il resto fino all'ultimo secondo. Che il mio percorso possa continuare ad avere case nei luoghi a cui già appartengo e nelle persone del presente e a trovarne di nuovi, ovunque essi siano.

Acronyms

ALICE A Large Ion Collider Experiment. iv, 30, 60

AMS-02 Alpha Magnetic Spectrometer. i, iv, 1, 12, 15–18, 20–22, 24, 265

ATLAS A Toroidal LHC ApparatuS. iv, 29–31, 215

BDT Boosted Decision Tree. 195

BGI Beam Gas Imaging. 50

CERN *Conseil européen pour la recherche nuclèaire*. i, iv, vii, 1, 3, 4, 9, 27–29, 265

CMB Cosmic Microwave Background. 8, 10

CMS Compact Muon Solenoid. iv, 29–31, 215

CPU Central Processing Unit. 241

CR Cosmic Ray. 7, 10–17, 19–22, 24, 53, 60, 124, 173, 207, 226

DIRA DIRection Angle. 107, 108, 115, 130, 132

DM Dark Matter. 7–9, 15, 16, 19, 22, 60

DOCA Distance Of Closest Approach. 107, 129, 130, 239, 263

DV Decay Vertex. 99, 100, 107, 108, 129, 130, 132, 182, 263

EB Event Building. 215, 217, 256

ECAL Electromagnetic Calorimeter. 40, 42, 44, 214

ECs Electron Clouds. 223

EFF Event Filter Farm. 215, 217

FD Flight Distance. 107, 108

FPGA Field Programmable Gate Array. 45, 215, 233

GEC Global Event Cut. 231

GEM Gas Electron Multipliers. 45

GFS Gas Feed System. 220

GMM Gaussian Mixture Model. 177, 179, 180, 188, 189, 199

GPU Graphic Processing Unit. 180, 189, 216, 217, 231, 232, 235, 240, 251, 255–257, 260

HCAL Hadronic Calorimeter. 40, 42, 44, 45

HEP High-Energy Physics. 10, 27, 216, 217

HLT High-Level Trigger. ix, 47, 50, 55, 65, 216–218, 230–232, 254, 256–259, 261, 262

HPDs Hybrid PhotoDetectors. 40, 214

IP Impact Parameter. 57, 62, 107, 108, 129, 130, 132, 209, 211, 246

ISM InterStellar Medium. 11, 13–15, 19, 20

ISR Intersecting Storage Ring. 27, 28

IT Inner Tracker. 38, 117, 209, 211

KS Kolmogorov-Smirnov. 186, 187, 195

L0 Level 0. 45, 47, 55

LEP Large Electron-Positron. 28

LHC Large Hadron Collider. i, iv–vii, 1, 7, 9, 22, 27–36, 40, 45–47, 49–53, 56, 98, 207–210, 215–217, 219, 220, 222–224, 227, 228, 251, 261, 265, 267

LHCb Large Hadron Collider beauty. i, iv–vii, ix, 4, 20, 22, 23, 25–27, 30–33, 35–40, 42–55, 57, 58, 63, 82–84, 98, 107, 121, 128, 153–155, 159, 163, 178–181, 183, 190, 195, 199, 207–210, 214, 215, 217–220, 222, 224–228, 231, 232, 235, 239, 258, 262, 266

MIB Machine Induced Background. 223

MISID MISIDentified. 66, 68, 70, 76, 110, 121

MLP Multi Layer Perceptron. 180

MWPC MultiWire Proportional Chambers. 45, 215

NN Neural Network. 180, 188, 189

OT Outer Tracker. 38, 117, 209, 211

OV Origin Vertex. 60, 62, 64, 68, 84, 85, 89, 94, 95, 99, 107, 108, 111, 115, 117–120, 162, 181, 187, 231

PAMELA Payload for Antimatter Matter Exploration and Light-nuclei Astrophysics. 12, 15–18, 20, 21

PDF Parton Distribution Function. 6, 53, 224, 225

pdf probability density function. 178–180, 189

PID Particle IDentification. ix, 40–43, 47, 56, 57, 66–68, 92, 94, 100, 110, 111, 113–115, 117, 121, 160, 173, 177–179, 181–184, 187, 188, 190, 191, 195, 199, 201, 214–217, 262

POCA Position Of Closest Approach. 183, 187, 191, 193, 239–242, 244, 245

pot protons on target. 52, 53

PS PreShower. 40, 42, 209

pS Proto-Synchrotron. 27, 29

PV Primary Vertex. 55, 57, 62, 65, 66, 68, 85, 89, 94, 95, 98, 99, 107, 108, 113, 129, 130, 132, 136, 144, 155, 159, 184, 185, 187, 239–246, 248, 249

QCD Quantum ChromoDynamics. 2, 4–7, 9, 267

QED Quantum ElectroDynamics. 2, 4, 5, 31

QGP Quark-Gluon Plasma. 30, 224, 225

RGA Residual Gas Analiser. 223

RICH Ring Imaging CHerenkov detector. vii, 16, 17, 40–42, 45, 47, 55, 66, 111, 173, 178, 181–184, 186–188, 191, 195, 199, 202, 209, 214, 217, 231

SciFi Scintillating Fibre. 208, 209, 211, 214, 229–231, 248, 251, 255

SIMD Single Instruction Multiple Data. 232

SIMT Single Instruction Multiple Threads. 216, 232, 251

SiPM Silicon PhotoMultiplier. 211

SM Standard Model of Particle Physics. 2, 3, 7–9, 29, 208

SMOG System for Measuring Overlap with Gas. iv–vii, ix, 1, 49–55, 57, 181, 191, 207, 219–227, 231, 235, 240–243, 246, 248, 251, 254, 256, 258, 261, 262

SOA Structure Of Array. 232, 233, 240

SPD Scintillating Pad Detector. 40, 42, 45, 55, 65, 95, 100, 107, 115, 116, 130, 136, 137, 178, 186, 187, 209

SpS Super Proto-Synchrotron. iv, 23, 28, 29, 53, 224

SUSY SuperSymmetry. 9

SV Secondary Vertex. 239

TBLV *TrackBeamLineVertexFinder*. 240, 242, 243, 246

TOF Time-Of-Flight. 16, 17

TRD Transition Radiation Detector. 16, 17

TT Trigger Tracker. 33, 37–39, 117, 209, 211, 249

UPC Ultra-Peripheral Collisions. 225

UT Upstream Tracker. 208, 209, 211, 213, 229–231, 248, 249, 251, 254, 255

VDM Van Der Meer. 49, 50

VELO VErtex LOcator. vii, ix, 33–37, 39, 50–52, 55, 65, 89, 93–95, 107, 108, 115, 117, 128, 130, 157, 159, 160, 162, 208, 209, 211–213, 219–223, 229–233, 235, 237, 239–241, 243, 244, 246, 248, 249, 251, 255, 261, 262

WFS Wake Field Suppressor. 223

WIMP Weakly-Interacting Massive Particles. 9

WM Wrong Mass. 115, 131, 132

Bibliography

[1] PBC forum: C. Barschel *et. al.* , *LHC fixed-target experiments* , CERN, Geneva, 2019. doi: 10.23731/CYRM-2020-004.

[2] R. K. Ellis *et al.*, *Physics Briefing Book: Input for the European Strategy for Particle Physics Update 2020*, arXiv:1910.11775.

[3] European strategy group, *2020 Update of the European Strategy for Particle Physics*, https://cds.cern.ch/record/2721370, 2020.

[4] S. Chatrchyan *et al.*, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Physics Letters B **716** (2012) 30.

[5] G. Aad *et al.*, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Physics Letters B **716** (2012) 1.

[6] Particle Data Group, P. A. Zyla *et al.*, *Review of particle physics*, Prog. Theor. Exp. Phys. **2020** (2020) 083C01.

[7] J. D. Bjorken, *Asymptotic sum rules at infinite momentum*, Phys. Rev. **179** (1969) 1547.

[8] G. Altarelli and G. Parisi, *Asymptotic freedom in parton language*, Nuclear Physics B **126** (1977) 298.

[9] V. N. Gribov, *A reggeon diagram technique*, http://jetp.ras.ru/cgi-bin/dn/e_026_02_0414.pdf, 1968.

[10] T. Pierog *et al.*, *EPOS LHC: test of collective hadronization with data measured at the CERN Large Hadron Collider*, Phys. Rev. **C92** (2015) 034906, arXiv:1306.0121.

[11] S. Ostapchenko, *QGSJET-II: towards reliable description of very high energy hadronic interactions*, Nuclear Physics B - Proceedings Supplements **151** (2006) 143.

[12] J. Albrecht *et al.*, *The muon puzzle in cosmic-ray induced air showers and its connection to the large hadron collider*, arXiv:2105.06148.

[13] F. Zwicky, *The redshift of extragalactic nebulae*, arXiv:1711.01693.

[14] F. Zwicky, *On the masses of nebulae and of clusters of nebulae*, Astrophysical journal **86** (1937).

[15] V. Rubin, *Rotation of the andromeda nebula from a spectroscopic survey of emission regions*, Astrophysical journal **15** (1979).

[16] M. Markevitch *et al.*, *Direct constraints on the dark matter self-interaction cross-section from the merging galaxy cluster 1e 0657-56*, The Astrophysical Journal **606** (2004) 819.

[17] N. Sugiyama, *Cosmic background anisotropies in cold dark matter cosmology*, The Astrophysical Journal Supplement Series **100** (1995) 281.

[18] WMAP collaboration, E. Komatsu *et al.*, *Seven-year WMAP observations: colsmological interpretation*, The Astrophysical Journal Supplement Series **192** (2011) 18.

[19] P. A. R. Ade *et al.*, *Planck 2015 results*, Astronomy & Astrophysics **594** (2016) A13.

[20] N. Aghanim *et al.*, *Planck 2018 results*, Astronomy & Astrophysics **641** (2020) A6.

[21] T. D. Brandt, *Constraints on macho dark matter from compact stellar systems in ultra-faint dwarf galaxies*, The Astrophysical Journal **824** (2016) L31.

[22] S. Hannestad, A. Mirizzi, G. G. Raffelt, and Y. Y. Y. Wong, *Neutrino and axion hot dark matter bounds after WMAP-7*, Journal of Cosmology and Astroparticle Physics **2010** (2010) 001.

[23] R. Bernabei *et al.*, *The DAMA/LIBRA apparatus*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **592** (2008) 297.

[24] R. Bernabei *et al.*, *First results from DAMA/LIBRA and the combined results with DAMA/NaI*, The European Physical Journal C **56** (2008) 333.

[25] R. Bernabei *et al.*, *DAMA/LIBRA-phase2 results and implications on several dark matter scenarios*, International Journal of Modern Physics A **35** (2020) 2044023.

[26] A. Berlin and F. Kling, *Inelastic dark matter at the LHC lifetime frontier: ATLAS, CMS, LHCb, CODEX-b, FASER, and MATHUSLA*, Physical Review D **99** (2019) .

[27] Particle Data Group, C. Patrignani *et al.*, *Review of particle physics*, Chin. Phys. **C40** (2016) 100001.

[28] J. A. Aguilar, *Particle astrophysics lecture 3*, `https://w3.iihe.ac.be/~aguilar/PHYS-467/PA3.html`.

[29] F. Nozzoli and C. Cernetti, *A data-driven approach for the measurement of ^{10}Be / ^{9}Be flux ratio in Cosmic Rays with magnetic spectrometers*, PoS **ICRC2021** (2021) 138.

[30] P. Picozza *et al.*, *PAMELA - a payload for antimatter matter exploration and light-nuclei astrophysics*, Astroparticle Physics **27** (2007) 296.

[31] A. Kounine, *The Alpha Magnetic spectrometer on the International Space Station*, International Journal of Modern Physics E **21** (2012) 1230005, arXiv:https://doi.org/10.1142/S0218301312300056.

[32] S. Torii, *The CALorimetric Electron Telescope (CALET): a High-Energy Astroparticle Physics Observatory on the ISS*, PoS **ICRC2015** (2016) 581.

[33] Auger collaboration, *The Pierre Auger Cosmic Rays observatory*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **798** (2015) 172.

[34] R. A. Ong *et al.*, *The gaps experiment to search for dark matter using low-energy antimatter*, arXiv:1710.00452.

[35] V. L. Ginzburg and S. I. Syrovatskii, *The Origin of Cosmic Rays*, https://www.elsevier.com/books/the-origin-of-cosmic-rays/ginzburg/978-0-08-013526-7, 1964.

[36] E. Fermi, *On the origin of the cosmic radiation*, 1949. doi: 10.1103/PhysRev.75.1169.

[37] M. Kachelriess, *Lecture notes on high energy cosmic rays*, arXiv:0801.4376.

[38] A. W. Strong and I. V. Moskalenko, *Propagation of cosmic-ray nucleons in the galaxy*, The Astrophysical Journal **509** (1998) 212.

[39] AMS collaboration, M. Aguilar *et al.*, *Precision Measurement of the boron-to-carbon flux ratio in Cosmic Rays from 1.9 GV to 2.6 TV with the Alpha Magnetic Spectrometer on the International Space Station*, Phys. Rev. Lett. **117** (2016) 231102.

[40] N. Przybilla, M.-F. Nieva, and K. Butler, *A cosmic abundance standard: Chemical homogeneity of the solar neighborhood and the ISM dust-phase composition*, The Astrophysical Journal **688** (2008) L103.

[41] R. Munini *et al.*, *Short-term variation in the galactic cosmic ray intensity measured with the PAMELA experiment*, doi: 10.22323/1.301.0091.

[42] M. Vecchi *et al.*, *The electromagnetic calorimeter of the AMS-02 experiment*, arXiv:1210.0316.

[43] PAMELA collaboration, O. Adriani *et al.*, *Ten years of PAMELA in space*, Riv. Nuovo Cim. **40** (2017) 473, arXiv:1801.10310.

[44] O. Adriani *et al.*, *An anomalous positron abundance in cosmic rays with energies 1.5-100 GeV*, Nature **458** (2009) 607.

[45] AMS Collaboration, L. Accardo *et al.*, *High statistics measurement of the positron fraction in primary cosmic rays of 0.5–500 GeV with the alpha magnetic spectrometer on the international space station*, Phys. Rev. Lett. **113** (2014) 121101.

[46] Particle Data Group, M. Tanabashi *et al.*, *Review of particle physics*, Phys. Rev. **D98** (2018) 030001, and 2019 update.

[47] AMS Collaboration, M. Aguilar *et al.*, *First result from the alpha magnetic spectrometer on the international space station: Precision measurement of the positron fraction in primary cosmic rays of 0.5–350 GeV*, Phys. Rev. Lett. **110** (2013) 141102.

[48] A. Ibarra, D. Tran, and C. Weniger, *Indirect searches for decaying dark matter*, International Journal of Modern Physics A **28** (2013) 1330040.

[49] P.-F. Yin, Z.-H. Yu, Q. Yuan, and X.-J. Bi, *Pulsar interpretation for the AMS-02 result*, Phys. Rev. D **88** (2013) 023001, arXiv:1304.4128.

[50] D. Gaggero *et al.*, *Three-dimensional model of cosmic-ray lepton propagation reproduces data from the alpha magnetic spectrometer on the international space station*, Phys. Rev. Lett. **111** (2013) 021102.

[51] A. S. Beach *et al.*, *Measurement of the Cosmic-Ray antiproton-to-proton abundance ratio between 4 and 50 gev*, Phys. Rev. Lett. **87** (2001) 271101.

[52] PAMELA Collaboration, O. Adriani *et al.*, *New measurement of the antiproton-to-proton flux ratio up to 100 GeV in the cosmic radiation*, Phys. Rev. Lett. **102** (2009) 051101.

[53] AMS Collaboration, M. Aguilar *et al.*, *Antiproton flux, antiproton-to-proton flux ratio, and properties of elementary particle fluxes in primary cosmic rays measured with the alpha magnetic spectrometer on the international space station*, Phys. Rev. Lett. **117** (2016) 091103.

[54] M. Aguilar *et al.*, *The Alpha Magnetic Spectrometer (AMS) on the international space station: Part II - Results from the first seven years*, Physics Reports **894** (2021) 1.

[55] M. Korsmeier, F. Donato, and M. Di Mauro, *Production cross-sections of cosmic antiprotons in the light of new data from the NA61 and LHCb experiments*, Physical Review D **97** (2018) .

[56] F. Donato, M. Korsmeier, and M. Di Mauro, *Prescriptions on antiproton cross-section data for precise theoretical antiproton flux predictions*, Physical Review D **96** (2017) .

[57] F. Donato *et al.*, *Antiprotons from spallations of Cosmic Rays on Interstellar Matter*, The Astrophysical Journal **563** (2001) 172.

[58] R. P. Duperray, C.-Y. Huang, K. V. Protasov, and M. Buénerd, *Parametrization of the antiproton inclusive production cross-section on nuclei*, Physical Review D **68** (2003) .

[59] G. Giesen *et al.*, *AMS-02 antiprotons, at last! secondary astrophysical component and immediate implications for Dark Matter*, Journal of Cosmology and Astroparticle Physics **2015** (2015) 023.

[60] A. Aduszkiewicz *et al.*, *Measurements of $\pi^{\pm}$, $k^{\pm}$, p and $\overline{p}$ spectra in proton-proton interactions at 20, 31, 40, 80 and 158 GeV/c with the NA61/SHINE spectrometer at the CERN SPS*, The European Physical Journal C **77** (2017) .

[61] LHCb collaboration, R. Aaij *et al.*, *Measurement of antiproton production in pHe collisions at $\sqrt{s_{NN}} = 110$ GeV*, Phys. Rev. Lett. **121** (2018) 222001, arXiv:1808.06127.

[62] T. Anticic *et al.*, *Inclusive production of protons, anti-protons and neutrons in pp collisions at 158 GeV/c beam momentum*, The European Physical Journal C **65** (2009) .

[63] L. C. Tan and L. K. Ng, *Parametrization of $\overline{p}$ invariant cross-section in $p-p$ collisions using a new scaling variable*, Phys. Rev. D **26** (1982) 1179.

[64] M. W. Winkler, *Cosmic ray antiprotons at high energies*, Journal of Cosmology and Astroparticle Physics **2017** (2017) 048.

[65] M. W. Winkler, *Cosmic rays in space*, https://agenda.infn.it/event/18734/contributions/98724/attachments/66188/80868/Winkler_lhcb.pdf, 2019. Presented at "Second LHCb Heavy Ion Workshop".

[66] M. Boudaud *et al.*, *AMS-02 antiprotons' consistency with a secondary astrophysical origin*, Physical Review Research **2** (2020) .

[67] G. Fidecaro, *The Discoveries of Rare Pion Decays at the CERN Synchrocyclotron*, Adv. Ser. Dir. High Energy Phys. **23** (2015) 397.

[68] C. courier, *The discovery of the weak neutral currents*, https://cerncourier.com/a/the-discovery-of-the-weak-neutral-currents/, 2004.

[69] L. Di Lella and C. Rubbia, *The Discovery of the W and Z Particles*, Adv. Ser. Direct. High Energy Phys. **23** (2015) 137.

[70] T. Berners-Lee, *Information management: A proposal*, https://www.w3.org/History/1989/proposal.html, 1989.

[71] L. Evans and P. Bryant, *LHC machine*, Journal of Instrumentation **3** (2008) S08001.

[72] CMS collaboration, S. Chatrchyan *et al.*, *The CMS Experiment at the CERN LHC*, JINST **3** (2008) S08004.

[73] ATLAS collaboration, G. Aad *et al.*, *The ATLAS experiment at the CERN LHC*, JINST **3** (2008) S08003.

[74] ALICE collaboration, K. Aamodt *et al.*, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002.

[75] LHCb collaboration, A. A. Alves, Jr. *et al.*, *The LHCb detector at the LHC*, JINST **3** (2008) S08005.

[76] C. Moller, *General properties of the characteristic matrix in the theory of elementary particles*, http://gymarkiv.sdu.dk/MFM/kdvs/mfm%2020-29/mfm-23-1.pdf, 1945.

[77] G. Papotti *et al.*, *Observations of beam-beam effects at the LHC*, arXiv:1409.5208.

[78] C. Barschel, *Precision luminosity measurement at LHCb with beam-gas imaging*, https://cds.cern.ch/record/1693671. Presented 05 Mar 2014.

[79] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv:0710.3820.

[80] LHCb collaboration, C. Elsässer, *$b\bar{b}$ production angle plots*, https://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb_ProductionAngles.html.

[81] LHCb collaboration, *LHCb magnet: Technical Design Report*, CERN-LHCC-2000-007, 2000.

[82] LHCb collaboration, *LHCb VELO (VErtex LOcator): Technical Design Report*, CERN-LHCC-2001-011, 2001.

[83] LHCb collaboration, *LHCb inner tracker: Technical Design Report*, CERN-LHCC-2002-029, 2002.

[84] LHCb collaboration, *LHCb outer tracker: Technical Design Report*, CERN-LHCC-2001-024, 2001.

[85] R. Arink *et al.*, *Performance of the LHCb Outer Tracker*, JINST **9** (2014) P01002, arXiv:1311.3893.

[86] R. Aaij *et al.*, *Performance of the LHCb Vertex Locator*, JINST **9** (2014) P09007, arXiv:1405.7808.

[87] LHCb collaboration, R. Aaij *et al.*, *LHCb detector performance*, Int. J. Mod. Phys. **A30** (2015) 1530022, arXiv:1412.6352.

[88] LHCb collaboration, *Tracking and alignment plots for conferences*, https://twiki.cern.ch/twiki/bin/view/LHCb/ConferencePlots.

[89] LHCb Silicon Tracker collaboration, *Material for publications*, https://lhcb.physik.uzh.ch/ST/public/material/index.php, 2016.

[90] LHCb Collaboration, L. M. Garcia, L. Henry, B. Kishor, and A. Oyanguren, *Tracking performance for long-lived particles at LHCb*, J. Phys. : Conf. Ser. **1525** (2019) 012095. 5 p, arXiv:1910.06171, Proceeding for Connecting the Dots and Workshop on Intelligent Trackers (CTD/WIT 2019).

[91] E. E. Bowen, B. Storaci, and M. Tresch, *VeloTT tracking for LHCb Run II*, LHCb-PUB-2015-024, 2016.

[92] M. De Cian, *Track Reconstruction Efficiency and Analysis of $B^0 \to K^{*0}\mu^+\mu^-$ at the LHCb Experiment*, CERN-THESIS-2013-145, 2013, Presented 14 Mar 2013.

[93] R. E. Kalman, *A new approach to linear filtering and prediction problems*, https://www.unitedthc.com/DSP/Kalman1960.pdf, 1960.

[94] E. Rodrigues, *Dealing with clones in the tracking*, LHCb-2006-057, 2006.

[95] M. De Cian, S. Farry, P. Seyfert, and S. Stahl, *Fast neural-net based fake track rejection in the LHCb reconstruction*, LHCb-PUB-2017-011, 2017.

[96] LHCb collaboration, *LHCb RICH: Technical Design Report*, CERN-LHCC-2000-037, 2000.

[97] LHCb collaboration, *LHCb calorimeters: Technical Design Report*, CERN-LHCC-2000-036, 2000.

[98] LHCb collaboration, *LHCb muon system: Technical Design Report*, CERN-LHCC-2001-010, 2001.

[99] H. Bichel, D. E. Groom, and S. R. Klein, *Passage of particles through matter*, Particle Data Group Review **32** (2013).

[100] LHCb RICH Collaboration, A. Papanestis and C. D'Ambrosio, *Performance of the LHCb RICH detectors during the LHC Run II*, LHCB-PUB-2017-012, 2017.

[101] M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, Eur. Phys. J. **C73** (2013) 2431, arXiv:1211.6759.

[102] LHCb collaboration, E. P. Olloqui, *LHCb preshower(PS) and scintillating pad detector (SPD): Commissioning, calibration, and monitoring*, Journal of Physics: Conference Series **160** (2009) 012046.

[103] LHCb MuonID group, F. Archilli *et al.*, *Performance of the Muon Identification at LHCb*, LHCB-DP-2013-001, 2013.

[104] LHCb collaboration, *LHCb online system, data acquisition and experiment control: Technical Design Report*, CERN-LHCC-2001-040, 2001.

[105] LHCb collaboration, *LHCb trigger system: Technical Design Report*, CERN-LHCC-2003-031, 2003.

[106] LHCb collaboration, *LHCb computing: Technical Design Report*, CERN-LHCC-2005-019, 2005.

[107] P. Mato, *GAUDI-Architecture design document*, LHCb-98-064, 1998.

[108] LHCb collaboration, *LHCb starterkit*, https://lhcb.github.io/starterkit-lessons/first-analysis-steps/dataflow.html.

[109] LHCb Collaboration, R. Aaij *et al.*, *RTA and DPA dataflow diagrams for Run 1, Run 2, and the upgraded LHCb detector* , LHCb-FIGURE-2020-016, 2020.

[110] R. Aaij *et al.*, *The LHCb trigger and its performance in 2011*, JINST **8** (2013) P04022, arXiv:1211.3055.

[111] R. Aaij *et al.*, *Design and performance of the LHCb trigger and full real-time reconstruction in run 2 of the lhc*, Journal of Instrumentation **14** (2019) 04013.

[112] *LHC computing grid web site*, https://wlcg.web.cern.ch/.

[113] G. Dujany and B. Storaci, *Real-time alignment and calibration of the LHCb Detector in Run II*, J. Phys. Conf. Ser. **664** (2015) 082010.

[114] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462** (2001) 152.

[115] Geant4 collaboration, S. Agostinelli *et al.*, *Geant4: A simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250.

[116] Geant4 collaboration, J. Allison *et al.*, *Geant4 developments and applications*, IEEE Trans. Nucl. Sci. **53** (2006) 270.

[117] M. Clemencic *et al.*, *The LHCb simulation application, Gauss: Design, evolution and experience*, J. Phys. Conf. Ser. **331** (2011) 032023.

[118] LHCb collaboration, R. Aaij *et al.*, *Precision luminosity measurements at LHCb*, Journal of Instrumentation **9** (2014) P12005.

[119] C. Rubbia, *Measurement of the luminosity of p-$\bar{p}$ collider with a (generalised) Van der Meer Method*, CERN-$p\bar{p}$-Note-38, 1977.

[120] M. Ferro-Luzzi, *Proposal for an absolute luminosity determination in colliding beam experiments using vertex detection of beam-gas interactions*, Nucl. Instrum. Methods Phys. Res. , A **553** (2005) 388.

[121] A. Bursche *et al.*, *Physics opportunities with the fixed-target programme of the LHCb experiment using an unpolarised gas target*, LHCb-PUB-2018-015, 2018.

[122] LHCb collaboration, R. Aaij *et al.*, *First measurement of charm production in fixed-target configuration at the LHC*, Phys. Rev. Lett. **122** (2019) 132002, arXiv:1810.07907.

[123] LHCb Collaboration, G. Graziani, *Results on heavy ion physics at LHCb*, J. Phys. : Conf. Ser. **1271** (2019) 012008. 10 p, arXiv:1904.04130, on behalf of the LHCb collaboration. Prepared for the fifth biennial Workshop on Discovery Physics at the LHC(Kruger2018), 3-7 December 2018, Hazyview (South Africa).

[124] ALICE collaboration, J. Adam *et al.*, *Enhanced production of multi-strange hadrons in high-multiplicity proton-proton collisions*, Nature Phys. **13** (2017) 535, arXiv:1606.07424.

[125] T. Pierog, C. Baus, and R. Ulrich, *CRMC library package*, web.ikp.kit.edu/rulrich/crmc.htm.

[126] M. Gyulassy and X.-N. Wang, *HIJING 1.0: A Monte Carlo programme for parton and particle production in high energy hadronic and nuclear collisions*, Computer Physics Communications **83** (1994) 307–331.

[127] T. Pierog and K. Werner, *EPOS model and ultra high energy cosmic rays*, Nuclear Physics B - Proceedings Supplements **196** (2009) 102.

[128] L. Gladilin, *Fragmentation fractions of c and b quarks into charmed hadrons at LEP*, The European Physical Journal C **75** (2015) .

[129] W. Verkerke, *RooFit*, EPJ Web of Conferences **4** (2010) .

[130] M. Pivk and F. R. Le Diberder, *A statistical tool to unfold data distributions*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **555** (2005) 356 .

[131] T. W. *et. al.* , *The disintegration of V0 particles*, Phys. Rev. Lett. **90** (1953) 329.

[132] LHCb collaboration, R. Aaij *et al.*, *Measurement of V^0 production ratios in pp collisions at $\sqrt{s}$ =0.9 and 7 TeV*, JHEP **08** (2011) 034, arXiv:1107.0882.

[133] LHCb collaboration, *Measurement of the track reconstruction efficiency at LHCb*, Journal of Instrumentation **10** (2015) 02007.

[134] LHCb collaboration, R. Aaij *et al.*, *Measurement of the track reconstruction efficiency at LHCb*, JINST **10** (2015) P02007, arXiv:1408.1251.

[135] G. Graziani *et al.*, *A Neural-Network-defined Gaussian Mixture Model for particle identification applied to the LHCb fixed-target programme*, arXiv:2110.10259.

[136] Y. Xie, *sFit: a method for background subtraction in maximum likelihood fit*, arXiv:0905.0724.

[137] C. Langenbruch, *Parameter uncertainties in weighted unbinned maximum likelihood fits*, arXiv:1911.01303.

[138] M. Borisyak and N. Kazeev, *Machine learning on data with splot background subtraction*, Journal of Instrumentation **14** (2019) P08020–P08020.

[139] F. Pedregosa *et al.*, *Scikit-learn: machine-learning in Python*, Journal of machine-learning Research **12** (2011) 2825.

[140] F. Chollet *et al.*, *Keras*, https://keras.io, 2015.

[141] M. Abadi *et al.*, *TensorFlow: Large-scale machine-learning on heterogeneous systems*, https://www.tensorflow.org/, 2015.

[142] Hecht-Nielsen, *Theory of the backpropagation neural network*, doi: 10.1109/IJCNN.1989.118638.

[143] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, *Automatic differentiation in machine-learning: a survey*, arXiv:1502.05767.

[144] A. Rogozhnikov, *hep_ml*, https://github.com/arogozhnikov/hep_ml, 2020.

[145] J. Brewer, A. Mazeliauskas, and W. van der Schee, *Opportunities of OO and pO collisions at the LHC*, arXiv:2103.01939.

[146] R. Aaij *et al.*, *A comparison of CPU and GPU implementations for the LHCb experiment Run3 trigger*, arXiv:2105.04031.

[147] LHCb collaboration, R. Aaij *et al.*, *Test of lepton flavor universality by the measurement of the $B^0 \to D^{*-}\tau^+\nu_\tau$ branching fraction using three-prong τ decays*, Phys. Rev. **D97** (2018) 072013, arXiv:1711.02505.

[148] LHCb collaboration, R. Aaij *et al.*, *Test of lepton universality in beauty-quark decays*, arXiv:2103.11769, Submitted to Nature Physics.

[149] LHCb collaboration, R. Aaij *et al.*, *Differential branching fraction of the rare decay $B_s^0 \to \phi\mu^+\mu^-$ and search for $B_s^0 \to f_2'(1525)\mu^+\mu^-$*, Phys. Rev. Lett. **127** (2021) 151801. 11 p, arXiv:2105.14007.

[150] LHCb Collaboration, T. Nakada, O. Ullaland, and W. Witzelling, *Expression of Interest for an LHCb Upgrade*, CERN-LHCC-2008-007, LHCC-G-139, 2008.

[151] *Letter of Intent for the LHCb Upgrade*, CERN-LHCC-2011-001, LHCC-I-018, 2011.

[152] LHCb collaboration, *LHCb VELO Upgrade Technical Design Report*, CERN-LHCC-2013-021, 2013.

[153] LHCb collaboration, *LHCb Tracker Upgrade Technical Design Report*, CERN-LHCC-2014-001, 2014.

[154] LHCb collaboration, *Physics case for an LHCb Upgrade II — Opportunities in flavour physics, and beyond, in the HL-LHC era*, arXiv:1808.08865.

[155] F. P. Albicocco *et al.*, *Long-term operation of the multi-wire-proportional-chambers of the LHCb muon system*, Journal of Instrumentation **14** (2019) P11031.

[156] LHCb collaboration, *LHCb PID Upgrade Technical Design Report*, CERN-LHCC-2013-022, 2013.

[157] J. Albrecht, V. V. Gligorov, and G. Raven, *Review Document: Full Software Trigger*, LHCb-PUB-2014-036, 2014, On behalf of the the HLT software group.

[158] LHCb collaboration, *LHCb Trigger and Online Upgrade Technical Design Report*, CERN-LHCC-2014-016, 2014.

[159] LHCb collaboration, *LHCb Upgrade Software and Computing*, CERN-LHCC-2018-007, 2018.

[160] LHCb collaboration, *LHCb Upgrade GPU High Level Trigger Technical Design Report*, CERN-LHCC-2020-006, 2020.

[161] R. Aaij *et al.*, *Allen: A High-Level Trigger on GPUs for LHCb*, Computing and Software for Big Science **4** (2020) .

[162] R. Aaij *et al.*, *Performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC*, JINST **14** (2019) P04013, arXiv:1812.10790.

[163] LHCb collaboration, *Computing Model of the Upgrade LHCb experiment*, CERN-LHCC-2018-014, 2018.

[164] LHCb collaboration, *LHCb SMOG Upgrade*, CERN-LHCC-2019-005, 2019.

[165] N. Caraban Gonzalez, *SMOG2 installation in the LHCb cavern*, https://cds.cern.ch/record/2727007, 2020.

[166] C. Boscolo Meneguolo *et al.*, *Calculation of the allowed aperture for a gas storage cell in IP8*, https://cds.cern.ch/record/2651289, 2018.

[167] F. Becattini, L. P. Csernai, and D. J. Wang, Λ *polarization in peripheral heavy ion collisions*, Phys. Rev. C **88** (2013) 034905.

[168] P. Li, E. Rodrigues, and S. Stahl, *Tracking Definitions and Conventions for Run 3 and Beyond*, LHCb-PUB-2021-005, 2021.

[169] S. Mariani, *LHCb - SMOG2 reconstruction performance: fixed-target physics at the LHCb experiment from Run3*, CERN-Poster-2020-1025, 2020.

[170] T. Bird *et al.*, *VP Simulation and Track Reconstruction*, LHCb-PUB-2013-018, 2013.

[171] A. Hennequin *et al.*, *A fast and efficient SIMD track reconstruction algorithm for the LHCb upgrade 1 VELO-PIX detector*, Journal of Instrumentation **15** (2020) P06018.

[172] D. H. Càmpora Pèrez, N. Neufeld, and A. Riscos Nuñez, *A fast local algorithm for track reconstruction on parallel architectures*, doi: 10.1109/IPDPSW.2019.00118.

[173] LHCb Collaboration, R. Aaij *et al.*, *Performance of VELO clustering and VELO pattern recognition on FPGA*, LHCb-FIGURE-2019-011, 2019.

[174] LHCb Collaboration, R. Aaij *et al.*, *Physics performances of the FPGA VELO clustering algorithm*, LHCb-FIGURE-2021-002, 2021.

[175] J. Tukey, *Approximate Weights*, The Annals of Mathematical Statistics **19** (1948) 91 .

[176] R. Aaij *et al.*, *Upgrade trigger: Biannual performance update*, LHCb-PUB-2017-005, 2017.

[177] M. De Cian *et al.*, *Status of HLT1 sequence and path towards 30 MHz*, LHCb-PUB-2018-003, 2018.

[178] R. Fruhwirth and W. Waltenberger, *Adaptive Multi-vertex fitting* , CERN, Geneva, 2004. doi: 10.5170/CERN-2005-002.280.

[179] E. Bowen and B. Storaci, *VeloUT tracking for the LHCb Upgrade*, LHCb-PUB-2013-023, 2014.

[180] P. Fernandez Declara *et al.*, *A parallel-computing algorithm for high-energy physics particle tracking and decoding using GPU architectures*, IEEE Access **7** (2019) 91612.

[181] P. Billoir, M. De Cian, P. A. Günther, and S. Stemmle, *A parametrised Kalman filter for fast track fitting at LHCb*, Computer Physics Communications **265** (2021) 108026.

[182] NVIDIA collaboration, *Profiler user guide*, https://docs.nvidia.com/cuda/profiler-users-guide/, 2019.

[183] E. Govorkova *et al.*, *A new scheduling algorithm for the LHCb upgrade trigger application*, J. Phys. Conf. Ser. **1525** (2020) 012052.