CBPF
Centro Brasileiro
de Pesquisas Físicas

Caio Victor Nogueira Damasceno

Charge asymmetry study of the $D_s^\pm \to \pi^\mp\pi^\pm\pi^\pm$ decay using LHCb Data

Dissertação de Mestrado

Advisor: Prof. Alberto Corrêa dos Reis

Rio de Janeiro
September 2023

"Charge asymmetry study of the DsPPP decay using LHCb Data"

CAIO VICTOR NOGUEIRA DAMASCENO

Dissertação de Mestrado em Física apresentada no Centro Brasileiro de Pesquisas Físicas do Ministério da Ciência Tecnologia e Inovação. Fazendo parte da banca examinadora os seguintes professores:

Alberto Correa dos Reis – Orientador/CBPF

Maria Clemencia Rosário Mora Herrera –UERJ

Irina Nasteva – UFRJ

Rio de Janeiro, 15 de setembro de 2023.

Acknowledgments

Completing this work was a huge journey, and I could not have done it without the support of so many important people.

First of all, I would like to thank God, who, acting like a weaver, has woven every step of my life up to this point in a brilliant and creative way. Who lent me intellect so I could develop knowledge and ideas, and use that to my enjoyment. Who took care of me throughout this process with His infinite mercy.

To my advisor Alberto dos Reis for reaching out to me and helping me understand and develop my understanding during this work. For guiding me through the entire process, ensuring I could be proud of my work.

To the Charm Rio Group, which was a safe haven every time I needed help, making themselves available to improve this work. To Carla Göbel, Erica Polycarpo and Sandra Amato my most sincere acknowledgements. Also to the fellow students in the group, for making this challenging process lighter, and especially for the super fun Munchkin games. To Beatriz, Juan and specially Felipe, for being so kind and helpful when I struggled with issues.

To CAPES and CBPF, for the aids granted, without which this work could not have been accomplished.

To Kathryn Coldham, who was an exceptional mentor, clearing my doubts and alleviating my concerns, always with kindness and helpfulness.

To Marta Cambeiro, my psychologist, who has been helping me understand myself and my struggles, making me able to face anxieties and challenges in the best possible way.

To my family, for their enduring love and care for me, specially my parents and my sister, Décio, Luciana and Isabella. My most sincere gratitude for your support during the difficult times.

And last but not least to my wife, Cecília. Thank you for believing in me even when I did not. Thank you for being my crutch when I could not move my legs. Thank you for holding my hand and never letting go.

Thanks to everyone who contributed, directly or indirectly, to the completion of this dissertation. To my friends, for making me laugh and have fun during moments of relaxation.

Abstract

Damasceno, Caio; dos Reis, Alberto C. (Advisor). **Charge asymmetry study of the $D_s^{\pm} \to \pi^{\mp}\pi^{\pm}\pi^{\pm}$ decay using LHCb Data**. Rio de Janeiro, 2023. 98p. Dissertação de Mestrado – Coordenação de Física de Altas Energias, Centro Brasileiro de Pesquisas Físicas.

The search for explanations to the asymmetry between matter and antimatter in the Universe has been a longstanding effort in Physics. The fact that antimatter is relatively scarce compared to matter leads to the question of why the universe appears to favor matter over antimatter. The source of this unbalance is the violation of the CP symmetry, which arises from the presence of a complex phase in the Cabibbo-Kobayashi-Maskawa (CKM) matrix of the Standard Model of particle physics. The presence of CP violation in the charm sector is very small, so it can be useful to provide evidence for new physics beyond the Standard Model.

In this work a fundamental step is performed in order to search for CP violation in charmed mesons. Using LHCb experiment data, this study examines charge asymmetry in the Cabibbo favoured $D_s^{\pm} \to \pi^{\mp}\pi^{\pm}\pi^{\pm}$ decay. The goal is to validate this channel as a reliable control for detecting CP violation in the singly Cabibbo suppressed decay $D^{\pm} \to \pi^{\mp}\pi^{\pm}\pi^{\pm}$.

The analysis is performed using the data collected in pp collisions at $\sqrt{s} = 13$ TeV by LHCb in the years 2016-2018 (Run 2), corresponding to $5.6\,\text{fb}^{-1}$ of integrated luminosity.

While the majority of the phase space displayed non-significant charge asymmetries, some areas require fine-tuning for a more accurate understanding. Despite minor adjustments needed, our findings provide a promising insight into the sources of charge asymmetry, emphasising the importance of understanding non-CP asymmetries in the $D^{\pm} \to \pi^{\mp}\pi^{\pm}\pi^{\pm}$ channel for observing CP asymmetries in this decay.

Keywords

Table of contents

List of figures

List of tables

, .

1
Introduction

At least since the proposition of the atom in ancient Greece, dating around the fifth century BC, humanity has sought to understand and explain how the matter that makes up our universe works.

But it was in 1928 that Paul Dirac[1] introduced a paradigm shift in this discussion. He proposed the existence of antiparticles as a solution to Equation 1-1, that would later be called the Dirac Equation, and which proposes to combine special relativity and quantum mechanics to describe the behaviour of elementary particles. Here, i is the imaginary unit, m is the mass of the particle, ψ is the wave function describing the quantum state of the particle, and $\slashed{\partial}$ is a shorthand notation for $\gamma^\mu \partial_\mu$, where γ^μ are the gamma matrices and ∂_μ are the four-gradient operators. The equation is written in natural units where $\hbar = c = 1$.

$$(i\slashed{\partial} - m)\psi = 0 \tag{1-1}$$

As early as 1932, just four years after Dirac's theoretical prediction, Carl Anderson[2] obtained the first experimental evidence of the existence of an antimatter particle, the positron, antiparticle of the electron. In Figure 1.1, recorded by Anderson, it is possible to observe the trail of a particle coming from the lower part of the image, which, due to the curvature in its trajectory caused by a magnetic field perpendicular to the plane of the photo, we know has a positive charge.
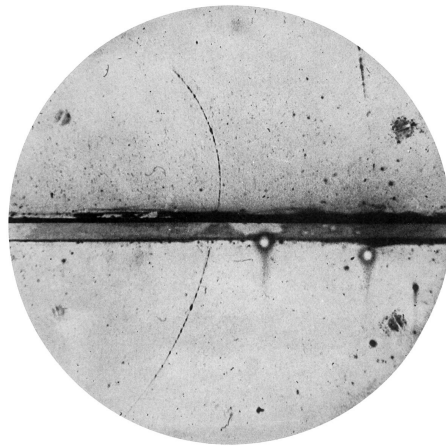
Figure 1.1: Photograph of a cloud chamber, obtained by Anderson in 1932 and which provided the first experimental evidence for the existence of the positron, in his experiment with cosmic rays.[2]

Each antiparticle has a mass equal to that of its corresponding particle, but with opposite electric charge and internal quantum numbers (such as isospin, strangeness, baryon number, and lepton number). Particle-antiparticle pairs can annihilate each other and give rise to new particles, as shown in the Feynman diagram in Figure 1.2.
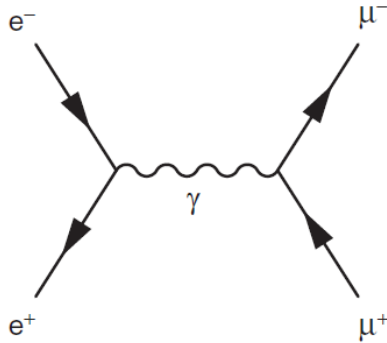
Figure 1.2: Feynman diagram of an electron-positron annihilation, giving rise to a photon that subsequently decays into a muon-antimuon pair.[3]

Some particles are their own antiparticles, such as the photon. Also, in the same way that combinations of matter particles can form different types of matter, the same is true for antiparticles. For example, an anti-hydrogen atom is composed of a positron and an antiproton.

Although antimatter exists, it is scarce in nature compared to matter. This raises the question of why we observe mostly matter in the universe. In fact, if matter and antimatter had the same abundance, they would annihilate each other completely. Thus, there must be some mechanism that caused an imbalance in the amount of matter and antimatter in the primordial universe, so that one of them could prevail.

The lack of antimatter cannot be explained simply by annihilation with matter, but also by some mechanism that produced matter in a slightly larger amount over antimatter, so that one could prevail over another. Also, the idea that there may be regions in the universe with an excess of antimatter is not supported by observations. The reason for the universe's apparent preference for matter is still an open question in physics.

In 1967, the Russian physicist Andrei Sakharov[4] proposed three requirements for the presence of asymmetry between matter and antimatter: (a) the existence of an interaction that violates the C and *CP* symmetries (which will be explained shortly); (b) an interaction that violates the baryon number conservation; (c) interactions out of thermal equilibrium.

Conditions (a) and (b) are important because without both, there is no way to produce matter and antimatter at unequal rates. The same applies to

condition (c): with thermal equilibrium, a reaction $i \to f$ will be as frequent as a reaction $f \to i$, so that the resulting asymmetry will be null. This work pays close attention to one of those conditions: CP violation.

In the Standard Model (SM) of particle physics, CP violation arises due to the presence of a complex phase in the quark mixing matrix, known as the Cabibbo-Kobayashi-Maskawa (CKM) matrix. This phase leads to interference between different decay amplitudes involving different weak and strong phases, which can manifest as CP violation effects in the decay products. However, the magnitude of these effects is typically very small, making them difficult to observe experimentally.

The search for CP violation in the charm sector is particularly challenging, due to its relatively small weak phase. Also, CP violation in charm occurs only in Cabibbo-suppressed decays, as will be further discussed in Chapter 2. But despite the challenges, there is strong motivation to search for CP violation in the charm sector, as it could provide evidence for new physics beyond the SM[5]. By searching for deviations from its predictions, we can gain new insights into the nature of matter.

In this work, we focus on identifying and quantifying the asymmetries in the $D_s^\pm \to \pi^\mp \pi^\pm \pi^\pm$ decay (from now on, charge conjugation of decays will be implied, unless otherwise stated), which is not predicted to exhibit CP violation as it has no Cabibbo-suppression. Our goal is to use this decay as a control channel for the search for CP violation in the Cabibbo-suppressed $D^+ \to \pi^- \pi^+ \pi^+$ decay, which is of particular interest not only for being able to have CP violation predicted by the SM but also for being able to exhibit evidence of new physics. We first introduce the theoretical background, then describe the LHCb detector and the data sample used in our analysis, as well as the event selection and reconstruction procedures. We present our methods for labeling the asymmetries in the $D_s^+ \to \pi^- \pi^+ \pi^+$ decay and discuss the implications of our work for the ongoing efforts to search for CP violation in charm meson decays and the potential for future studies to reveal new insights into the behavior of subatomic particles and the fundamental laws of nature.

2
Theoretical Fundamentals

2.1
The Standard Model

The SM is a quantum field theory based on a set of gauge symmetries, which determines the way elementary particles interact with each other. Even though it describes only the three fundamental forces that have a mediating field (leaving gravity out), the SM provides a very successful description of nature, showing good agreement with most of the current experimental data.

The aforementioned elementary particles can occur in three types: quarks, leptons, vector bosons and a scalar boson. Each of the first two consist of six particles, as shown in the two blocks on the left in Figure 2.1, in addition to their respective antiparticles. Also, in each of those blocks of six, particles are classified in pairs, called "generations" or "families". The lightest and most stable particles compose the first generation, and the heaviest and less-stable constitute the second and third generations. In Figure 2.1, these generations are given by the columns of the blocks at left; for example, the first generation is composed by the quarks and leptons of the first column, i.e., the quarks up (u) and down (d), the electron (e^-) and the electron neutrino (ν_e).

In the world of particle physics, quarks are never found alone; they always team up to form larger particles known as hadrons. When three quarks come together, they form a baryon, while a pair consisting of a quark and an antiquark makes a meson. This behavior is explained by a principle called 'confinement,' a key feature of Quantum Chromodynamics (QCD), which governs the strong nuclear force. Essentially, the energy needed to pull quarks apart increases so much that it becomes more efficient to spontaneously create new quark-antiquark pairs instead. This is why quarks are always part of larger particles and never exist in isolation.

Besides quarks and leptons, the SM also describes force-carrier particles, called gauge bosons, from whose exchange result three of the fundamental forces. The strong interaction is carried by the gluons (g), as the electromagnetic force is carried by the photon (γ), and the $W^\pm$ and Z^0 bosons carry the weak force.

Last but not least, another pivotal element of the SM is the Higgs boson, which provides the mechanism by which the elementary particles acquire mass. In contrast to fermions like quarks and leptons —particles with half-integer
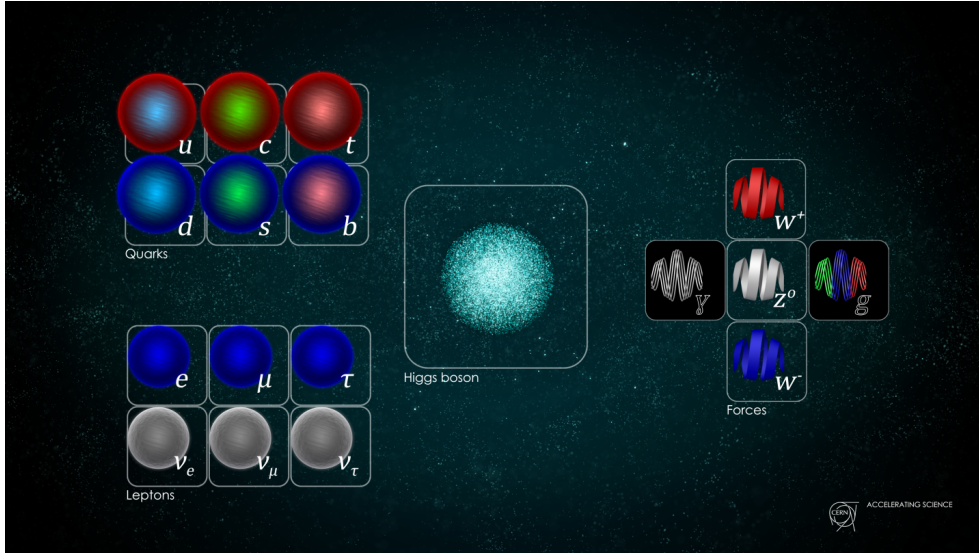
Figure 2.1: Particles of the Standard Model of particle physics[6]

spin that follow to the Pauli Exclusion Principle — bosons are particles with integer spin that can exist in the same state in unlimited numbers. Unlike other bosons, the Higgs boson is unique in having a spin of 0.

Another important feature of the SM resides in the fact that for each of the twelve fermions there exists an antiparticle, with the same mass but quantum numbers with opposite sign. The antiparticle of the electron for example is the positron, denoted by e^+. Similarly, the antiparticle of the up quark is the anti-up quark $\bar{u}$. The asymmetry between matter and antimatter is the heart of the discussion brought by this work, on the search for *CP* violation, as will be better explained on the next section.

2.2
P, C and *CP* symmetries

Until 1956, it was believed that the laws of physics were "two-handed": that changing the algebraic sign of the coordinate system of any physical process, it should be described by the same equations, i.e., that parity was always conserved. The parity transformation, implemented by the operator P, is:

$$\mathbf{r}_i \xrightarrow{P} \mathbf{r}'_i = -\mathbf{r}_i, \tag{2-1}$$

where $\mathbf{r}_i$ is the position vector, given by:

$$\mathbf{r}_i = x_i\hat{\imath} + y_i\hat{\jmath} + z_i\hat{k} \tag{2-2}$$

Lee and Yang[7] proposed an experimental test, performed by Wu[8], that aimed at verifying the conservation of the P symmetry in weak interactions,

which was not confirmed until then.

Wu's experiment consisted in aligning cobalt-60 nuclei spins in a specific direction, z, for example, through a magnetic field, according to Figure 2.2, and register the directions of the emitted electrons via beta decay:

$$^{60}\text{Co} \longrightarrow {}^{60}\text{Ni}^* + \text{e}^- + \overline{\nu}_e \tag{2-3}$$
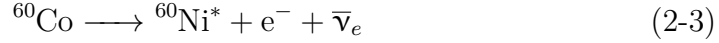
Through parity transformation, the momenta $\mathbf{p}$ of the particles have their signs changed, while the orbital angular momenta $\mathbf{r} \times \mathbf{p}$, as well as the spin remain the same. In that sense, since the schemes on Figure 2.2 are the P transformation of each other, for this symmetry to be conserved, both must be observed at the same rate.
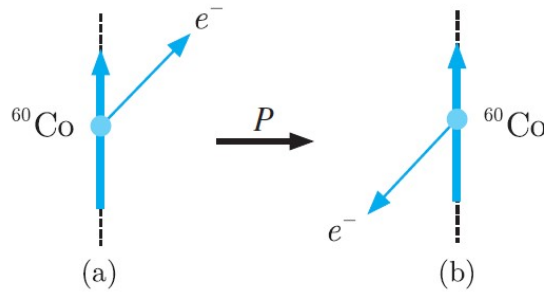
Figure 2.2: Wu Experiment illustration.[9]

However, they observed that most of the electrons were emitted in the opposite direction to the spin (Figure 2.2.b), leading to the conclusion that parity is *not* conserved in weak interactions.

The charge conjugation operation, represented by C, on the other hand, turns a particle into its antiparticle:

$$C \left| p \right\rangle = \left| \overline{p} \right\rangle \tag{2-4}$$

Regardless of being called *charge* conjugation, this transformation can be carried out also in electrically neutral particles, by changing sign of its "internal" quantum numbers (like baryon and lepton number, strangeness, charm, beauty ...) but not affecting momentum and spin[10].

Parity and charge conjugation are conserved symmetries of strong and electromagnetic interactions, and both are violated in weak interactions.

Before discussing examples, it is important to make some definitions. The helicity h of a particle consists of the normalised component of its spin along its momentum. But a property frequently confused with helicity is chirality, which is a quantum number that describes the transformation properties of a particle's wavefunction under parity inversion. Unlike helicity, which is related to the direction of spin in relation to momentum, chirality represents an

inherent feature of a particle. The wavefunction of any fermion, for example, can be described as a combination of chirality eigenstates. Although helicity and chirality are distinct properties, they become equivalent in the ultra-relativistic limit ($E \gg m$), helicity eigenstates are equivalent to chirality eigenstates.

It is important to point out that in SM only left-handed chirality particles and right-handed chirality antiparticles participate in the charged weak interactions.

So, by observing the π^+ decay in Equation 2-5, for example, as the emitted antimuon is always right-handed, a parity transformation should give us a left-handed antimuon, which can not be found in nature,

$$\pi^+ \rightarrow \mu^+ + \nu_\mu. \tag{2-5}$$

Then, applying C in Equation 2-5, we get:

$$\pi^- \rightarrow \mu^- + \overline{\nu}_\mu \tag{2-6}$$

As C does not change chirality, this gives us a right-handed muon, which does not exist either, as they come out always left-handed. Hence, charge conjugation is also violated.

Nonetheless, if we combine both operations in Equation 2-5, CP turns left-handed muons into right-handed antimuons, exactly as observed in nature. Thus, comes up a transformation that seemed to be lead to a fundamental symmetry of nature.

The Fitch-Cronin experiment[11], in 1964, though, gave the first evidence of violation of the CP symmetry, in decays of neutral kaons. The CP violation differs from the parity breaking, for example, because it is a subtle effect, and therefore requires sensitive measurement, while parity is maximally violated in weak interactions.

As an example, while *all* neutrinos are left-handed, violating P, the decay $B^0 \rightarrow K^+\pi^-$ is 13% more frequent than its CP transformed version[12], given by $\overline{B}^0 \rightarrow K^-\pi^+$. In the vast majority of B decays, CP is conserved, even though the quark-b sector has a more intense CP violation compared to others. Nowadays, it is understood that the actual conserved symmetry is given by CPT, which includes also the time reversal symmetry T, and is conserved in every Lorentz invariant theory.

The key point is to notice that CP violation, being one of the Sakharov's conditions, allows the production of matter and antimatter at different rates, making it an important subject when it comes to trying to understand the reasons why there is so much more matter than antimatter in the Universe. However, the CP violation provided by the SM is not sufficient to explain

baryon asymmetry[13]. It is necessary that some other CP breaking mechanism beyond the Standard Model exists, to justify such a significant unbalance. In this way, the study of CP violation also becomes a potential gateway to the observation of new physics.

Since the Fitch-Cronin experiment, CP violation was found again in neutral kaons at Fermilab[14] and CERN[15] in 1999. In 2001, it was observed also in the experiments Belle[16] and BaBar[17], in B mesons. Finally, it was also detected at LHCb, in B_S^0 decays[18], in 2013, and also in D^0 decays[19] in 2019.

It is important to realize how subtle the effect of CP violation is in the charm sector, so that the first observation only took place 55 years after Fitch and Cronin's discovery. And in this work we perform one of the steps in the search for CP violation in the $D^+ \to \pi^- \pi^+ \pi^+$ decay.

2.3
CKM Matrix

The key point that culminated in what we now know as the CKM matrix is the so-called Cabibbo hypothesis. Although the weak interaction presents a universal coupling constant for the leptonic case, for quarks the situation is somehow different. The coupling constants for the charged current weak interaction, that occur through the emission of a W boson, with vertices ud and us, for example, are not only different from the leptonic case, but also different between them. This feature, therefore, opposes the idea that there could be a universal weak coupling for the quark sector[3].

Thus, in 1963, Cabibbo presented his hypothesis to explain the absence of the $K^0 \to \mu^+ \mu^-$ decay[20]. At the time of his proposal, only the quarks up, down and strange were known, and his idea was that the weak coupling constant for the quarks would have the same value as for the leptonic case, but the weak eigenstates of quarks would differ from the mass eigenstates. In this way, the mass eigenstates, called d' and s', are related to the weak eigenstates through the following matrix:

$$\begin{pmatrix} d' \\ s' \end{pmatrix} = \begin{pmatrix} \cos\theta_c & \sin\theta_c \\ -\sin\theta_c & \cos\theta_c \end{pmatrix} \begin{pmatrix} d \\ s \end{pmatrix}, \tag{2-7}$$

where θ_c is called the Cabibbo angle.

This mechanism can explain, for example, the reason why the decay rate of the process $K^-(u\bar{s}) \to \mu^- \bar{\nu}_\mu$ is about twenty times smaller than for the process $\pi^-(u\bar{d}) \to \mu^- \bar{\nu}_\mu$. In light of the matrix in Equation 2-7, we can notice that the K^- decay is suppressed by a factor of $\tan^2\theta_c$. Now we now, based on existing measurements, that $\theta_c \approx 13°$[3].

Since the Cabibbo mechanism allows the couplings ud and us to happen, one may expect for example that the long-lived neutral kaon K_L^0, which is a weak eigenstate, into $K_L^0 \to \mu^+\mu^-$ can occur through the exchange of a virtual up quark, as shown in Figure 2.3
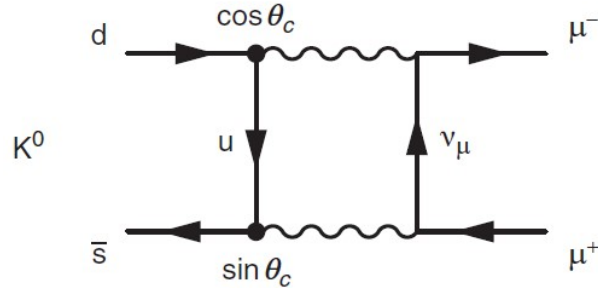
Figure 2.3: $K_L \to \mu^+\mu^-$ decay through exchange of a virtual u quark.[3]

However, the observed decay rate for such a process is much smaller than expected for the case with only the contribution of the diagram in Figure 2.3. Thus, in 1970, Glashow, Iliopoulos and Maiani[21] postulated the existence of a fourth quark that would couple to the weak eigenstate s'. In this way, this decay would also be able to happen through the exchange of a quark that would be called *charm*, but with new coupling constants, as shown in Figure 2.4. It is important to notice that those two contributions have opposite sign, and cancel it other; this cancelling is not perfect though, as the quarks have different masses. This is the so-called GIM (Glashow, Iliopoulos and Maiani) mechanism, and it was responsible for explaining why the $K_0 \to \mu^+\mu^-$ decay has a smaller branching ratio than was expected.
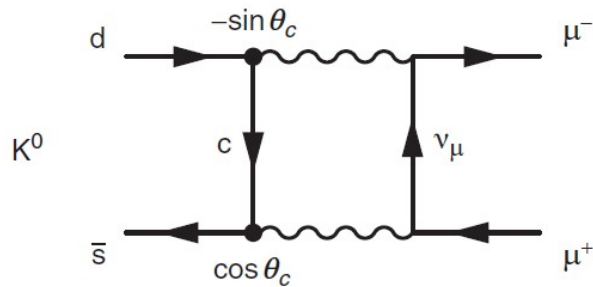
Figure 2.4: $K_L \to \mu^+\mu^-$ decay, but now through exchange of a virtual c quark.[3]

The generalisation of the Cabibbo mechanism for the three families of quarks of the SM gives rise to what we call the CKM matrix, as shown in Equation 2-8. This insight came before the discovering of the third generation of quarks, firstly as an explanation for *CP* violation through the presence of an irreducible complex phase, and got Kobayashi and Maskawa a Nobel Prize

in 2008. The CKM matrix is a 3×3 unitary matrix, and gives us the relation between the weak and the mass eigenstates of quarks, but now for the case with the three generations.

$$
\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{2-8}
$$

A 3×3 matrix can be constructed in its most general form in terms of 9 complex numbers, which gives it 18 real parameters. Through the unitarity condition, 9 constraints arise, remaining only 9 independent parameters.

Given its characteristics, if the CKM matrix had only real elements, it would correspond to the rotation matrix R, and could be written in terms of 3 rotation angles, θ_{12}, θ_{13} e θ_{23}:

$$
R = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \times \begin{pmatrix} c_{13} & 0 & s_{13} \\ 0 & 1 & 0 \\ -s_{13} & 0 & c_{13} \end{pmatrix} \times \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{2-9}
$$

where $c_{ij} = \cos \theta_{ij}$ and $s_{ij} = \sin \theta_{ij}$, for $i, j = 1, 2, 3$.

However, because it is a complex matrix, 6 more degrees of freedom are needed, that is, 6 more complex phases with the form $e^{i\delta}$. We can also rewrite those phases through a reparameterization in terms of a global phase δ, so that only one complex phase remains, leading us to a matrix of the form:

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

$$
= \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \times \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta'} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta'} & 0 & c_{13} \end{pmatrix} \times \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2-10}
$$

Thus, as discussed in section 2.4, the presence of a complex phase in the CKM matrix allows the existence of *CP* violation.

The experimental values for the elements of the CKM matrix are compatible with the unitarity condition, and are the following[22]:

$$
|V_{\mathrm{CKM}}| = \begin{pmatrix} 0.97435 \pm 0.00016 & 0.22500 \pm 0.00067 & 0.00369 \pm 0.00011 \\ 0.22486 \pm 0.00067 & 0.97349 \pm 0.00016 & 0.04182^{+0.00085}_{-0.00074} \\ 0.00857^{+0.00020}_{-0.00018} & 0.04110^{+0.00083}_{-0.00072} & 0.999118^{+0.000031}_{-0.000036} \end{pmatrix},
$$

$$
\tag{2-11}
$$

where is easy to see that out-of-diagonal elements are relatively small, meaning that the coupling between quarks of different generations is suppressed. So, we say that weak interactions between quarks from different generations are *Cabibbo-suppressed* (or *doubly Cabibbo-suppressed* if they are two families apart). From the moduli of the CKM matrix elements, the angles are given by $\theta_{12} = 13°$, $\theta_{13} = 2.3°$ e $\theta_{23} = 0.2°$[3].

It is important to notice, though, that the existence of a complex phase that can not be eliminated through reparameterization depends on the number of quark families in the CKM matrix. A $n \times n$ unitary matrix can be defined in terms of $2n$ complex phases, where $2n - 1$ of those phases can be redefined. Thus, the number of real parameters needed is given by $n^2 - (2n-1) = (n-1)^2$. As the matrix has also to be orthogonal, it contains $n(n-1)/2$ parameters corresponding to the independent rotation angles. Then, the minimum number of complex phases in an unitary matrix of n dimension is given by[23]:

$$N_{\text{phases}} = (n-1)^2 - \frac{1}{2}n(n-1) = \frac{1}{2}(n-1)(n-2) \qquad (2\text{-}12)$$

We can notice then, through Equation 2-12, that for two generations of quarks ($n = 2$), it is possible to redefine the CKM matrix in order to rule out any complex phase, as $N_{\text{phases}} = 0$. For the three quark families in the SM ($n = 3$), on the other hand, there will always be a complex phase in the CKM matrix, so that this is the minimum number of families necessary for the existence of *CP* violation through this mechanism. Clearly, for more generations of quarks, more complex phases would arise[23].

Since the elements of the CKM matrix have very different magnitudes (see Equation 2-11), it is possible to parameterise them through the expansion of $\lambda = \sin\theta_C$ (the sine of the Cabibbo angle). In the so-called Wolfenstein parameterization[24], the elements of the matrix are written in terms of the four real parameters λ, A, ρ and η, all of the order of unity ($\lambda \simeq 0.23, A \simeq 0.82, \rho \simeq 0.22, \eta \simeq 0.34$), such that the number of powers in each element indicates its relative size. The CKM matrix can be written until fourth order contributions, through the Wolfenstein parameterization, as:

$$V_{\text{CKM}} = \begin{pmatrix} 1 - \lambda^2/2 & \lambda & A\lambda^3\left(\rho - i\eta + i\eta\lambda^2/2\right) \\ -\lambda & 1 - \lambda^2/2 - i\eta A^2\lambda^4 & A\lambda^2\left(1 + i\eta\lambda^2\right) \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + O\left(\lambda^5\right)$$

(2-13)

This parameterization helps us observe the "hierarchical" structure of the CKM matrix mentioned before, in which interactions between quarks of different families can be singly or doubly suppressed. The matrix elements between generations one and two are of the order of λ, those between generations

two and three are of the order of λ^2 and those between generations one and three are of the order of λ^3.

We can elaborate more formally this need of a phase that can not be redefined. As only functions of V_{ij} that are invariant over a global phase redefinition can have physical meaning, physical quantities can be written in terms of the moduli of the matrix elements and of $V_{ij}V_{kl}V_{il}^*V_{kj}^*$ (with no implicit summation)[23]. Besides that, the unitarity of V_{CKM}, in the three-quark-generations case, defines that the non-reparameterizable imaginary part of all those four-products of elements of V_{CKM} must have the same modulus[25]:

$$J_{CP} \equiv \left| \text{Im}\left(V_{ij}V_{kl}V_{il}^*V_{kj}^* \right) \right| = s_{12}s_{13}s_{23}c_{12}c_{13}^2 c_{23}\sin\delta_{CP} =$$
$$= A^2\lambda^6\eta\left(1 - \lambda^2/2\right) + \mathcal{O}\left(\lambda^{10}\right) \approx 3 \times 10^{-5} \quad (2\text{-}14)$$
$$\text{Im}\left(V_{ij}V_{kl}V_{il}^*V_{kj}^* \right) = J_{CP}\sum_{m,n}\epsilon_{ikm}\epsilon_{ijn},$$

where ϵ_{ikm} is the antisymmetric tensor.

This way, any quantity that violates CP in the SM must be proportional to J_{CP}, which is known as the Jarlskog invariant[26]. This is because only one non-removable complex phase shows up in the CKM matrix for three families of quarks[23]. The maximum value allowed for J_{CP} would be $1/(6\sqrt{3}) \simeq 0.1$, but its experimental value is substantially smaller, of the order of 10^{-5}. This is evidence that CP violation is largely suppressed due to the hierarchy of the CKM matrix elements[25], in Equation 2-11.

Another interesting aspect in Equation 2-14 is the fact that at least four types of different quarks must be involved in the process (real or virtual) for CP violation to occur.

In the case where transitions may occur, the CP violating effects are expected to be stronger when any competing CP-conserving mechanism has a small amplitude. That is the reason why D meson decays ($c \to u$), for example, are good candidates[23].

Finally, one way of describing weak phases is through unitarity triangles, that represent the unitarity of the CKM matrix, which gives us relations of the form:

$$\sum_{i=1}^{3} |V_{ij}|^2 = 1; \quad j = 1, \dots, 3 \quad (2\text{-}15)$$

$$\sum_{i=1}^{3} V_{ji}V_{ki}^* = 0 = \sum_{i=1}^{3} V_{ij}V_{ik}^*; \quad j, k = 1, \dots, 3, \quad j \neq k. \quad (2\text{-}16)$$

The first equation expresses the so-called *weak universality*, which means that the sum of the couplings of an up-type quark with all down-type quarks
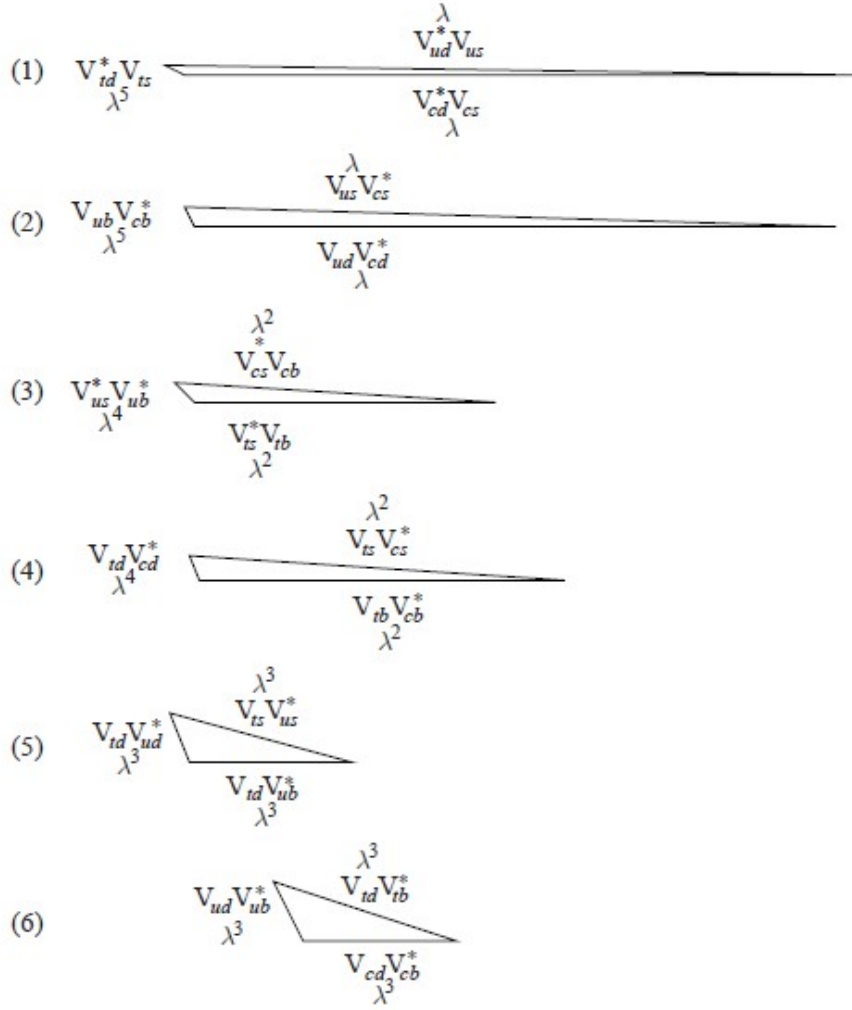
Figure 2.5: The six triangles originated by the unitarity relations of the CKM matrix elements, in Equations 2-15 and 2-16.[5]

is the same whatever the up quark is; the second equation expresses the unitarity of V_{CKM}. Since the CKM matrix elements are complex, the relations in Equations 2-15 and 2-16 tell us that they form triangles in the complex plane.

A first aspect about those triangles is that a global phase change can, at most, rotate them in the complex plane, not changing their shape. Besides that, they all have the same area, given by $|J_{CP}|/2$. It becomes clear that, due to the fact that the triangles have a non-zero area (or they would be just a line), they necessarily have an imaginary part, and because of that, manifest the existence of CP violation. Taking the inverse path, we notice that without CP violation, $J_{CP} = 0$, and the area of the triangles becomes null. Also, the fact that all triangles have the same area is a consequence of the existence of only one mandatory complex phase. The different shapes of those triangles is shown in Figure 2.5.

2.4
Complex Phases and *CP* violation

For a physical quantity to be conserved, it is necessary that the Hamiltonian of the process in question is invariant under such symmetry transformation. Hence, for CP to be conserved, we must have:

$$H_{CP} \equiv (CP)^\dagger H(CP) = H \quad \text{or} \quad [H, CP] = 0, \tag{2-17}$$

where $H^\dagger = (H^*)^T$.

Suppose, for example, a Lagrangian density with a Yukawa coupling, that describes the interaction between scalar and fermion fields in a quantum field theory:

$$\mathcal{L}^{(Y)} = g\bar{\psi}_L \phi \psi_R + g^* \bar{\psi}_R \phi^\dagger \psi_L \tag{2-18}$$

The CP transformation of such term gives us:

$$\mathcal{L}^{(Y)} \xrightarrow{CP} \mathcal{L}^{(Y)}_{CP} = g\bar{\psi}_R \phi^\dagger \psi_L + g^* \bar{\psi}_L \phi \psi_R \tag{2-19}$$

So in order to CP to be conserved, it is necessary that $\mathcal{L}^{(Y)} = \mathcal{L}^{(Y)}_{CP}$, that determines the strength of the interaction. That is: CP violation is described by complex terms appearing in the Lagrangian density, so that $g \neq g^*$. For the strong interaction, described by a Yukawa coupling, $g = \sqrt{4\pi\alpha_S}$, where $\alpha_S \approx 0.118$. As g is a real parameter in this case, CP is conserved in strong interactions.

However, not any complex phase has a physical meaning, given that some of them can be ruled out through a global phase definition, as shown Section 2.3. Thus, CP violation only happens when it is not possible to eliminate all the complex parameters through a phase redefinition. As, to date, CP violation has only been observed in weak interactions, those complex phases are called *weak phases*[23].

One difficulty that arises from this is the fact that experimentally observed quantities are expressed as squared matrix elements, in a way that obscures the presence of these complex phases. A quantity that can indeed be observed, on the other hand, is the phase difference, i.e., interferences between complex phases[23], in a way that they either reinforce or cancel each other out.

Supposing an interference of two amplitudes in a given process:

$$A(i \to f) = |A_1| \, e^{i\phi_1} + |A_2| \, e^{i\phi_2}, \tag{2-20}$$

with i being the initial state and f the final state.

Thus, a term related to the phase difference in the squared modulus of the amplitude appears:

$$|A(i \to f)|^2 = |A_1|^2 + |A_2|^2 + 2\,|A_1|\,|A_2|\cos\left(\phi_1 - \phi_2\right) \qquad \text{(2-21)}$$

Nonetheless, to get the value of the third term through the transition rate, it is necessary to know the values of the first two, which is not always possible to have. Therefore, another approach is necessary.

Additional phases may come up due to final state interactions, such as strong interactions between the decay products. Those, in opposition to the weak phases, are called *strong phases*. They are fundamental for *CP* violation detection, since in a *CP* transformed process, they do not change sign, while weak phases do.

Thus, performing the *CP* transformation of a process with a strong and a weak phase:

$$A(i \to f) = |A_1|\,e^{i\phi_i}e^{i\delta_1} + |A_2|\,e^{i\phi_2}e^{i\delta_2}$$
$$A(\bar{i} \to \bar{f}) = |A_1|\,e^{-i\phi_i}e^{i\delta_1} + |A_2|\,e^{-i\phi_2}e^{i\delta_2} \qquad \text{(2-22)}$$

So the difference between the transition rates is:

$$|A(\bar{i} \to \bar{f})|^2 - |A(i \to f)|^2 = 2\,|A_1|\,|A_2|\sin\left(\phi_1 - \phi_2\right)\sin\left(\delta_1 - \delta_2\right), \qquad \text{(2-23)}$$

and goes to 0 for the case when $\phi_1 = \phi_2$ (assuming that $\delta_1 \neq \delta_2$). That is the reason why it is possible to observe the presence of phases that violate $CP^{[23]}$.

With that in mind, we are now able to proceed to an analysis of the CKM matrix and its terms, knowing the importance of those weak phases for the presence of *CP* violation.

2.5
Dalitz plot

This work studies three-body decays, which consist in processes that can be written as:

$$X \to abc. \qquad \text{(2-24)}$$

In these processes, different energy-sharing relationships between the daughter particles can be defined by their three-momenta relative orientation and magnitude. As there are three final-state particles, their four-momenta give us twelve parameters, from which only nine of them are independent, due to momentum conservation. Also, the relation between energy and momentum reduces this number to four. Finally, since in the rest frame of the parent particle the three-momenta of the daughter particles add up to the null vector, they must be co-planar. It takes then three angles to determine the orientation of this plane, reducing the number of free parameters to two, leaving only two degrees of freedom to a three-body decay.
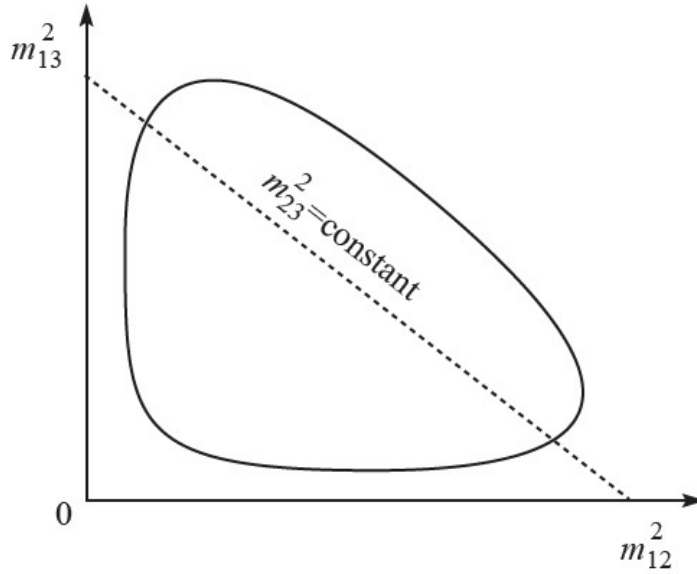
Figure 2.6: Schematic of a generic Dalitz plot, with the invariant masses m_{12}^2 and m_{13}^2 as the plot axes.[28]

Knowing that, an important tool used in the analysis of three-body decays is the so-called Dalitz plot (DP), named after Richard Dalitz[27]. It allows us to represent those decays in a two-dimensional plot by assigning coordinates with respect to the two axis to each decay event. This way we can not only observe the event distribution over the phase space associated to the process but also obtain important information about the mechanisms present in it.

The two independent parameters used in the DP are the two-particle invariant masses m_{ij}^2 of the final-state particles, which implies a constant density of states across the space phase, and are given by:

$$m_{ij}^2 = s_{ij} = (p_i + p_j)^2 = (E_i + E_j)^2 - (\vec{p_i} + \vec{p_j})^2, \qquad (2\text{-}25)$$

where $i, j = 1, 2, 3$, with each index related to the number of the respective daughter particle, as illustrated in Figure 2.6. Straight lines like the dashed one correspond to the loci of the configurations with constant m_{23}^2, in the case where m_{12}^2 and m_{13}^2 correspond to the graph axes.

By observing Figure 2.6, it is possible to notice that there is a closed region where the allowed events by kinematics take place. It can be proved that the density of points should be uniform when there are no dynamic factors affecting the decay[29], and because of that, every structure in the density of the plots must be due to dynamic rather than kinematic features. Suppose, for example, a reaction via an quasi-stable two-body intermediate state:

$$X \to (ab)c \to abc. \qquad (2\text{-}26)$$
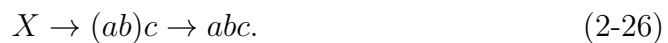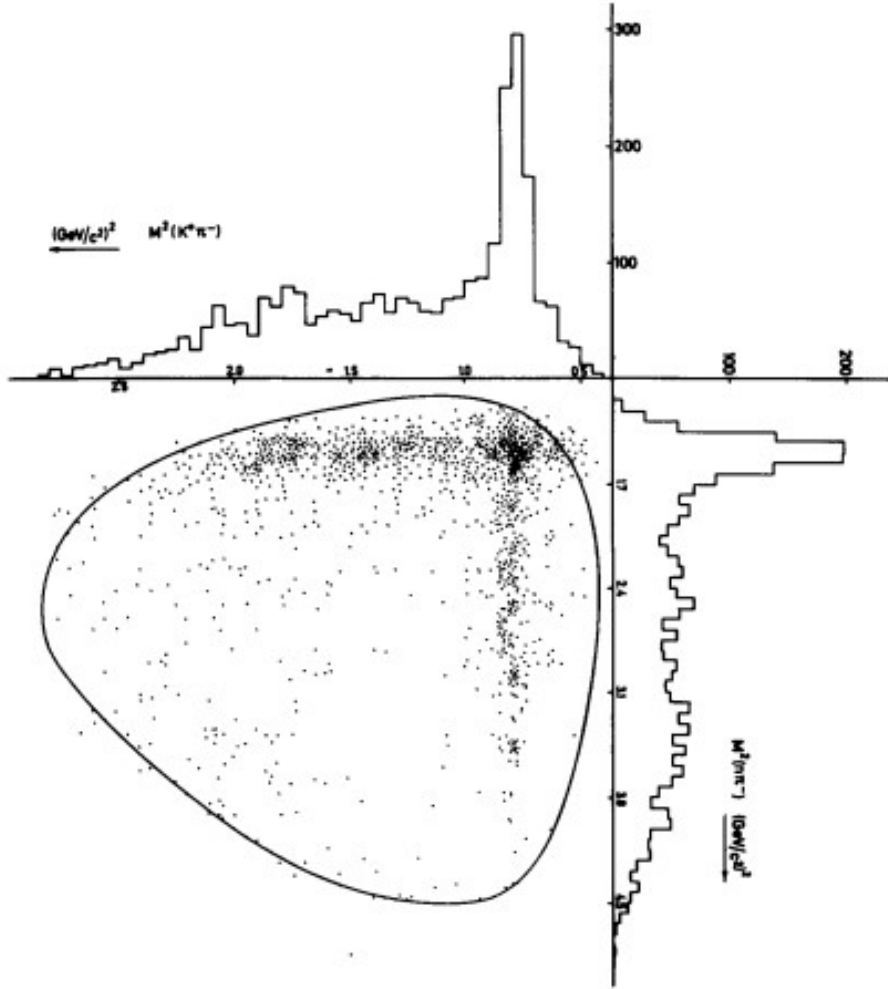
Figure 2.7: Dalitz plot of the $\overline{K}^0\pi^-$ and $n\pi^-$ invariant mass squared for the process $K^-n \to \overline{K}^0\pi^-n$.[30]

In this case, the distribution of m_{ab}^2 is centred around the invariant mass of the intermediate-state particle, so that a band of higher than average density of points appears on the DP, as can be seen in Figure 2.7, for the $K^-n \to \overline{K}^0\pi^-n$ process.

2.6
The $D^+ \to \pi^-\pi^+\pi^+$ and $D_s^+ \to \pi^-\pi^+\pi^+$ decays

As discussed above, processes for which there is a transition between quarks one family apart are called Cabibbo-suppressed. This is the case for the $D^+ \to \pi^-\pi^+\pi^+$ decay, where a D^+ meson $(c\overline{d})$ goes into a π^+ meson $(u\overline{d})$ and a resonant intermediate state $d\overline{d}$, that decays into a pair $\pi^+\pi^-$, as shown in Figure 2.8.

A phenomenon called mixing, where a particle switches between different types or states over time, can influence on the type of CP violation. Direct CP violation happens when a particle and its antiparticle decay into the same
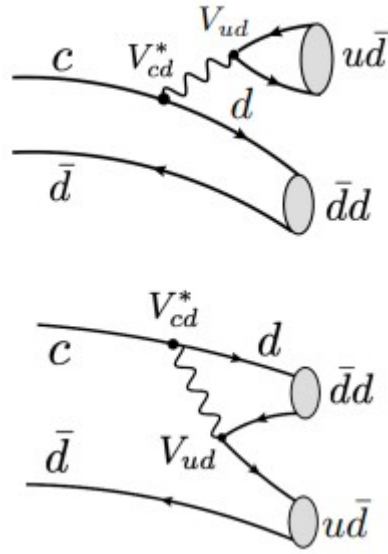
Figure 2.8: Dominant tree-level diagram leading to the $D^+ \rightarrow \pi^-\pi^+\pi^+$ decay.[31]

end products but at different rates, without any mixing going on. Indirect *CP* violation, on the other hand, occurs when the different decay rates are influenced by the particle mixing into different states.

In the case of charged mesons, only direct *CP* violation is possible because mixing can't happen with charged particle, as electric charge has to be conserved. Thus, when searching for *CP* violation in the $D^+ \rightarrow \pi^-\pi^+\pi^+$ decay, we look for differences in the decay rate of the decays of D^+ and D^-. However, this difference might be caused by other effects, such as charge asymmetries in the production of D^+ and D^-, and final-state particles detection and identification.

In order to account for the nuisance asymmetries, we use a control channel for which no *CP* violation is predicted by the SM, but whose non-CP violation related asymmetries are assumed to be the same as the $D^+ \rightarrow \pi^-\pi^+\pi^+$ decay. In this study, we use $D_s^+ \rightarrow \pi^-\pi^+\pi^+$ as control channel.

Both decays have the same dominating diagrams, but the produced intermediate resonances are obtained differently, as shown in Figures 2.8 and 2.9. While in the former, it happens through a $d\bar{d}$, in the latter it is through a $s\bar{s}$ state. This makes these decays have slightly different resonant structures, but addressing this is beyond the scope of this work. From Figure 2.9, we also can see that, as there is a $V_{cs}^*V_{ud}$ interaction, this process is *Cabibbo favoured*, because there are only transitions of quarks of the same generation.

This work focuses on the study of the charge asymmetries present on the $D_s^+ \rightarrow \pi^-\pi^+\pi^+$ decay. Our goal is to identify and quantify the contribution
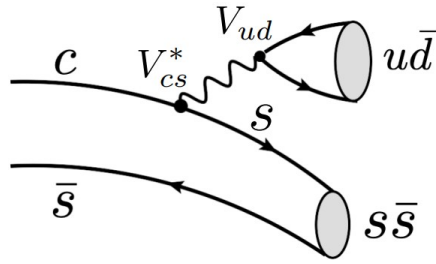
Figure 2.9: Dominant tree-level diagrams leading to the $D_s^+ \rightarrow \pi^-\pi^+\pi^+$ decay.[31]

of each of the non-CP violation effects listed above across the DP. Then, having control of each of those effects, we are then able to study the signal channel $D^+ \rightarrow \pi^-\pi^+\pi^+$, understanding that what differs from the asymmetries observed in the control channel could be an indicative of existence of *CP* violation.

3
The LHCb Experiment

3.1
The Large Hadron Collider

The LHCb is one of the particle physics experiments located at the Large Hadron Collider (LHC), which is the world's largest particle accelerator, achieving the highest energies. It takes place inside a tunnel located near CERN, Geneva, that has 27 km circumference and lies between 45 m and 170 m below the surface[32], being formerly used for the CERN LEP machine[33]. The apparatus aims at colliding proton beams with a centre-of-mass energy of 14 TeV and instantaneous luminosity of 10^{34}cm^{-2}s^{-1} and also Pb ions with an energy of 2.8 TeV per nucleon and a peak luminosity of 10^{27}cm^{-2}s^{-1}. Its circular shape is intended to allow the particles to be accelerated along the path, but with no need to build an excessively long linear tunnel[34].

Inside the accelerator, two proton beams travel in opposite directions at speeds close to the speed of light, each in a separate pipe, kept at ultrahigh vacuum. The beam trajectory is guided by superconducting electromagnets
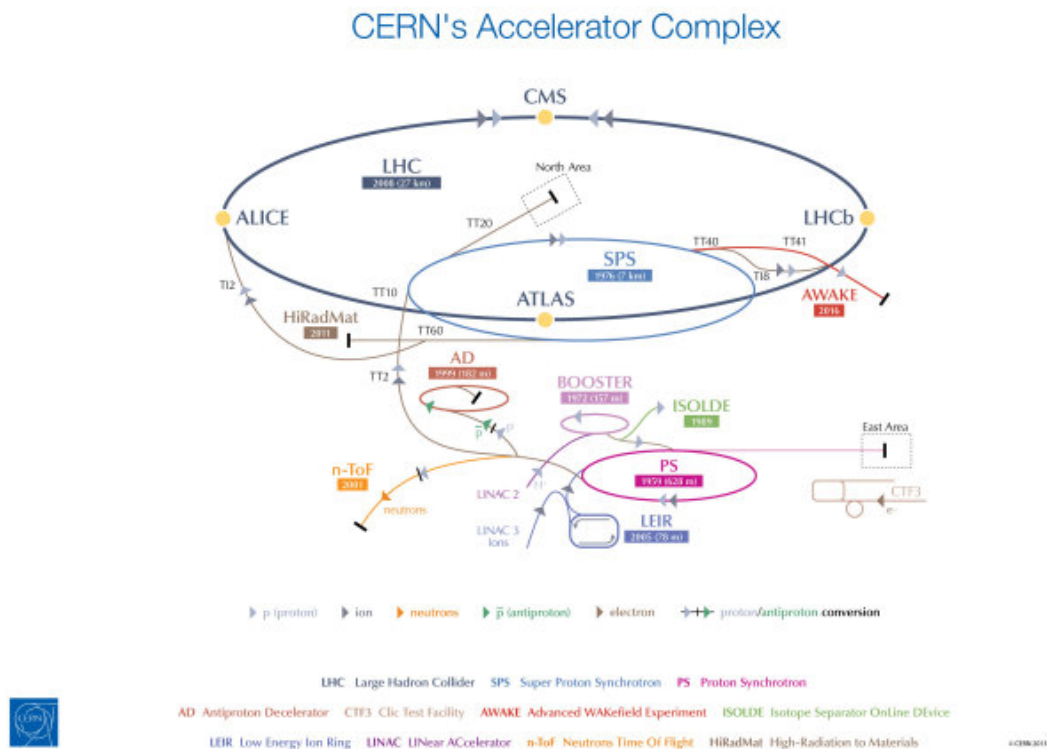
Figure 3.1: CERN's Accelerator Complex and Experiments[35]

chilled down to -271°C by a distribution system of liquid helium.

One may ask where the proton beams come from, before entering the LHC pipes. The beams flow from the so called CERN's Accelerator Complex, portrayed in Figure 3.1, which works so that their energy is boosted at each step, before being injected in the next machine. The last element of the chain, of course, is the LHC[36].

The first component of the Accelerator Complex is the linear accelerator. Although 3.1 shows Linac2, since 2020, the current source of the beams is the Linear Accelerator 4 (Linac4). The Linac4 accelerates negative hydrogen ions (one proton and two electrons), before injecting them into the Proton Synchrotron Booster (PSB), where those two electrons are stripped from the atoms. After that, the protons are accelerated to 2 GeV and injected into the Proton Synchrotron (PS). There, the protons are boosted to 26 GeV and sent to the Super Proton Synchrotron (SPS), where they reach an energy of up to 450 GeV and are finally injected in the beam pipes of LHC.

Inside the LHC rings, the two beams circulate in opposite directions, and after only 20 minutes they reach the energy of approximately 7 TeV, giving a total energy at the collision point of 13.6 TeV, through the use of a dipole field. Finally, the beams are intended to collide in four specific locations of the ring, which correspond to the positions of the four main detectors: ATLAS, CMS, ALICE and LHCb.

3.2
The LHCb detector[1]

The LHCb is an experiment for flavour physics. It studies the *CP* violation in hadron decays with quarks *b* and *c*, as well as rare decays, where indirect evidence of new physics could be found. This is a topic of the utmost importance, as it may help us find sources of *CP* violation outside the Standard Model, thus, we may find better explanations for the asymmetry between matter and antimatter in the Universe.

In addition to ATLAS and CMS, LHCb is more focused in collisions of proton beams (differently from ALICE, focuses on ion collisions). But a huge difference between their structure is that LHCb is a single-arm spectrometer, while the others surround the entire collision point with an enclosed detector[37]. In other words, LHCb detects particles thrown forwards by the collision in one direction. This is because at high energies, b- and $\bar{\text{b}}$-quarks

---

[1]The description of LHCb given here refers to its features at the time of the analysis (2016-2018). Further improvements have been done since then.
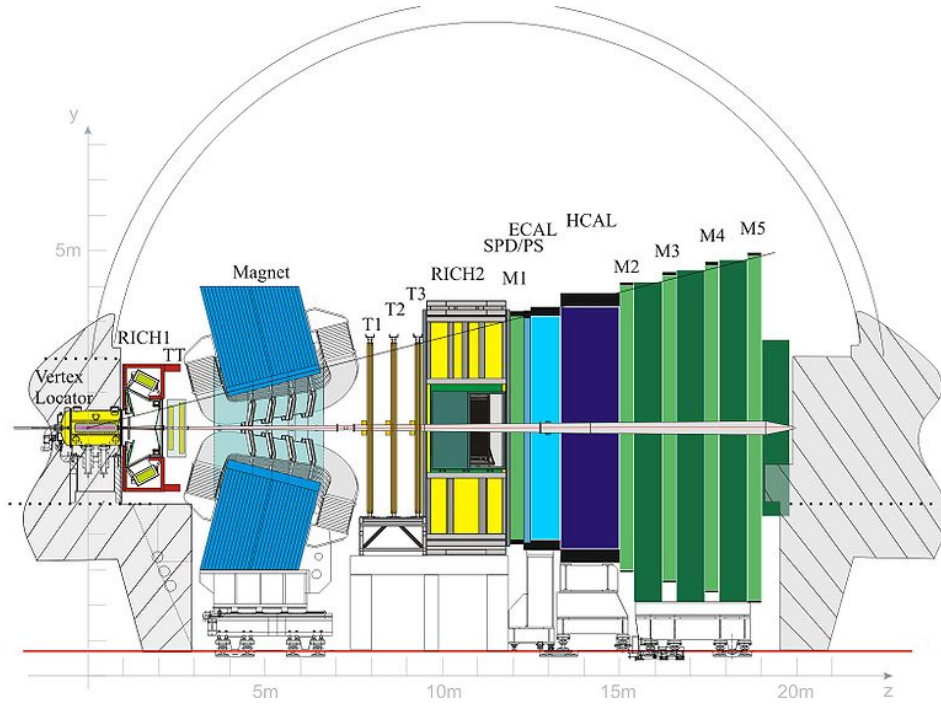
Figure 3.2: Side view of the components of the LHCb subsystem[38]

(and also c- and c̄-quarks) are predominantly produced in the same forward or backward cone.

The detector consists of a series of sub-detectors. The first of them is located closest to the collision point, and the others are placed one after the other, over a distance of 20 meters. The main components of the LHCb subsystem, sketched in Figure 3.2 are briefly discussed in the following subsections.

3.2.1
Magnet

A magnet is used in the LHCb to help measure the momentum of charged particles. The magnetic field bends the paths of positive and negative particles in opposite directions, making it possible to also identify their charge and then calculate their momenta.

The LHCb uses a warm dipole magnet which provides a bending power of about 4 Tm and covers the forward acceptance of ± 250 mrad vertically and of ± 300 mrad horizontally[39]. This magnet consists of two identical coils of conical saddle shape placed mirror-symmetrically to each other, each composed fifteen pancakes arranged in five triplets produced of pure Al 99.7%[39].

Also, the polarity of the magnetic field can be inverted, taking data with the field pointing either up (MagUp) or down (MagDown), to cancel asymmetries caused by overall detector efficiency.

3.2.2
Vertex Locator (VELO)

Right around the collision point, the first component of the so-called Tracking System is the VErtex LOcator (VELO), which aims at providing precise measurements of track coordinates close to the interaction region[39]. It is inside the VELO sub-detector that the proton beams collide.

The VELO is constituted of 42 silicon modules disposed along the beam axis, as shown in Figure 3.3, providing a precise measurement of r and ϕ, in azimuthal coordinates.
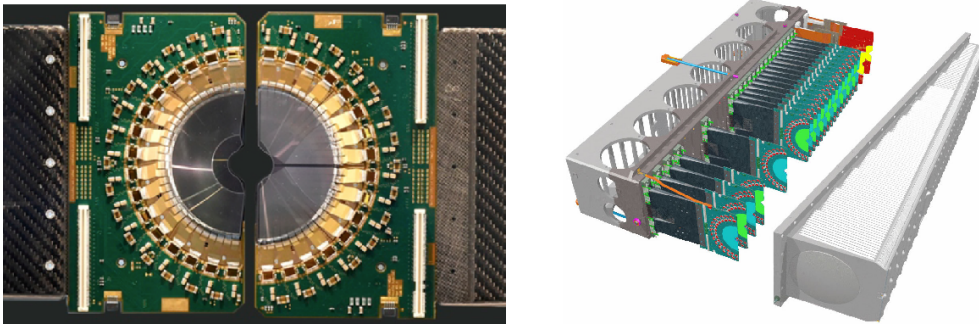
Figure 3.3: Two of the VELO silicon modules (left) and one side of their disposition on the detector (right)[40].

As B- and D-mesons have very short lifetimes, their trajectories are very close to the beam, and for this reason, VELO modules are positioned at a radial distance of just 7 mm from the collision point. Thereby, the task of VELO is to reconstruct those particles from the multitude of particles produced by the collision by measuring the distance between the collision point and the point where they decay[41].

3.2.3
Silicon Tracker

The Tracking System is also composed of a Silicon Tracker, which comprises the Tracker Turicensis (TT) and the Inner Tracker (IT), both constructed of p$^+$-on-n silicon microstrip detectors.[42]. The main goal of the Silicon Tracker is to provide efficient reconstruction of charged-particle tracks. For TT and IT, respectively, the total hit efficiency is shown to be more than 99.7% and 99.8%. Using data from 2011, the hit resolution was calculated to be 52.6 μm for the TT and 50.3 μm for the IT.

Each TT module has read-out sectors with one, two, three or four sensors bonded together, in a way that the single-sensor detectors are closer to the beam-pipe. Those sensors are located upstream of the LHCb dipole magnet, covering the full acceptance of the experiment, i.e. the range of particle
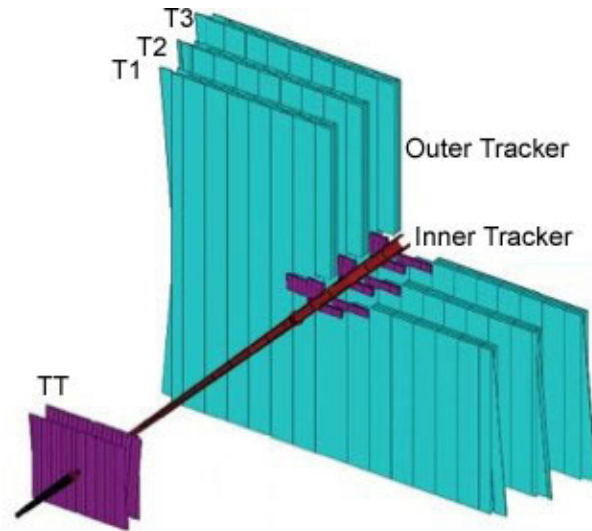
Figure 3.4: Illustration of the disposition of the TT, IT and OT modules on the experiment.[43]

momenta and trajectories that the detector is able to effectively measure, as shown in Figures 3.2 and 3.4.

On the other hand, the IT has two types of sensor. Some are thicker (410 μm) and bonded together on either side. The others are thinner (320 μm) and placed above and below the beam-pipe. They cover a cross-shaped region in the centre of three large planar tracking stations downstream of the magnet, as illustrated in Figure 3.4, namely T_1, T_2 and T_3.

3.2.4
Outer Tracker

The blue region on Figure 3.4 consists of the Outer Tracker, a straw-tube detector that composes the outer parts of the tracking stations T1–T3, and its main goal is the tracking of charged particles and the measurement of their momentum over a large acceptance area[39].

When charged particles pass through the tubes, they ionise gas along their trajectory, allowing us to measure the drift-times of the ionisation electrons to the wire located at the centre of the straw. This information is used then to reconstruct the particles trajectories by getting to know its relative position. The tubes have an inner diameter of 4.9mm and are filled with a gas mixture of Ar (70%) and CO_2 (30%).

3.2.5
RICH

Located on each side of the magnet, as shown in Figure 3.2, the Ring-Imaging Cherenkov (RICH) detectors are two of the components responsible for the particle identification on LHCb. They work by measuring the Cherenkov radiation emitted by the passage of charged particles, which consists of a cone of light that is produced when a particle moves through a medium at a speed greater than the phase velocity of light in that medium. When this cone hits the detector, it produces a ring, so that the momentum and the ring's radius allows us to distinguish different types of hadrons. This detector is a crucial tool when it comes to distinguishing pions and kaons, a very important task in the study of beauty and charm hadrons.

RICH-1, the upstream Cherenkov detector, covers the low momentum charged particle range $\sim 1 - 60$ GeV/c, while RICH-2, covers the high momentum range from ~ 15 GeV/c up to and beyond 100 GeV/c[39]. The former is placed after the VELO, while the latter is placed after the magnet and the tracking system. Figures 3.5 and 3.6 illustrate the structure of each of the detectors.
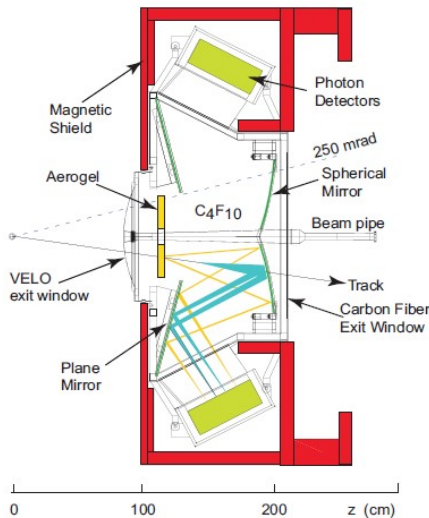
Figure 3.5: Schematic sideview of RICH-1 detector, which uses silica aerogel and C4F10 gas radiators.[39]
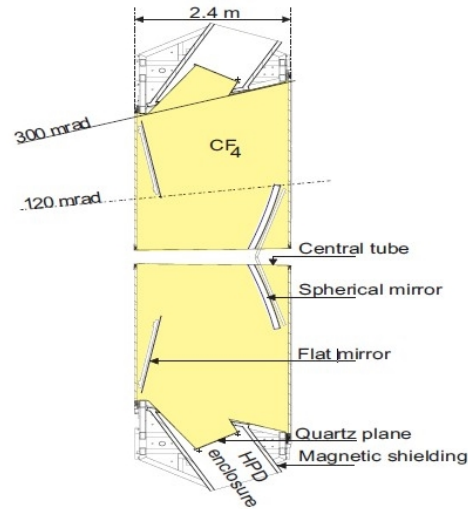
Figure 3.6: Schematic topview of RICH-2 detector, which uses a CF4 gas radiator.[39]

3.2.6
Calorimeter System

Another component that works on particle identification, as well as energy measurements, is the Calorimeter System. It identifies electrons, photons, and hadrons by stopping the particles as they pass through the detector

and measuring the amount of energy lost by them as each one stops. The passing particles produce showers of secondary particles inside the detector, which excite the medium and produce scintillation light, transmitted to a photomultiplier[44].

The identification of neutral pions and prompt photons performed at this stage is a fundamental step, as calorimeters are the main way of identifying neutral particles.

An electromagnetic calorimeter (ECAL) is followed by a hadron calorimeter (HCAL), so that the former measures the energy of electrons and photons, and the latter samples the energy of hadrons.

In addition to the two calorimeters, the system is composed of other layers: the Scintillating Pad Detector (SPD) and the Pre-Shower Detector (PS). The SPD and PS are composed of scintillating pads 15 mm thick. While the SPD determines the charge of the incident particles, by detecting their ionization patterns and analyzing the curvature of their trajectories in a magnetic field, the PS auxiliates the identification of electromagnetic showers initiated by electrons or photons at the ECAL. In association with the ECAL, they are used at trigger level to indicate the presence of electrons, photons, and neutral pions.

As the hit density varies over the calorimeter surface, the layers adopt a variable lateral segmentation. While SPD, PS and ECAL possess a segmentation in three sections, HCAL is segmented in only two zones, but with larger cell sizes, as shown in Figure 3.7.
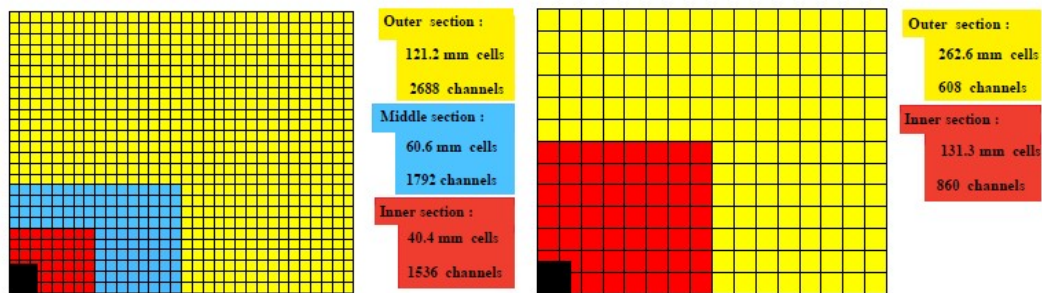
Figure 3.7: Illustration of one quarter of the front face of the calorimeter system layers. In the left, the lateral segmentation of the SPD, PS and ECAL, but with the cell dimensions corresponding to ECAL. On the right, the segmentation of the HCAL[39].

After the SPD and the PS, the following layer corresponds to the ECAL. It uses a 'shashlik' technology of alternating scintillating tiles and lead plates and its cell sizes varies from 4 x 4 cm to 12 x 12 cm as it goes further. The HCAL is positioned after the ECAL, and it is composed of thin iron plates

interspaced with scintillating tiles paralel to the beam pipe. Similarly to ECAL, its cell sizes vary from 13 x 13 cm to 26 x 26 cm as it goes further.

3.2.7
Muon Detector

The last component of the LHCb subsystem is the Muon Detector. Muons are present in the final states of many decays, and provide a tag of the initial state flavour of the accompanying neutral mesons[45], making muon detection a fundamental step for the LHCb experiment.

Five rectangular stations make up the muon system, and their size increases from one to the other, such that together, they all cover a total area of 435 m^2, about the same as a basketball court. The positions of these stations are shown in Figure 3.2.

Each of the muon system stations is divided into four regions, with different logical-pad dimensions, each twice the size of the previous one and with increasing distance from the beam axis, as shown in Figure 3.8, that illustrates the view of one quadrant of muon station 2. Despite their different sizes, each region have approximately the same geometrical acceptance.

Figure 3.8: Scheme of one quadrant of the muon station 2.

The detectors provide binary (yes or no) information about the muon detection, employing distinct readout methods in different parts of the detector, namely anode wire readout, cathode pad readout or both. Anode wire readout involves measuring electrical signals generated by muons passing near fine wires under high voltage, while cathode pad readout captures charge dis-

tributions on metal pads to map the particle's trajectory as it moves through the detector.

3.2.8
Trigger

Approximately 40 million proton-proton collisions occur every second inside the LHCb detector, making it a really complicated task to register the information extracted from all those events, as it would require an enormous storage capacity. So in order to select the events of interest, it is necessary to employ an electronic system called "trigger", that uses information from the sub-detectors described above to choose the events to be recorded[42].

This selection is made in two levels, by two different systems: the Level-0 (L0) and the High Level Trigger (HLT). The former makes decisions in real-time based on information from the calorimeter and muon systems, reducing the original rate of 10 million events per second to only 1 million, discarding the information of the remaining events. The latter takes more time to make its decisions, as the events filtered by the first level trigger are fed to 2,000 state-of-the-art computers, located at the LHCb site, in order to select the interesting events to save. At the end of this second step, the rate of events falls down from 1 million to 2,000 per second.

The criteria used by the L0 trigger is to select high transverse energy electrons, photons or hadrons from the calorimeters or high transverse momentum muons from the muon system. Then, this information is processed by a Decision Unit that derives the final Level-0 decision for each bunch.

Differently, the HLT is divided in two sequential sub-levels: HLT1 and HLT2, which outputs the desired fraction of events. The HLT1 trigger confirms the selection made at L0, by performing a partial event reconstruction adding information from the tracks from VELO. On the other hand, the HLT2 trigger performs a complete event reconstruction, by implementing more rigorous algorithms and using tracks in the VELO as seeds for the rest of the tracking.

4
Data Selection

This chapter outlines the methodology employed for selecting $D_s^+ \to \pi^-\pi^+\pi^+$ data and obtaining the final sample used in the analysis. It also discusses the generation of the simulations used as a proxy in the selection, for the study of specific background contributions (contamination by other D_s^+ decays).

The information that comes from the LHCb detectors has to be reconstructed, in order to select true $D_s^+ \to \pi^-\pi^+\pi^+$ events. This process starts with an online event selection performed by the trigger systems during data taking, followed by an offline selection. The criteria for selection are carefully chosen to ensure the largest statistical significance in the sample number of events while avoiding distortions throughout the phase space caused by wrongly selected events.

4.1
Data and Monte Carlo samples

4.1.1
Data sample

The analysis is performed using the data collected in *pp* collisions at $\sqrt{s} = 13$ TeV by LHCb in the years 2016-2018, corresponding to $5.6 \, \text{fb}^{-1}$ of integrated luminosity. From approximately 40 million collisions per second, the trigger reduces to around 10 thousand per second, taking in consideration the main features of B and D decays.

Trigger categories can be of two types: TIS or TOS. TIS stands for "Trigger Independent of Signal", and refers to events where the trigger decision is independent of the decay of interest. TOS on the other hand stands for "Trigger On Signal", and refers to events where the trigger decision is based on both the decay of interest and the patterns or characteristics in the data that are indicative of that particular decay. The data set we use comes directly from the HLT2 algorithm, which requires the D_s^+ candidates to be TOS on HLT1 and TIS on any L0 algorithm[46]. The criteria used in each selection step presented throughout this chapter are shown in Appendix A.

The HLT1 algorithm employs a multivariate analysis (MVA) to select the candidates. It consists of a statistical technique used to analyse data sets that involve multiple variables, so that the relationships between two or more variables are studied simultaneously in order to understand how they

are related to each other. The MVA integrates various aspects of the event, like the shape of tracks, momentum, and vertex information, to discriminate between signal and background.

To prepare the samples for analysis, a preliminary event selection is performed using mild requirements, displayed on Table A.4 in Appendix A. This step was executed by the central production[47]. The three-body invariant mass distribution of the data sample that meets all the selection criteria up to the central production level is displayed in Fig. 4.1 and the approximate number of signal candidates is 117.4 million.
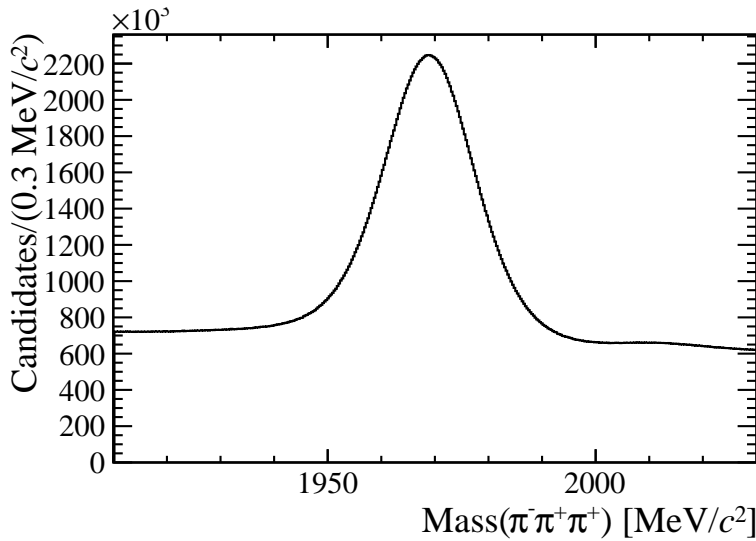
Figure 4.1: Invariant mass distribution of candidates which satisfy all the selection criteria up to the central working group production.

4.1.2
Simulated samples

Monte Carlo (MC) simulations are used for determining the parameters of the signal probability density functions (PDFs) $\mathcal{P}_{sig}(m)$ in invariant mass fits and for modelling the contribution of a specific background channel, namely $D_s^+ \to \eta'\pi^+$.

Events are generated with an amplitude based on a resonant model. This is performed using PYTHIA[48], to generate the *pp* collision. Additionally, the GEANT[49] toolkit is used to simulate the detector response. To improve the efficiency of generation, we discard the events that would not be entirely reconstructed, by applying cuts in the generation step. The number of MC events for the signal channel is shown in Table 4.1

Generator level cuts of the MC samples for the background channel $D_s^+ \to \eta'\pi^+$ were also applied, and the reconstructed events were filtered using

Table 4.1: Number of simulated events available for the signal channel

Year	Polarity	N. of events
2016	Up	5406187
	Down	5442096
2017	Up	5886522
	Down	5889346
2018	Up	5290156
	Down	5256483

the $D_s^+ \to \pi^-\pi^+\pi^+$ HLT2 line. The number of MC events of this background decay available for analysis is shown in Table 4.2.

Table 4.2: Number of simulated events that passed the HLT2 $D_s^+ \to \pi^-\pi^+\pi^+$ line, available for the background channels.

Year	Polarity	N. of events
2016	Up	320239
	Down	324473
2017	Up	320408
	Down	327480
2018	Up	320417
	Down	330805

4.2
Offline Event Selection

4.2.1
Fiducial and cloned tracks cuts

To eliminate regions with a high level of asymmetry not related to the physical phenomena, fiducial cuts are made, through which few events are discarded. This also helps us reduce the amount of background noise. We use the same cuts on both the signal and control samples, so that the control channel acts as a good representation of the asymmetries we expect to see in the signal sample. To improve signal purity and further reduce background noise, we use a MVA.

Low momentum particles can be driven away from the detector's range by the magnetic field, which creates a charge asymmetry. The control sample is used to define cuts that reject regions with large asymmetries, which can be applied to the signal sample as well. The three final particles must also fall

within a momentum and pseudorapidity range where the RICH detectors yield reliable PID, and events with a high SPD hit count are rejected. Additionally, some same-charge pion tracks may be duplicates of a single track, with only slight differences between them. This contribution of cloned π^+ tracks is studied using the distribution of two slope difference variables, dTX_{23} and dTY_{23}[1], which are defined as:

$$dTX_{23} = \frac{p_{x_2}}{p_{z_2}} - \frac{p_{x_3}}{p_{z_3}} \qquad dTY_{23} = \frac{p_{y_2}}{p_{z_2}} - \frac{p_{y_3}}{p_{z_3}}, \tag{4-1}$$

and they are selected by requesting $|dTX_{23}| > 6 \times 10^{-5}$ and $|dTY_{23}| > 6 \times 10^{-5}$.

4.2.2
Specific background contributions

One of the goals of the offline selection process in this study is also to effectively reduce background contributions to the data samples. To achieve this, specific selection criteria are applied to the candidate events in order to minimise the impact of certain sources of background.

Firstly, to reduce the contribution from D_s^+ candidates originating from b-hadron decays, we filter by the value of the impact parameter χ^2 of these candidates. This variable represents the perpendicular distance between the path of a particle before and after an interaction with a target. In addition, to minimise the contamination from partially reconstructed semi-leptonic decays of D_s^+ mesons, candidates for which any of the final state tracks match hits in the muon chambers are rejected. A loose particle identification (PID) cut (ProbNNpi>0.3) is also applied to all daughter particles, in order to reduce kaon contamination from other three-body decays.

Another potential source of background is the $D^+ \to K^- \pi^+ \pi^+$ feed-through, when a true kaon from the opposite charge daughter is misidentified as a pion in the $D_s^+ \to \pi^- \pi^+ \pi^+$ decay. Stringent PID requirements (ProbNNk < 0.2) are put in place to reduce this contamination. To ensure charge symmetry, the same PID cut is applied to all three tracks. The tight PID requirement is imposed on the pions from $D_s^+ \to \pi^- \pi^+ \pi^+$, which significantly reduces the $D^+ \to K^- \pi^+ \pi^+$ background. This requirement results in an efficiency of 98% for D_s^+ signal candidates.

After applying all of the above selection criteria, there still exists a background contribution to the $D_s^+ \to \pi^- \pi^+ \pi^+$ sample, arising from the combination of a $D^0 \to \pi^- \pi^+$ decay with a random π^+. For each D_s^+ candidate, the invariant masses of the $\pi^- \pi^+$ pairs, m_{low} and m_{high}, are calculated (giving rise to the squared invariant masses s_{low} and s_{high} as well). This is done by

combining each of the two positively charged pions $(\pi^+\pi^+)$ and the negatively charged pion (π^-) to form two possible pairs of pions. The lowest invariant mass is defined as m_{low}, and the other pair is defined as m_{high}. This background is reduced to a low level with the requirement $m_{high} < 1.84\,\text{GeV}/c^2$, which keeps essentially 100% of the signal candidates.

Some specific background contributions may still remain in the samples after those selection steps. These are partially or fully reconstructed charm decays, which require the inclusion of a specific model in the mass PDF. Because of that, the only decay mode considered here for such background contributions is $D_s^+ \rightarrow (\eta' \rightarrow (\rho^0 \rightarrow \pi^-\pi^+)\gamma)\pi^+$, where the γ in the final product is not detected.

4.2.3
MVA selection

The MVA selection is a step used to improve the purity of samples by reducing the combinatorial background. We employ a N-fold technique where each year's sample is randomly divided into two folds, resulting in 6 folds (N = 6). The procedure is done by training a Gradient Boosted Decision Tree (BDTG) classifier using the TMVA package on each fold of the data, with simulated $D_s^+ \rightarrow \pi^-\pi^+\pi^+$ decays as the signal input and events from the sidebands as the background input. The training, test and application of the BDTG classifiers are done separately for the D^+ and D_s^+ channels, but the same input variables and hyper-parameters are chosen. The classifiers are then applied to each fold for both data and reweighted MC samples. The final weight for a given fold is randomly selected from one of the five classifiers trained with the other folds to avoid biases in the test and application processes. The yields are estimated by preliminary fits to test data samples with 2M events. The figure of merit to define the response requirement is the signal significance, given by Equation 4-2. The cut chosen was `valBDTG_rand` > -0.45, which is around 90% efficient. The fraction of events with multiple candidates is approximately 0.02%, and these candidates are kept in the sample.

$$\text{Significance} = \frac{S}{\sqrt{S + B}}. \tag{4-2}$$

Since MC samples are used as a proxy for the signal, it is necessary that the distributions of the variables used in the selection match those obtained from the data. However, differences were noticed in the data distributions, hence the procedure is carried out separately for each year and magnet polarity. To accomplish this, we employ the machine-learning algorithm Gradient Boost Reweighter[50].

4.2.4
Background matching procedure

Finally, an extra step is necessary for the specific background MC samples. In order to make sure that we are taking the right decay, we check the daughter particles genealogy through a matching procedure.

Since the decay we want to get is the $D_s^+ \rightarrow \eta'(\rho^0\gamma)\pi^+$, the mother particle of the first pion, for example, must be a $\rho(770)^0$, and the grandmother a η'. To guarantee that, we check Particle Data Group's (PDG) Monte Carlo Particle Numbering Scheme[51] and require that the values of the `MC_MOTHER_ID` and `MC_GD_MOTHER_ID` variables of each particle is equal to the corresponding number for the desired particle.

Our possibilities are:

- `p1_MC_MOTHER_ID` $= 113$,
  `p1_MC_GD_MOTHER_ID` $= 331$,
  `p2_MC_MOTHER_ID` $= 113$,
  `p2_MC_GD_MOTHER_ID` $= 331$ and
  `p3_MC_MOTHER_ID` $= 431$

  or

- `p1_MC_MOTHER_ID` $= 113$,
  `p1_MC_GD_MOTHER_ID` $= 331$,
  `p2_MC_MOTHER_ID` $= 431$,
  `p3_MC_MOTHER_ID` $= 113$, and
  `p2_MC_GD_MOTHER_ID` $= 331$,

so that we make sure that we are getting actual $D_s^+ \rightarrow \eta'(\rho^0\gamma)\pi^+$ events.

5
Study of the $D_s^+ \to \pi^-\pi^+\pi^+$ channel asymmetries

The main goal of this study is to understand the influence of instrumental and production charge asymmetries on the control channel $D_s^+ \to \pi^-\pi^+\pi^+$. We identify these sources of asymmetry and subtract them when we look the signal channel, as mentioned on Chapter 2. Then, the difference in their total asymmetry must be due to physical contributions, indicating the possibility of existence of CP violation in the $D^+ \to \pi^-\pi^+\pi^+$ decay.

In order to understand how each of these sources of asymmetry affects the DP locally, we use the Mirandizing technique[52], that provides a model independent mapping of local charge asymmetries, as will be explained later in this chapter.

To do that, we generate MC samples and inject the asymmetry contributions on them one by one, according to measurements previously performed[53][54][55], as will be discussed on the next session. Following this, we divide the DP in 2D cells and, for each of those, use the Mirandizing technique to compute the significance on the difference on the numbers of D_s^+ and D_s^- candidates.

5.1
Preparation of the Monte Carlo samples

As explained above, it is necessary to use simulated samples of $D_s^+ \to \pi^-\pi^+\pi^+$ to understand how each asymmetry component affects each region of the DP. Full simulation samples, though, typically require a long time for their generation, reconstruction and selection, in addition to the large storage requirements to save them. Because of that, we used RapidSim[56], a tool that allows us to quickly generate samples through a particle-gun-like approach, where only the decaying particle is simulated, not the underlying event, thus saving time and storage.

The parent production kinematics from Full MC are used to generate those samples, using the same samples sizes as data. Also, the same selection criteria as data were used whenever it was possible. Additionally, vertex and IP smearing expressions have been modified to correspond to the actual efficiency of the LHCb detector.

Because we need the RapidSim samples to be as realistic as possible in comparison to the data samples, it is necessary to carry out a reweighting procedure. It consists on assigning weights to the distribution of a specific

variable in the target sample. The goal is to adjust the target samples distribution of that variable to closely match the distribution of the same variable in the original (source) sample. In our case, the target distributions are variables of the RapidSim samples, and the original ones are those same variables of the data samples.

But prior to the reweighting, the background of the data sample is subtracted via $_s\mathcal{P}lot$ technique[57], to make sure we take into account in the procedure only the signal contribution. This technique consists in assuming that the events are characterised by two sets of variables: those for which the distributions of all the sources of events are known, called *discriminating* variables, and those for which the distributions of some sources of events are either truly unknown or considered as such, called *control* variables. In this sense, the $_s\mathcal{P}lot$ technique allows us to reconstruct the distributions for a control variable based on the distributions of a discriminating one.

This is done by performing an invariant mass fit on the data sample, and assigning a signal and a background weight to each event, based on the signal and background components of the distribution. The signal PDF is expressed by the the sum of one Gaussian and two Crystal Ball functions[58], which consist of a Gaussian core portion with a power-law low-end tail:

$$CB(m \mid \alpha, n, \mu, \sigma) = N \begin{cases} \exp\left(-\frac{(m-\mu)^2}{2\sigma^2}\right), & \text{for } \frac{(m-\mu)}{\sigma} > -\alpha \\ A\left(B - \frac{(m-\mu)}{\sigma}\right)^{-n}, & \text{for } \frac{(m-\mu)}{\sigma} \leq -\alpha \end{cases} \tag{5-1}$$

where

$$A = \left(\frac{n}{|\alpha|}\right)^n \exp\left(-\frac{|\alpha|^2}{2}\right), \quad B = \frac{n}{|\alpha|} - |\alpha|,$$

$$N = \frac{1}{\sigma(C+D)}, \quad C = \frac{n}{|\alpha|}\frac{1}{n-1}\exp\left(-\frac{|\alpha|^2}{2}\right),$$

$$D = \sqrt{\frac{\pi}{2}}\left(1 + \mathrm{erf}\left(\frac{|\alpha|}{\sqrt{2}}\right)\right),$$

so that the signal PDF ends up to be:

$$\mathcal{P}_{\mathrm{sig}}(m) = [(1 - f_{CB_1} - f_{CB_2}) \times G(\mu_1 + \Delta\mu, \sigma_1 + \Delta\sigma)] + $$
$$[f_{CB_1} \times CB_1(\mu_1, \sigma_1, \alpha_1, N_1)] + \tag{5-2}$$
$$[f_{CB_2} \times CB_2(\mu_2, \sigma_2, \alpha_2, N_2)],$$

where f_1 and f_2 are the PDF fractions for the two Crystal Balls, and $\Delta\mu$ and $\Delta\sigma$ are the mean and sigma offsets for the Gaussian function, related to the

values of the first Crystal Ball.

The background PDF, on the other hand, corresponds to a second-order Bernstein polynomial ($n = 2$):

$$\mathcal{P}_{bkg}(m) = \sum_{i=0}^{n} a_i \begin{pmatrix} n \\ i \end{pmatrix} m^i \cdot (1 - m)^{n-i}, \tag{5-3}$$

where $a_2 = 1$ was fixed.

The plots of the data with $_s\mathcal{P}lot$ weights are in Figure 5.1, with the samples separated by the year they were taken and the polarity of the LHCb magnet (Chapter 3) when they were measured.

Besides performing the $_s\mathcal{P}lot$ on the data sample, it is also necessary to take into account the particle identification (PID) efficiency on the RapidSim sample before the reweighting procedure. Those values are also obtained through calibration data[54], and each event receives a PID efficiency weight, depending on the daughter particle involved, and also the year and polarity of the detection. As the generated MC samples do not correspond to any measurement at first, it is necessary to define a sample to represent each year and each polarity, with the same size of the data ones, for us to consider the specific weights for the PID efficiency.

In possession of the signal weights for the data sample and the PID efficiency weights for the RapidSim sample, we are able to proceed to the reweighting. The variables to be reweighted are the momentum p, the transverse momentum p_T (that corresponds to the component of p perpendicular to the beam line) and the pseudorapidity η of the daughter particles, and also the invariant masses squared s_{12} and s_{13}. The pseudorapidity is calculated as:

$$\eta = -\ln \left[\tan \left(\frac{\theta}{2} \right) \right], \tag{5-4}$$

where θ is the angle between the positive direction of the z axis and the particle momentum. And the invariant masses squared s_{ij} (not to be confused with the sines of the angles of the CKM matrix) are given by:

$$s_{12} = (p_1 + p_2)^2 = (p - p_3)^2,$$
$$s_{13} = (p_1 + p_3)^2 = (p - p_2)^2, \tag{5-5}$$
$$s_{23} = (p_3 + p_3)^2 = (p - p_1)^2,$$

where

$$p = p_1 + p_2 + p_3. \tag{5-6}$$

The procedure is carried out with GBReweighter from the hep_ml library[50] of Python. For the predicted weights to be unbiased, besides the six original MC samples (corresponding to MagUp and MagDown polarities for the 2016, 2017 and 2018 samples), six other samples were generated
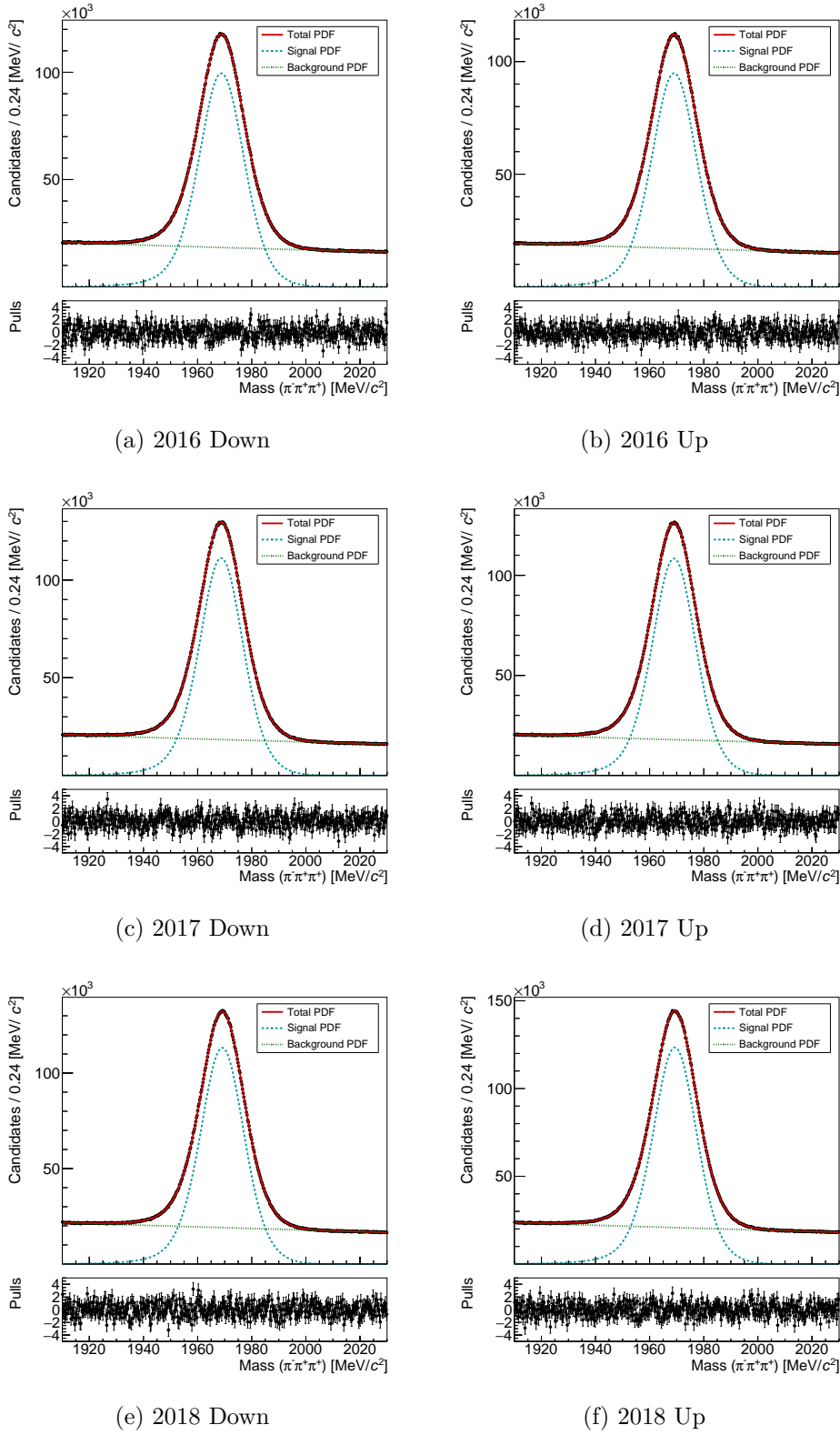
(a) 2016 Down

(b) 2016 Up

(c) 2017 Down

(d) 2017 Up

(e) 2018 Down

(f) 2018 Up

Figure 5.1: Components ${}_s\mathcal{P}lot$ invariant mass fit of each data sample of the $D_s^+ \to \pi^-\pi^+\pi^+$ channel, separated by year and polarity.

via RapidSim, with 4M events each. Those extra samples were used for the FoldingReweighter training, and then the predicted weights obtained in each of the six trainings were applied on the corresponding original RapidSim sample. The hyperparameters used are listed in Table 5.1.

Hyperparameter	Value
n_estimators	500
learning_rate	0.1
max_depth	6
sub_sample	0.7

Table 5.1: Hyperparameters used on the reweighting procedure.

Figures 5.2 and 5.3 show the distributions of the reweighted variables of the 2016 MagDown samples, before and after the reweighting procedure, respectively, for Data and MC. It is possible to notice how the technique succeeds in providing weights to those distributions so that the MC sample has the same kinematic distribution as the data sample.

The quality of the reweighting procedure can be quantified by observing its associated receiver operating characteristic (ROC) curve. It consists in a graph of the true positive rate (TPR) versus the false positive rate (FPR), which provides us the performance of a classification model. In other words, we can see how distinguishable the MC sample is in comparison to the data sample.

Figure 5.4 shows the ROC curve for the 2016 MagDown sample. The blue curve represents the MC sample before reweighting, and the orange one represents the sample after the procedure. We can see that after reweighting, the ROC curve approaches the dashed line, that represents a random classifier; the closer to this dashed line, the more indistinguishable the distributions are. To quantify that, we measure the area under the curve (AUC), that tells us how good the model is at distinguishing our samples. For example, if AUC = 1, the model is correctly distinguished 100% of the times; if AUC = 0, on the other hand, the model is never correctly distinguished. Given that, an indistinguishable model must have an AUC close to 0.5, that is, a 50% rate of success. We can see then in Figure 5.4 that our procedure was satisfactory, as we got an AUC of 0.505.

The plots for the other years and polarities are in Appendix B.

The final step of the RapidSim samples preparation is the inclusion of the instrumental and production asymmetries. It is assumed that those contributions to the so-called *raw asymmetry* are small, so that higher-order or cross terms would be so small that they could be neglected. Thus the asymmetry can be factorised as:
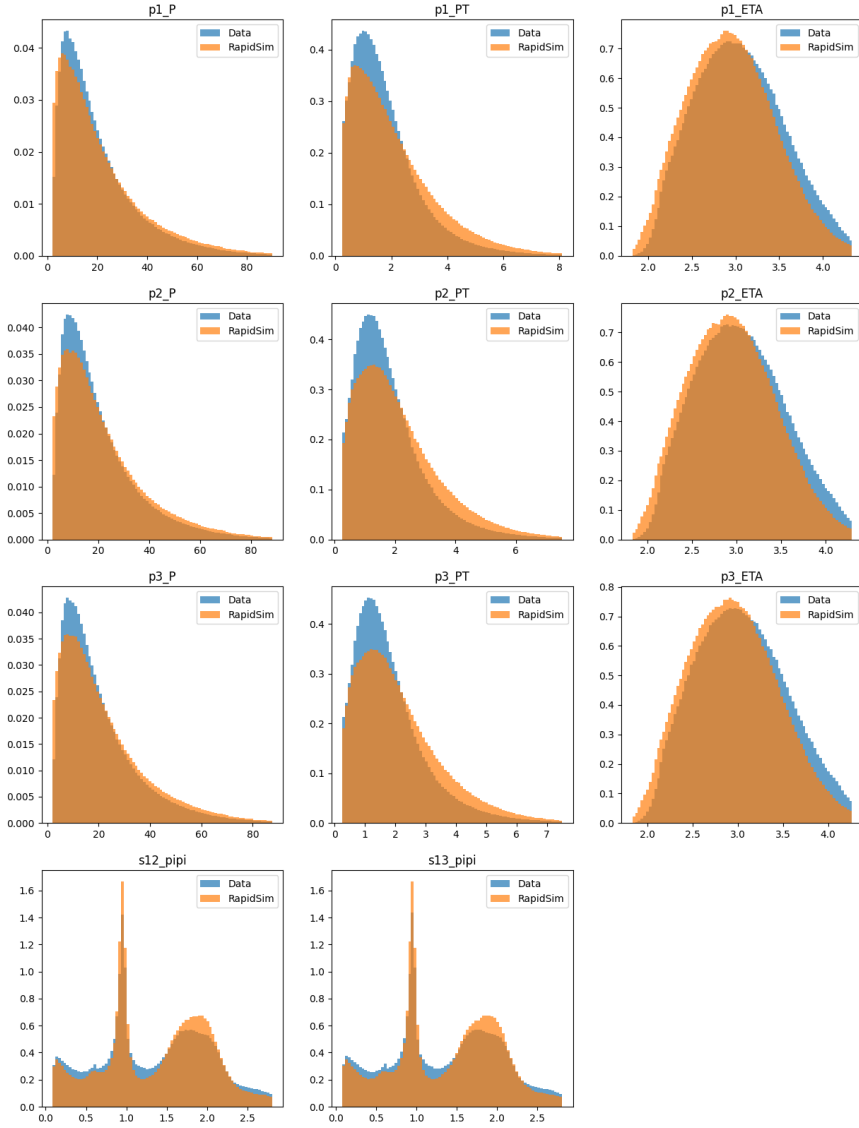
Figure 5.2: Comparison between the reweighted variables distributions of the Data and MC 2016 MagDown sample, before the reweighting procedure. All variables are shown in GeV.
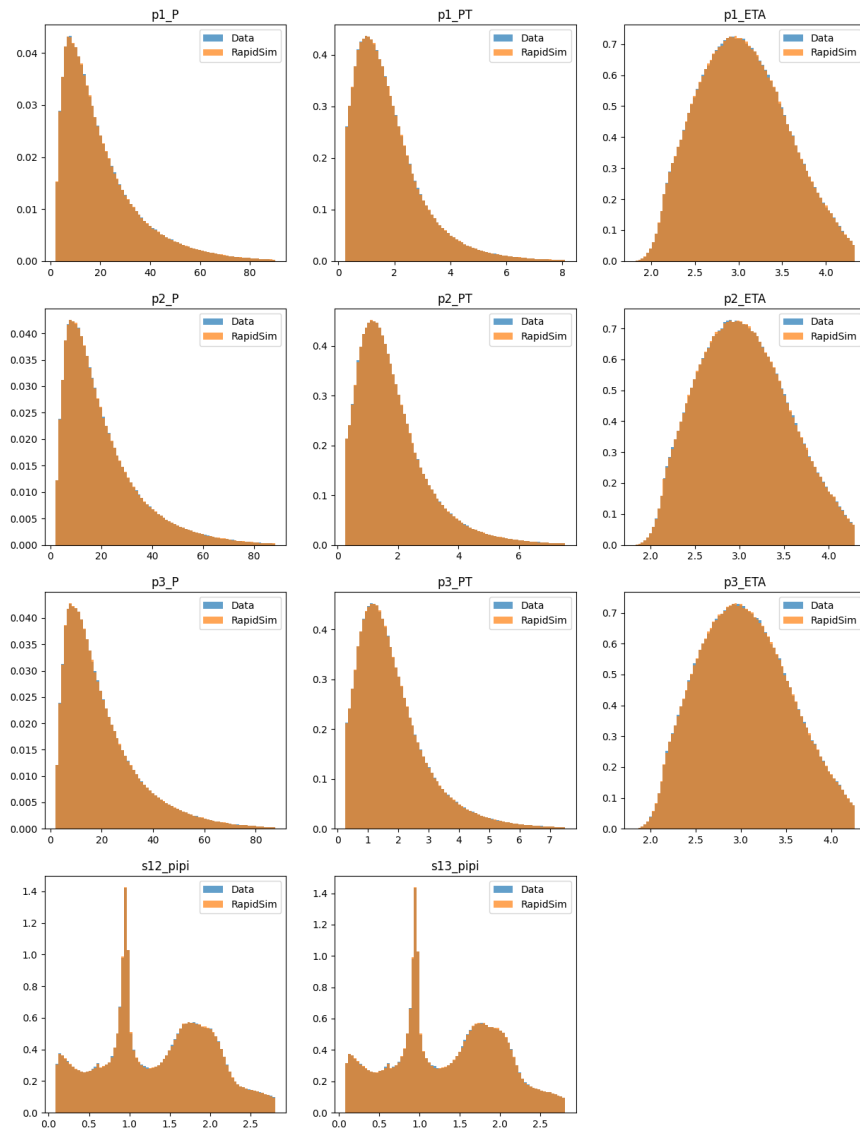
Figure 5.3: Comparison between the reweighted variables distributions of the Data and MC 2016 MagDown sample, after the reweighting procedure. All variables are shown in GeV. We can see then how similar the distributions get after the procedure.
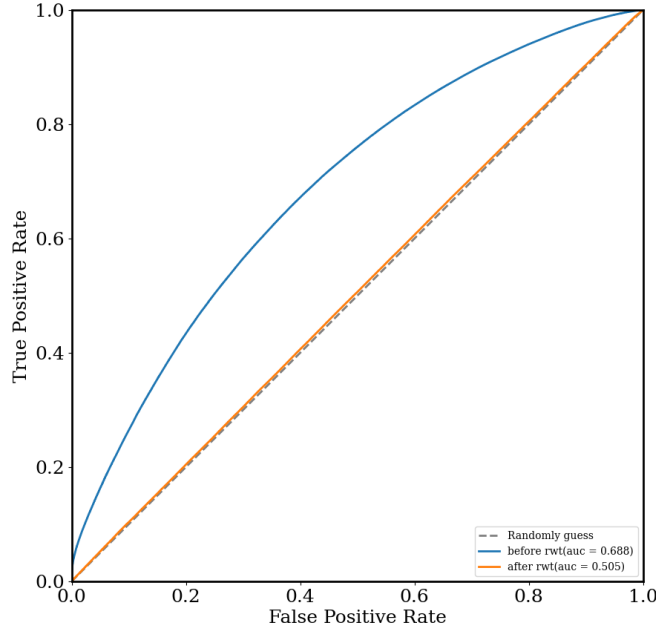
Figure 5.4: ROC curve of the 2016 MagDown reweighted MC sample.

$$A_{\mathrm{raw}} = \frac{N^+ - N^-}{N^+ + N^-} \approx A_{CP} + A_{\mathrm{inst}} + A_{\mathrm{prod}} , \qquad (5\text{-}7)$$

where N^+ and N^+ are the number of candidates for the positive and negative particle, A_{CP} is the CP asymmetry, A_{inst} is the instrumental asymmetry, that is related to the detector and the measurement procedure, and A_{prod} is the production asymmetry, which arises from the hadronisation process in the production of the initial state hadron. The instrumental asymmetry can be factorised in two components, namely the PID induced charge-asymmetry A_{PID} and the other detector related asymmetries, given by A_{det}. This gives us:

$$A_{\mathrm{raw}} \approx A_{CP} + A_{\mathrm{PID}} + A_{\mathrm{det}} + A_{\mathrm{prod}} . \qquad (5\text{-}8)$$

The algorithm for the asymmetry insertion takes the value of certain kinematic parameter for an event and finds the measured value for the corresponding asymmetry. For example, the value for the detection asymmetry A_{det} depends on the momenta of the three daughter particles. The PID asymmetry A_{PID}, on the other hand, takes in consideration both the momentum and the pseudorapidity of the three daughters. The production asymmetry A_{prod}, differently, depends on the transverse momentum of the mother particle, but the reference values are different depending on the rapidity y of this particle, which is given by:

$$y = \frac{1}{2} \log \left(\frac{E + p_Z}{E - p_Z} \right) . \qquad (5\text{-}9)$$

With the calculated values for each asymmetry in hand, it is necessary to implement the following algorithm in order to insert them into the RapidSim sample:

– Compute the total asymmetry A assuming factorisation (Equation 5-8);
– Calculate the ratio:

$$R = \frac{(1+A)}{(1-A)} = \frac{N^+}{N^-}, \tag{5-10}$$

corresponding to the calculated asymmetry.

- – If $R > 1$ and the event is a D_s^-, reject it with probability $(1-1/R)$;
- – If $R < 1$ and the event is a D_s^+, reject it with probability $(1-R)$;
- – Else ($R > 1$ and event is D_s^+ or $R < 1$ and event is D_s^-): accept event.

At the end of this procedure, the RapidSim samples are ready to use for our purposes, as they now have the necessary asymmetries introduced.

Then, rather than the difference between the D_s^+ and D_s^- candidates in each DP bin, as mentioned before, we compute the significance of this difference, called $\mathcal{S}_{CP}$.

The formula to calculate the $\mathcal{S}_{CP}$ is given by:

$$\mathcal{S}_{CP}^i = \frac{N_i^+ - \alpha N_i^-}{\sqrt{\alpha\left(\delta N_i^{+2} + \delta N_i^{-2}\right)}}, \quad \alpha \equiv \frac{N^+}{N^-}, \tag{5-11}$$

where $N_i^\pm$ and $\delta N_i^\pm$ are the number of signal candidates in the ith bin of the $D_s^\pm$ DP and its corresponding uncertainty. Similarly, $N^\pm$ is the total number of $D_s^\pm$ signal candidates, integrated over the DP, and the constant α removes the global net asymmetry.

The importance of this measure is that the p-values that result from it give us the degree of confidence that the differences between the D_s^+ and D_s^- Dalitz plots are caused only by statistical fluctuations. The p-value is calculated as the probability of obtaining a test statistic as extreme as, or more extreme than, what was observed, assuming the null hypothesis is true. For that reason, a p-value smaller than 3×10^{-7} corresponds to an observation of CPV with a significance larger than 5σ.

But since our RapidSim samples possess only signal events, there is no uncertainty associated to the unsure the candidates correspond to signal or background. For that reason we can use the error associated to the number of signal candidates in the data sample as input on the RapidSim $\mathcal{S}_{CP}$ calculation. To obtain those, we must perform a fit to its mass distribution, as will be explained on the next session.

5.2
Mass fit on the data sample

In order to obtain the values for $N_i^\pm$ and $\delta N_i^\pm$ in Equation 5-11, it is necessary to distinguish the signal and background contributions to the mass distribution of the sample in question.

While the RapidSim samples possess only signal events, the data samples contain also background contributions. To distinguish between the signal and background distributions, we perform a mass-per-bin fit on the data sample, parameterising the signal and the background, each with a different PDF.

Differently from the approach on the $_s\mathcal{P}lot$ step, here we perform the fits in each bin separately to take in consideration the differences among each region of the DP. As shown by Equation 5-11, the $\mathcal{S}_{CP}$ will be computed for each bin, so we need the best precision possible in the determination of the number of D_s^+ candidates, as well as its associated error.

This bin division could be done in many different ways, each of them with distinct motivations and advantages. In this work, we chose a physics motivated binning, in which each of the bins comprehends regions with different resonance contributions or specific kinematics, in order to take into account the local effects generated by each component of the resonant structure across the DP. The interference between the S-wave and the P-wave, which are components of the wave function, makes the CP asymmetry change signs when crossing the nominal mass of the resonances, as well as the helicity angle, which is the angle between a particle's direction of motion and its spin.

The used binning scheme is presented in Figure 5.5. To show how the bins are related to the resonant structure of the decays, Figures 5.5a and 5.5b display it placed over the $D^+ \to \pi^-\pi^+\pi^+$ and $D_s^+ \to \pi^-\pi^+\pi^+$ Dalitz plots, respectively. The $D^+ \to \pi^-\pi^+\pi^+$ decay was used as reference for this binning scheme, as it is ultimately our object of study. One may notice though that there are some differences in the shape and size of the DP's in Figures 5.5a and 5.5b. To take this into account, for the $D_s^+ \to \pi^-\pi^+\pi^+$ control channel, some of the bins in the border were extended to include the parts that were beyond the limits of the original binning scheme used for the signal channel.

With the binning scheme at hand, we have what we need to proceed to the mass-per-bin fits. In order to do that, we create mass histograms with the events of each bin, in order to fit those distributions separately. As we take in consideration the local differences of the mass distributions in each bin, each fit has its own particularities. In most of the cases we can use the same function for these contributions in all bins, and those differences result only in different parameters for the signal and background PDF's. But in some cases, specific
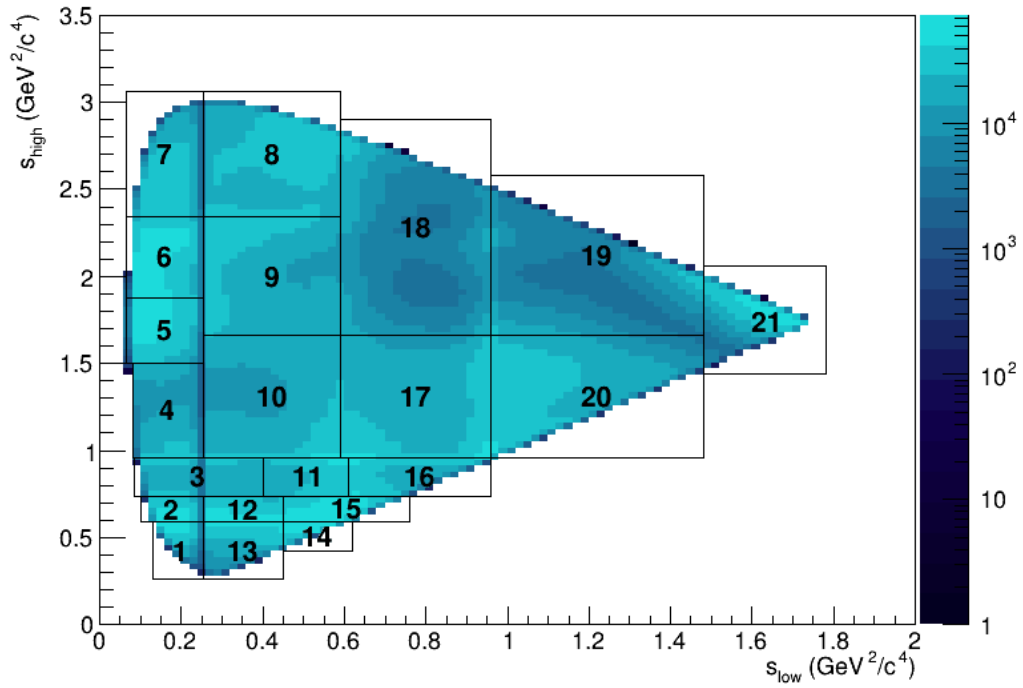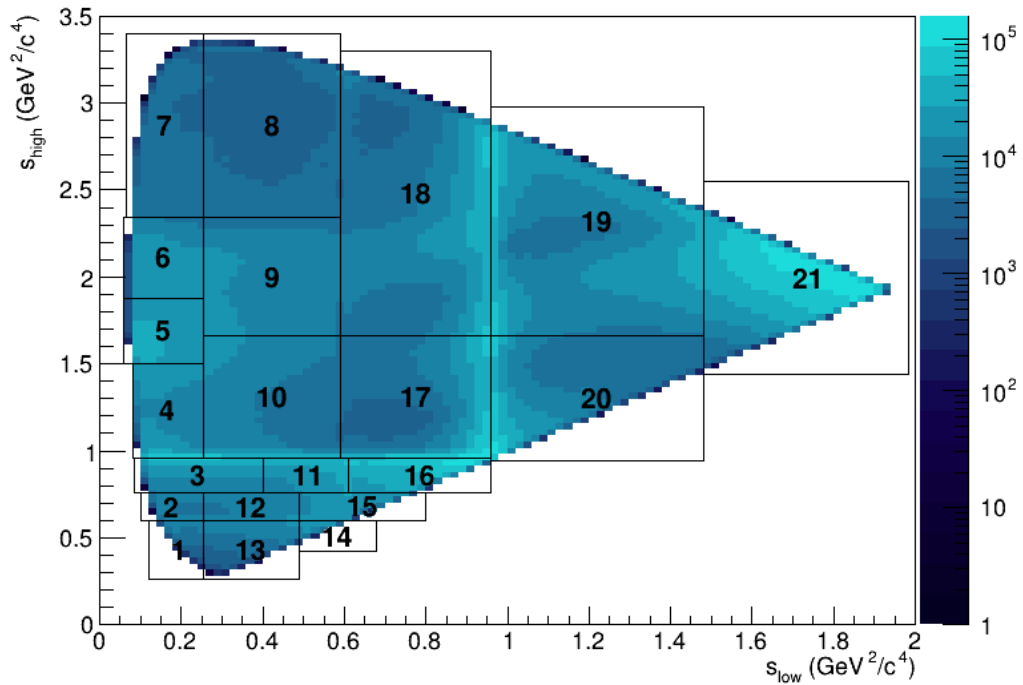
(a) $D^+ \to \pi^-\pi^+\pi^+$

(b) $D_s^+ \to \pi^-\pi^+\pi^+$

Figure 5.5: Physics motivated binning scheme used in this work, placed over the $D^+ \to \pi^-\pi^+\pi^+$ and $D_s^+ \to \pi^-\pi^+\pi^+$ Dalitz plots, respectively. The colors represent the number of candidates in that region

background contributions show up, and must be dealt with.

The signal PDF is described by the sum of one Gaussian and two Crystal Ball functions similarly to what we do in the $_s\mathcal{P}lot$ case, as shown in Equation 5-2.

On the other hand, the background PDF can have two different contributions. One of them corresponds to the *combinatorial background*, given by random three-track associations, and is expected to be the major background contribution to the data samples. As it comes from combinations of uncorrelated tracks which pass the selection algorithms, this background is expected not to introduce any structure on the DP, and have a broad distribution over the invariant mass of the three pions. This background is described by a second-order Bernstein polynomial, detailed in the $_s\mathcal{P}lot$ step above (Equation 5-3).

The other background contribution comes from partially reconstructed $D_s^+ \rightarrow (\eta' \rightarrow (\rho^0 \rightarrow \pi^-\pi^+)\gamma)\pi^+$ decays without misidentified particles, and is called *specific background*. One may notice that the intermediate process $\eta' \rightarrow \rho^0\gamma$ has a γ as sub-product, giving rise to a four-particle final state. As a consequence of that, the invariant mass of the 3 pions appears as a background contribution at the lower mass region of $D_s^+ \rightarrow \pi^-\pi^+\pi^+$, and is parameterised by a Gaussian function (more specifically its high mass tail), as shown in Figure 5.6. However, as it comes from a resonance, it appears only at the $\rho(770)^0$ region on the DP; thus, this contribution to the total PDF is only included at the bins in this DP region, as demonstrated in Figure 5.7.
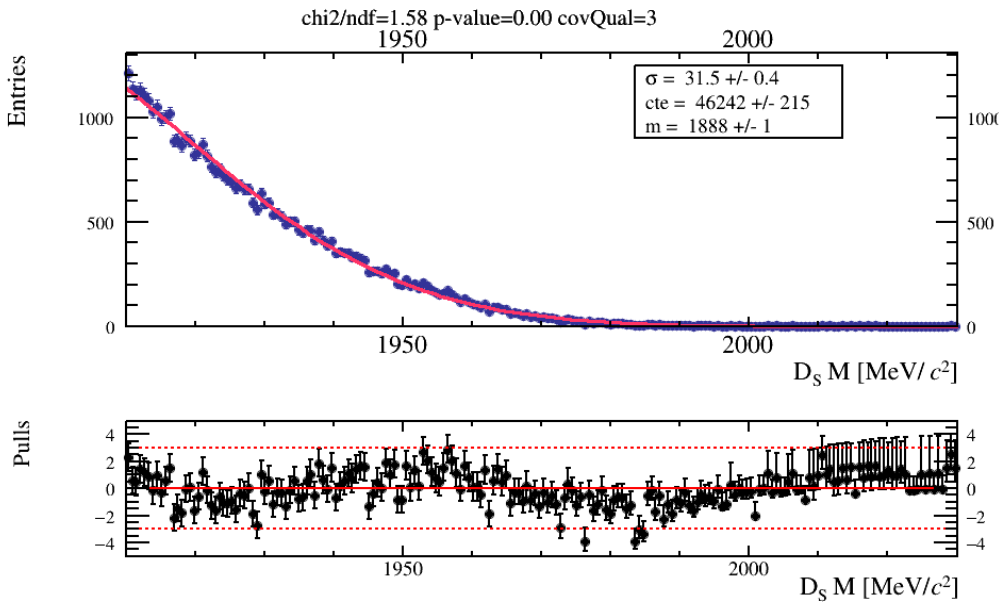
Figure 5.6: Full Monte Carlo Distribution of the $D_s^+ \rightarrow \eta'\pi^+$ decay, fitted with the right tail of a gaussian function.
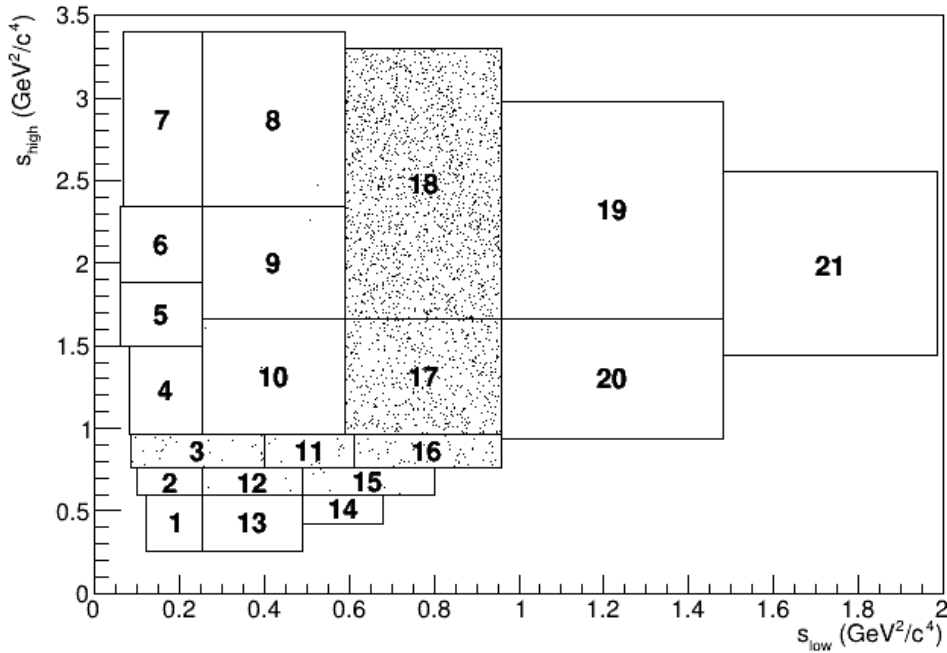
Figure 5.7: Distribution of the $D_s^+ \to \eta' \pi^+$ MC events across the $D_s^+ \to \pi^- \pi^+ \pi^+$ Dalitz plot.

But in order to fit those mass distributions, we keep constant the parameters that we already know the value, and leave free only the parameters that we do not know the value beforehand. For example, for the signal PDF, we can obtain the parameters of its shape through the fit of the Full MC simulations of the $D_s^+ \to \pi^- \pi^+ \pi^+$ decay, as it must have the same behaviour as the data samples, with the advantage of not having background contamination. Similarly, we can find the parameters of the Gaussian background through the fit of the $D_s^+ \to \eta' \pi^+$ MC sample. It is important to stress out that this $D_s^+ \to \eta' \pi^+$ MC sample must have passed through the same selection criteria used by the data sample, detailed in Chapter 4, so that we make sure that it has a similar behaviour to the actual specific background.

For the signal PDF (Equation 5-2) the ratio between the σ of the Gaussian function and the Crystal Balls, the Gaussian fraction, the α and N parameters of the Crystal Balls and the mean offset are fixed by the MC fit results.

When it came to the specific background PDF, on the other hand, all the parameters of the Gaussian function were fixed, with different values for each of the bins with $D_s^+ \to \eta' \pi^+$ contribution to the background. Also, an extra parameter was included, to take into account the fraction of the total background that comprehends the $D_s^+ \to \eta' \pi^+$ events.

Additionally, we can estimate how many $D_s^+ \to \eta' \pi^+$ events are expected

in each bin of the $D_s^+ \to \pi^-\pi^+\pi^+$ phase space. It can be done by using information from the Full MC samples.

The number of observed events from the $D_s^+ \to \eta'\pi^+$ decay can be obtained by:

$$N_{obs}^{(i)} = N_{prod} \times BR \times \epsilon_{tot}^{(i)}, \tag{5-12}$$

which consists on the product of the number of generated decays produced N_{prod}, the branching ratio associated to the process of interest BR and its total efficiency $\epsilon_{tot}^{(i)}$. The (i) index indicates that this equation is valid for the entire DP or for a specific bin i.

As we can not have direct access to the number of produced D_s^+ particles, we used previous results for an ongoing study of the $D_s^+ \to K^-K^+\pi^+$ channel. For this decay, $N_{obs}^{(i)}$ was obtained through a mass fit, so that we are able to reproduce Equation 5-12 for its case, in order to obtain N_{prod}. Rewriting Equation 5-12, we get to:

$$N_{obs}^{(i)}\left(D_s^+ \to \eta'\pi^+\right) =$$
$$= N_{obs}\left(D_s^+ \to K^-K^+\pi^+\right) \times \frac{BR\left(D_s^+ \to \eta'\pi^+\right)}{BR\left(D_s^+ \to K^-K^+\pi^+\right)} \times \frac{\epsilon_{tot}^{(i)}\left(D_s^+ \to \eta'\pi^+\right)}{\epsilon_{tot}\left(D_s^+ \to K^-K^+\pi^+\right)}, \tag{5-13}$$

where the branching ratios are taken from PDG data[22], with the particularity that for the $D_s^+ \to \eta'\pi^+$ decay, all of the sub-processes involved must be taken into consideration:

$$BR(D_s \to \left(\eta' \to \left(\rho^0 \to \pi\pi\right)\gamma\right)\pi) =$$
$$= BR(D_s \to \eta'\pi) \times BR(\eta' \to \rho^0\gamma) \times BR(\rho^0 \to \pi\pi) \tag{5-14}$$

The total efficiency $\epsilon_{tot}^{(i)}$ is obtained through:

$$\epsilon_{tot}^{(i)} = \frac{N_{events_{MC}}^{(i)}}{N_{requested}}, \tag{5-15}$$

where $N_{events_{MC}}^{(i)}$ is the number of events in the Full MC sample, and $N_{requested}$ is the number of requested MC events, before generation and all the performed cuts, listed in Chapter 4.

The estimated values of the yields of the $D_s^+ \to \eta'\pi^+$ decay are shown in Tables 5.2 and 5.3, and the results in Table 5.4.

It is important to notice that, as displayed in Figure 5.7, the $D_s^+ \to \eta'\pi^+$ contribution to the $D_s^+ \to \pi^-\pi^+\pi^+$ background does not occur in the entire phase space, but only in some of the bins. The bins in which there were enough specific background events to take it in consideration when performing the mass fits are bins 3, 11, 16, 17 and 18.

$BR(D_s^+ \to (\eta' \to (\rho^0 \to \pi^+ \pi^-)\,\gamma)\,\pi^+)$	
BR	$(11.62 \pm 0.75) \times 10^{-3}$
ϵ_{tot}	$(21.74 \pm 0.10) \times 10^{-6}$
$N_{events_{MC}}$	$(46.24 \pm 0.21) \times 10^3$
$N_{requested}$	$(21,265.97 \pm 0.46) \times 10^5$

Table 5.2: Values related to the $D_s^+ \to \eta' \pi^+$ used in the yield calculation.

$D_s^+ \to K^- K^+ \pi^+$	
BR	$(5.38 \pm 0.10) \times 10^{-2}$
ϵ_{tot}	$(482.96 \pm 0.14) \times 10^{-6}$
N_{obs}	$(1805.97 \pm 0.15) \times 10^5$
$N_{events_{MC}}$	$(1,098.12 \pm 0.33) \times 10^4$
$N_{requested}$	$(22,737.23 \pm 0.15) \times 10^6$

Table 5.3: Values related to the $D_s^+ \to K^- K^+ \pi^+$ decay used in the yield calculation.

Bin	Yields
Bin 1	76
Bin 2	5,280
Bin 3	**113,167**
Bin 4	38
Bin 5	228
Bin 6	76
Bin 7	0
Bin 8	1,216
Bin 9	1,558
Bin 10	2,241
Bin 11	**91,780**
Bin 12	22,413
Bin 13	304
Bin 14	152
Bin 15	38,368
Bin 16	**204,035**
Bin 17	**533,697**
Bin 18	**741,226**
Bin 19	304
Bin 20	342
Bin 21	152
TOTAL	**1,756,652**

Table 5.4: Calculated yields for the $D_s^+ \to \eta' \pi^+$ background contribution to the $D_s^+ \to \pi^- \pi^+ \pi^+$ data sample. The boldfaced bins are those for which the specific background contribution was substantial enough to be taken into consideration on the mass-per-bin fits.

Finally, the mass-per-bin fits were done by taking the mass distributions in each bin of the $D_s^+ \to \pi^-\pi^+\pi^+$ Dalitz plot, and performed simultaneously for each of the four components of each year: $D_s^+ \to \pi^-\pi^+\pi^+$ and $D_s^- \to \pi^+\pi^-\pi^-$, MagUp or MagDown. The number of events set to be the same for both polarities.

Figure 5.8 shows two examples: bins 18 and 19. In the former, the specific background was taken into account, while for the latter it was not necessary (Figure 5.7 and Table 5.4). The fit results of the other bins of the $D_s^+ \to \pi^-\pi^+\pi^+$ DP are in Appendix C.
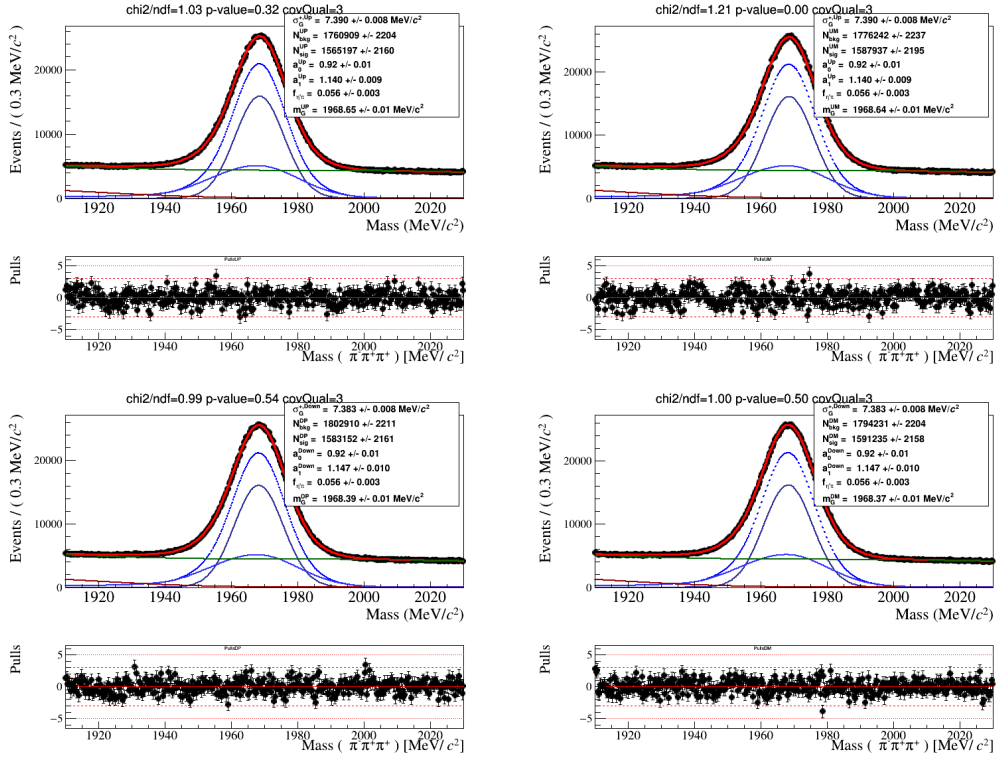
(a) Bin 18
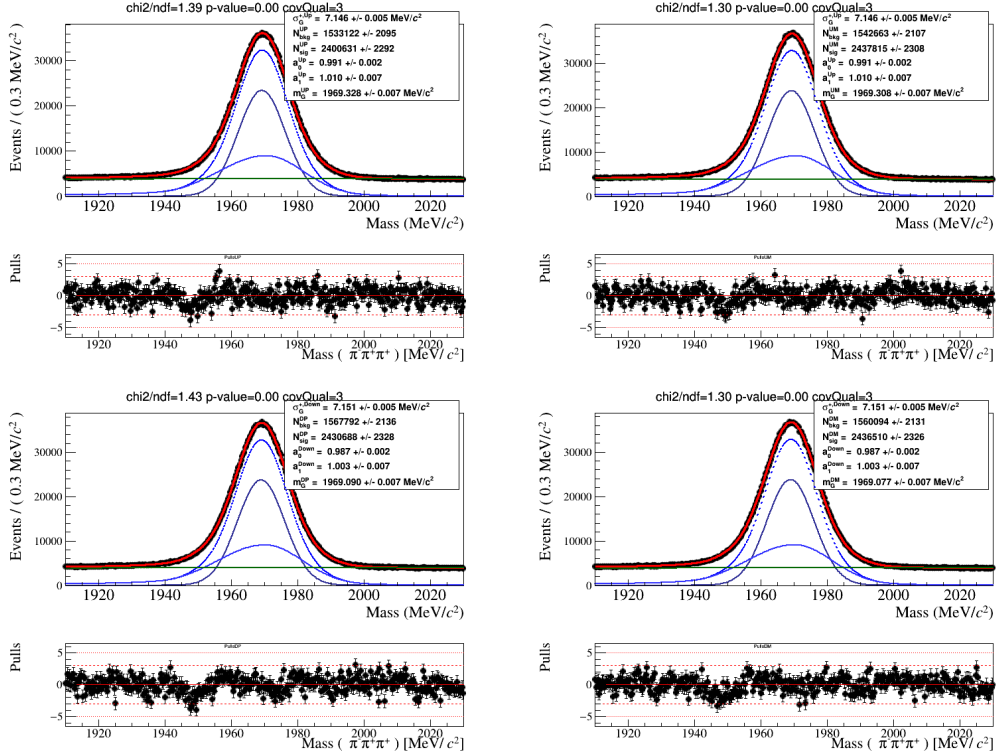
(b) Bin 19

Figure 5.8: Mass-per-bin fits of the four components of bins 18 and 19. For bin 18, an extra PDF (5.8a, narrow red line) was include to parameterise the specific background. The blue lines correspond to the signal components, and the green line to the combinatorial background. The red thick line corresponds to the summed PDF.

6
Results

With the MC samples prepared, as discussed in Section 5.1, and the signal yields obtained through the mass fits, as explained in Section 5.2, we are now able to proceed to the Mirandizing technique. We employ this to obtain the $\mathcal{S}_{CP}$ (Equation 5-11) in each bin of the $D_s^+ \to \pi^-\pi^+\pi^+$ phase space, to achieve the significance of the difference between the number of D_s^+ and D_s^- candidates.

In Figure 6.2, we can see the $\mathcal{S}_{CP}$ of the $D_s^+ \to \pi^-\pi^+\pi^+$ data sample, computed separately in each bin of the physics motivated binning presented in Figure 5.5b. For the RapidSim samples, the $\mathcal{S}_{CP}$ is shown in Figure 6.3. It is important to stress out that we chose not to use the errors from the mass fits yet, as we could not improve them enough at this point. The best results are presented using $\sqrt{N}$ as errors, where N is the number of the events in the sample. The $\mathcal{S}_{CP}$ results with the errors from the mass fits are shown in Appendix D.

Apart from that, we can notice that both $\mathcal{S}_{CP}$ results (for data and for simulation) have some statistical significance, since they present a p-value below 0.05. Looking at the value itself, most bins present a result that fails to reject the null hypothesis that there is no asymmetry, i.e., have a result under 3.

Some bins on the other hand, namely bin 17 for the data sample, got a result of 3.2, meaning that the asymmetry is statistically significant. It is important to notice though that this is exactly the region of the specific
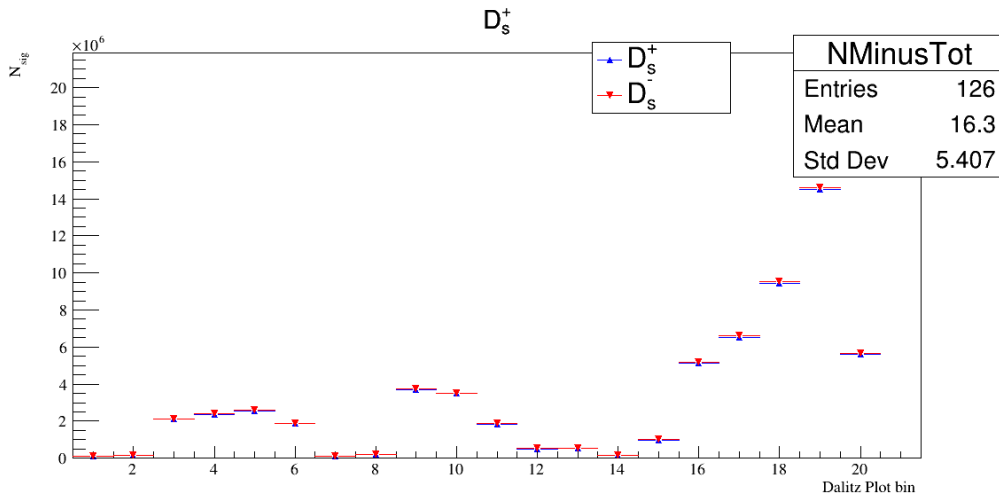
Figure 6.1: Counts for D_s^+ and D_s^- in each bin of the data sample
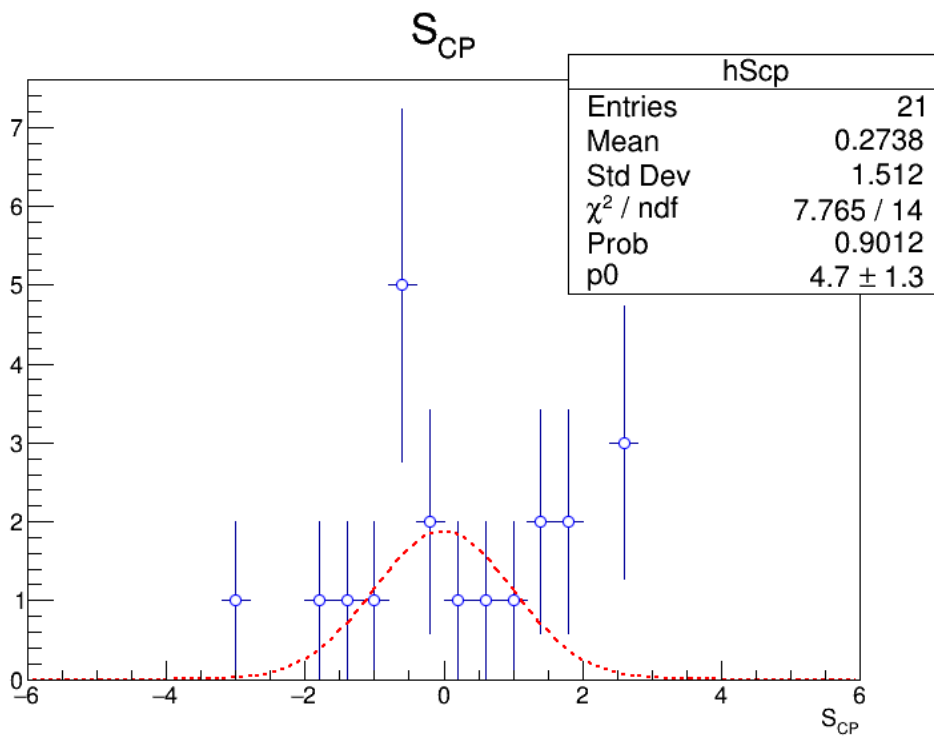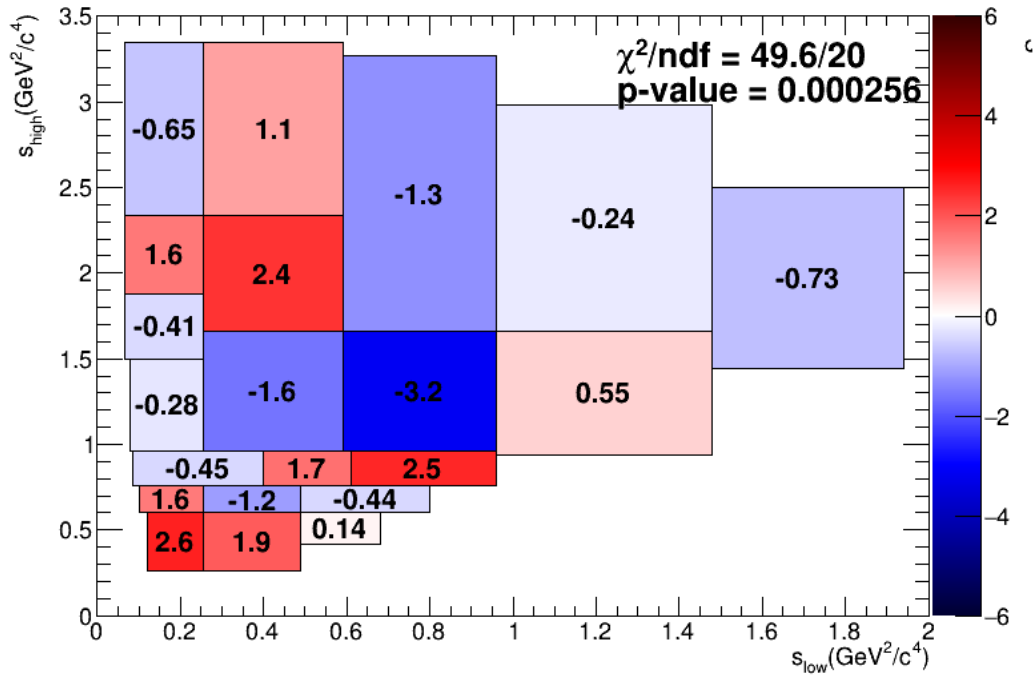
Figure 6.2: Observed asymmetry significance for the control channel $D_s^+ \rightarrow \pi^-\pi^+\pi^+$.
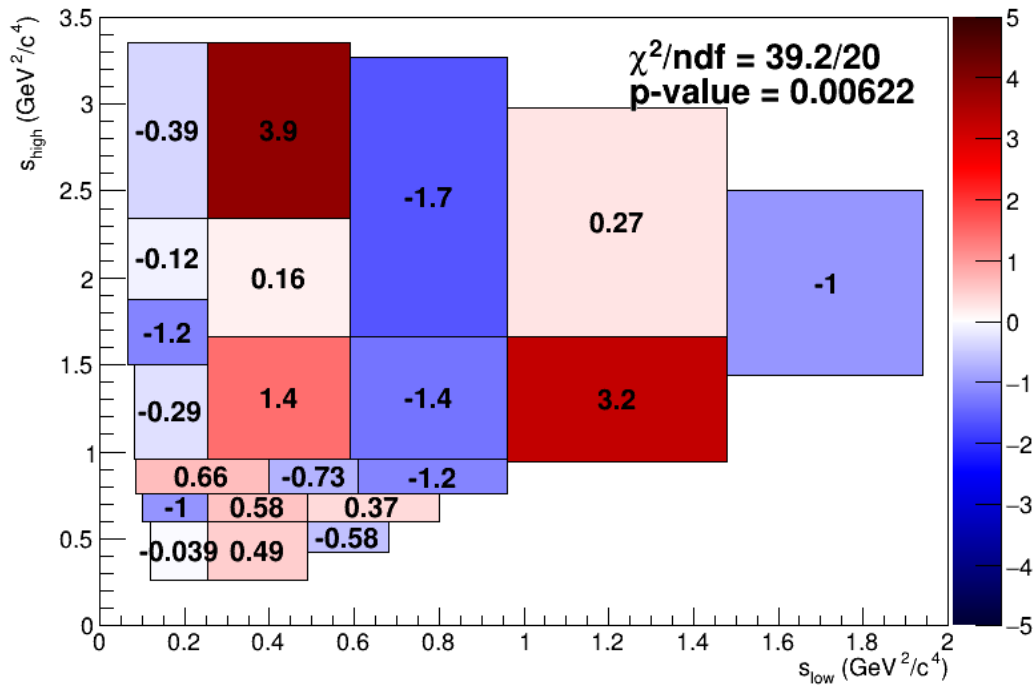
Figure 6.3: Observed asymmetry significance for the RapidSim sample of the control channel $D_s^+ \rightarrow \pi^- \pi^+ \pi^+$. The contributions of each asymmetry components are shown in Appendix D

background we discussed on section 5.2. The mass fit for this bin is shown in Figure 6.4. Although the fit result seems to be reasonable, and with pulls distributions mostly bellow 3σ, alternative strategies can be tested in order to be more certain that the value of $\mathcal{S}_{CP}$ obtained is correct.

When it comes to the $\mathcal{S}_{CP}$ results for the RapidSim sample, we can notice that bins 8 and 20 show a statistically significant result. For bin 8, by observing Figure 5.5b, we can notice that we have very few events, and also there is a substantial background component, with a small number of signal yields. Due to this, we must be careful when getting conclusions, as the sample we have is somehow sparse. In bin 20 on the other hand this does not happen. But alternative asymmetry models were tested, as shown and discussed in Appendix D, and in all of them, the $\mathcal{S}_{CP}$ obtained for this bin was compatible with the null hypothesis (as well as for bin 8).

It is important no notice though that not all of those results are final, as the analysis will continue from where this work stopped.
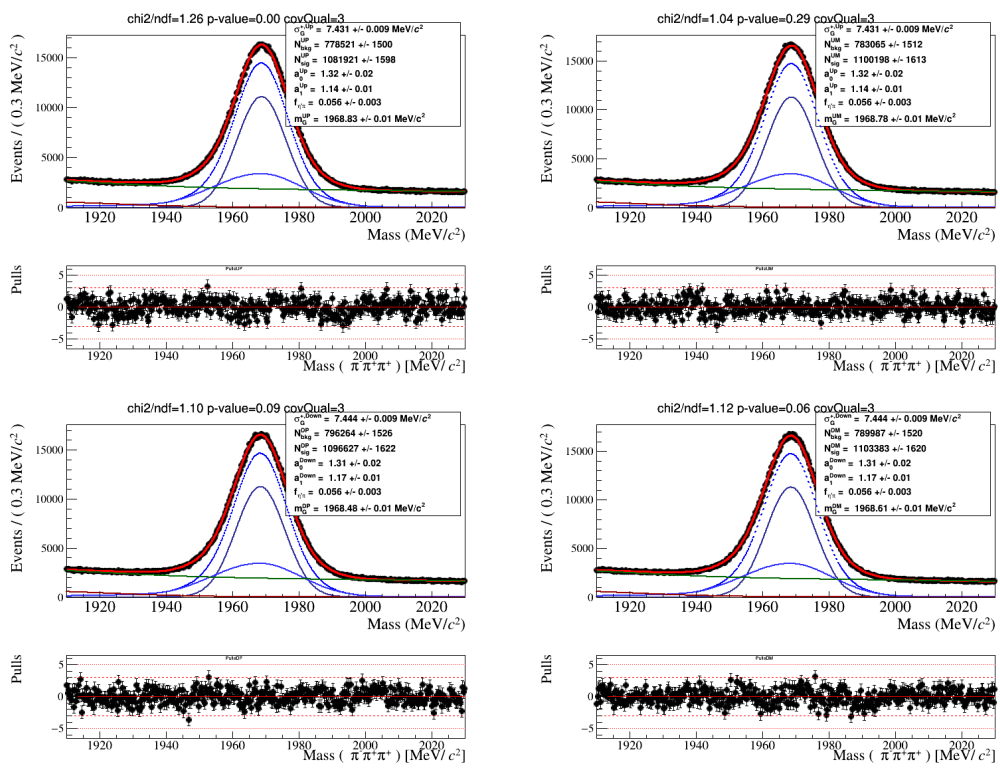
Figure 6.4: Mass fit for bin 17

7
Conclusion and future work

Throughout this work, we aimed at deeply understanding the processes and mechanisms that could cause instrumental and production charge asymmetries on a control channel, in order to give continuity to the study of the signal channel. As no CP effects are expected on the $D_s^+ \to \pi^- \pi^+ \pi^+$ decay by the SM, it is a well-suited choice to be used as control channel of the $D^+ \to \pi^- \pi^+ \pi^+$ decay. It is assumed that all non-CP related effects of charge asymmetry are the same for both processes, so an asymmetry on the signal channel that is not explained by what we observe on the control channel might be an evidence of CP violation.

A small value for the $\mathcal{S}_{CP}$, i.e., the significance of the difference between positive and negatively charged candidates, would mean that these instrumental and production charge asymmetries do not produce a statistically significant imbalance between positive and negative particles. And this being true for the control channel, it would be expected to be true for the signal channel as well. On the other hand, observing the presence of an actually significant asymmetry caused by those factors in a specific region of the phase space, would mean similarly that the same thing should happen on the signal channel.

However, based on the $\mathcal{S}_{CP}$ results, most regions of the DP proved to have non-significant charge asymmetries, meaning that they are most probably caused just by statistical fluctuations. Nonetheless, different approaches to the asymmetry injection lead to different conclusions for the $\mathcal{S}_{CP}$ in some DP bins. This means that to be more sure of what are the effects of those asymmetries across the phase space, some fine-tuning should be done in a few steps of the study.

Improvements on the specific background treatment for example could lead to an even better result on the mass-fit, giving a more realistic value for the signal yields and its associated error. Different line-shapes can be tested, in order to find one that better describes the $D_s^+ \to \eta' \pi^+$ distributions. In addition, multiple approaches to the fitting procedure can be experimented not only on the specific background but also on the combinatorial background.

Despite some minor changes that still need to be made, the findings presented here give a very encouraging indication of our understanding of the sources of charge asymmetry. Ensuring that we have a well-rounded knowledge of the non-CP asymmetries of the $D^+ \to \pi^- \pi^+ \pi^+$ channel is a vital step in our journey to observe CP asymmetries in this decay.

Bibliography

[1] P. Dirac. "The quantum theory of the electron." In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 117.778 (1928), pp. 610–624. DOI: `10.1098/rspa.1928.0023`.

[2] C. D. Anderson. "THE APPARENT EXISTENCE OF EASILY DEFLECTABLE POSITIVES." In: *Science* 76.1967 (1932), pp. 238–239. ISSN: 0036-8075. DOI: `10.1126/science.76.1967.238`. eprint: `https://science.sciencemag.org/content/76/1967/238.full.pdf`. URL: `https://science.sciencemag.org/content/76/1967/238`.

[3] M. Thomson. *Modern Particle Physics.* Modern Particle Physics. Cambridge University Press, 2013. ISBN: 9781107034266.

[4] A. D. Sakharov. "Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe." In: *Pisma Zh. Eksp. Teor. Fiz.* 5 (1967), pp. 32–35. DOI: `10.1070/PU1991v034n05ABEH002497`.

[5] I. Bigi and A. Sanda. *CP Violation.* Cambridge monographs on particle physics, nuclear physics, and cosmology. Cambridge University Press, 2009. ISBN: 9780521847940.

[6] D. Dominguez. *The standard model.* URL: `https://home.cern/science/physics/standard-model`.

[7] T. D. Lee and C. N. Yang. "Question of Parity Conservation in Weak Interactions." In: *Phys. Rev.* 104 (1 Oct. 1956), pp. 254–258. DOI: `10.1103/PhysRev.104.254`.

[8] C. S. Wu et al. "Experimental Test of Parity Conservation in Beta Decay." In: *Phys. Rev.* 105 (4 Feb. 1957), pp. 1413–1415. DOI: `10.1103/PhysRev.105.1413`.

[9] B. Martin and G. Shaw. *Particle Physics.* Manchester Physics Series. Wiley, 2013. ISBN: 9781118681664.

[10] D. Griffiths. *Introduction to Elementary Particles.* Physics textbook. Wiley, 2008. ISBN: 9783527618477.

[11] J. H. Christenson et al. "Evidence for the 2π Decay of the K_2^0 Meson." In: *Phys. Rev. Lett.* 13 (4 July 1964), pp. 138–140. DOI: `10.1103/PhysRevLett.13.138`.

[12] B. Aubert et al. "Directmml:math xmlns:mml="http://www.w3.org/1998/Math/MathML" display="inline"mml:miC/mml:mimml:miP/mml:mi/mml:mathViolating Asymmetry inmml:math xmlns:mml="http://www.w3.org/1998/Math/MathML" display="inline"mml:msupmml:miB/mml:mimml:mn0/mml:mn/mml:msupmml:mo→/mml:mo /mml:mo/mml:msup/mml:mathDecays." In: *Physical Review Letters* 93.13 (Sept. 2004). DOI: 10.1103/physrevlett.93.131801. URL: https://doi.org/10.1103%2Fphysrevlett.93.131801.

[13] M. B. GAVELA et al. "STANDARD MODEL CP-VIOLATION AND BARYON ASYMMETRY." In: *Modern Physics Letters A* 09.09 (Mar. 1994), pp. 795–809. DOI: 10.1142/s0217732394000629. URL: https://doi.org/10.1142%2Fs0217732394000629.

[14] A. Alavi-Harati et al. "Observation of Direct CP Violation in $K_{S,L} \xrightarrow{\pi} \pi$ Decays." In: *Physical Review Letters* 83.1 (July 1999), pp. 22–27. ISSN: 1079-7114. DOI: 10.1103/physrevlett.83.22.

[15] V. Fanti et al. "A new measurement of direct CP violation in two pion decays of the neutral kaon." In: *Physics Letters B* 465.1-4 (Oct. 1999), pp. 335–348. ISSN: 0370-2693. DOI: 10.1016/s0370-2693(99)01030-8.

[16] K. Abe et al. "Observation of LargeCPViolation in the NeutralBMeson System." In: *Physical Review Letters* 87.9 (Aug. 2001). ISSN: 1079-7114. DOI: 10.1103/physrevlett.87.091802.

[17] B. Aubert et al. "Measurement of CP-Violating Asymmetries in $B_0 DecaystoCPEigenstates$." In: *Physical Review Letters* 86.12 (Mar. 2001), pp. 2515–2522. ISSN: 1079-7114. DOI: 10.1103/physrevlett.86.2515.

[18] R. Aaij et al. "First Observation of CP Violation in the Decays of B_s^0 Mesons." In: *Phys. Rev. Lett.* 110.22, 221601 (May 2013), p. 221601. DOI: 10.1103/PhysRevLett.110.221601. arXiv: 1304.6173 [hep-ex].

[19] R. Aaij et al. "Observation of C P Violation in Charm Decays." In: *Phys. Rev. Lett.* 122.21, 211803 (May 2019), p. 211803. DOI: 10.1103/PhysRevLett.122.211803.

[20] N. Cabibbo. "Unitary Symmetry and Leptonic Decays." In: *Phys. Rev. Lett.* 10 (12 June 1963), pp. 531–533. DOI: 10.1103/PhysRevLett.10.531.

[21] S. L. Glashow, J. Iliopoulos, and L. Maiani. "Weak Interactions with Lepton-Hadron Symmetry." In: *Phys. Rev. D* 2 (7 Oct. 1970), pp. 1285–1292. DOI: 10.1103/PhysRevD.2.1285.

[22] P. D. Group et al. "Review of Particle Physics." In: *Progress of Theoretical and Experimental Physics* 2022.8 (Aug. 2022). 083C01. ISSN: 2050-3911. DOI: `10.1093/ptep/ptac097`. eprint: `https://academic.oup.com/ptep/article-pdf/2022/8/083C01/45434166/ptac097.pdf`. URL: `https://doi.org/10.1093/ptep/ptac097`.

[23] M. S. Sozzi. *Discrete symmetries and CP violation: from experiment to theory.* Oxford graduate texts. New York, NY: Oxford Univ. Press, 2008. DOI: `10.1093/acprof:oso/9780199296668.001.0001`.

[24] L. Wolfenstein. "Parametrization of the Kobayashi-Maskawa Matrix." In: *Phys. Rev. Lett.* 51 (21 Nov. 1983), pp. 1945–1947. DOI: `10.1103/PhysRevLett.51.1945`.

[25] A. Höcker and Z. Ligeti. "CP Violation and the CKM Matrix." In: *Annual Review of Nuclear and Particle Science* 56.1 (Nov. 2006), pp. 501–567. ISSN: 1545-4134. DOI: `10.1146/annurev.nucl.56.080805.140456`.

[26] C. Jarlskog. "Commutator of the Quark Mass Matrices in the Standard Electroweak Model and a Measure of Maximal CP Violation." In: *Phys. Rev. Lett.* 55 (1985), p. 1039. DOI: `10.1103/PhysRevLett.55.1039`.

[27] R. Dalitz. "CXII. On the analysis of -meson data and the nature of the -meson." In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 44.357 (1953), pp. 1068–1080. DOI: `10.1080/14786441008520365`. eprint: `https://doi.org/10.1080/14786441008520365`. URL: `https://doi.org/10.1080/14786441008520365`.

[28] A. Bettini. *Introduction to elementary particle physics; 1st ed.* Cambridge: Cambridge Univ. Press, 2008. DOI: `10.1017/CBO9780511809019`. URL: `https://cds.cern.ch/record/1111396`.

[29] A. Wuethrich. *Dalitz Plots and Hadron Spectroscopy.* 2005. DOI: `10.48550/ARXIV.HEP-PH/0507058`. URL: `https://arxiv.org/abs/hep-ph/0507058`.

[30] G. Bakker et al. "The reactions Kn → Kn and Kn → KoN at 3 GeV/c." In: *Nuclear Physics B* 16.1 (1970), pp. 53–60. ISSN: 0550-3213. DOI: `https://doi.org/10.1016/0550-3213(70)90336-6`.

[31] LHCb Collaboration. *Amplitude analysis of the $D_s^+ \to^- {}^{++}$ decay.* 2022. DOI: `10.48550/ARXIV.2209.09840`. URL: `https://arxiv.org/abs/2209.09840`.

[32] L. Evans and P. Bryant. "LHC Machine." In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08001–S08001. DOI: 10.1088/1748-0221/3/08/s08001. URL: https://doi.org/10.1088/1748-0221/3/08/s08001.

[33] R. Amann, M. Lamont, and S. Myers. "A brief history of the LEP collider." In: *Nuclear Physics B - Proceedings Supplements* 109.2 (2002). Proceedings of the 7th Topical Seminar, pp. 17–31. ISSN: 0920-5632. DOI: https://doi.org/10.1016/S0920-5632(02)90005-8. URL: https://www.sciencedirect.com/science/article/pii/S0920563202900058.

[34] CERN. *The large hadron collider.* 2022. URL: https://home.cern/science/accelerators/large-hadron-collider (visited on 04/06/2022).

[35] F. Marcastel. "CERN's Accelerator Complex. La chaîne des accélérateurs du CERN." In: (Oct. 2013). General Photo. URL: https://cds.cern.ch/record/1621583.

[36] CERN. *CERN's accelerator complex.* 2022. URL: https://home.cern/science/accelerators/accelerator-complex (visited on 04/11/2022).

[37] CERN. *LHCB.* 2022. URL: https://home.cern/science/experiments/lhcb (visited on 04/11/2022).

[38] R. Lindner. "LHCb layout$_2$.*LHCbschema*$_2$." LHCb Collection. Feb. 2008. URL: http://cds.cern.ch/record/1087860.

[39] A. A. Alves et al. "The LHCb Detector at the LHC." In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08005–S08005. DOI: 10.1088/1748-0221/3/08/s08005. URL: https://doi.org/10.1088/1748-0221/3/08/s08005.

[40] E. Buchanan. "The LHCb Vertex Locator (VELO) Pixel Detector Upgrade." In: *Journal of Instrumentation* 12 (2017).

[41] LHCb. *VELO.* 2022. URL: https://lhcb-outreach.web.cern.ch/detector/vertex-locator-velo/ (visited on 04/12/2022).

[42] "LHCb detector performance." In: *International Journal of Modern Physics A* 30.07 (Mar. 2015), p. 1530022. DOI: 10.1142/s0217751x15300227. URL: https://doi.org/10.1142%5C%2Fs0217751x15300227.

[43] Celladoor. *Daten, Fakten und Funktionen des Neuen Teilchenbeschleunigers von Cern Bei Genf.* URL: https://www.lhc-facts.ch/index.php?page=lhcb.

[44] CERN. *https://lhcb-outreach.web.cern.ch/detector/calorimeters/*. 2022. URL: `https : / / lhcb - outreach . web . cern . ch / detector / calorimeters/` (visited on 04/06/2022).

[45] R. Nobrega et al. "LHCb Trigger System : Technical Design Report." In: (Jan. 2003).

[46] R. A. *et al.* "Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC." In: *JINST* 44 (2019), P04013–P04013. DOI: `10.1088/1748-0221/14/04/p04013`.

[47] L. D. Packages. *CharmWGProd/productions/Run2_D2HHH*. URL: `https : / / gitlab . cern . ch / lhcb - datapkg / WG / CharmWGProd/ - /tree/master/productions/Run2_D2HHH`.

[48] *About PYTHIA*. URL: `https : / / pythia . org / about/` (visited on 04/06/2022).

[49] *Overview Geant4*. URL: `https://geant4.web.cern.ch/about/` (visited on 04/06/2022).

[50] A. Rogozhnikov. *Reweighting algorithms*. URL: `https://arogozhnikov. github.io/hep_ml/reweight.html`.

[51] R. L. Workman et al. "Review of Particle Physics." In: *PTEP* 2022 (2022), p. 083C01. DOI: `10.1093/ptep/ptac097`.

[52] I. Bediaga et al. "On a CP anisotropy measurement in the Dalitz plot." In: *Physical Review D* 80.9 (Nov. 2009). DOI: `10.1103/physrevd.80. 096006`. URL: `https://doi.org/10.1103%2Fphysrevd.80.096006`.

[53] L. Dufour and J. Van Tilburg. *Decomposition of simulated detection asymmetries in LHCb*. Tech. rep. Geneva: CERN, 2018. URL: `https: //cds.cern.ch/record/2304546`.

[54] *PIDCalib packages*. URL: `https://twiki.cern.ch/twiki/bin/view/ LHCb/PIDCalibPackage`.

[55] R. Aaij et al. "Measurement of D± s production asymmetry in pp collisions at \sqrt{s}=7 and 8 TeV." In: *Journal of High Energy Physics* 2018.8 (Aug. 2018). DOI: `10 . 1007 / jhep08(2018 ) 008`. URL: `https: //doi.org/10.1007%2Fjhep08%282018%29008`.

[56] G. Cowan, D. Craik, and M. Needham. "RapidSim: An application for the fast simulation of heavy-quark hadron decays." In: *Computer Physics Communications* 214 (May 2017), pp. 239–246. DOI: `10.1016/j.cpc. 2017.01.029`. URL: `https://doi.org/10.1016%2Fj.cpc.2017.01. 029`.

[57]   M. Pivk and F. L. Diberder. ": A statistical tool to unfold data distributions." In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 555.1-2 (Dec. 2005), pp. 356–369. DOI: `10.1016/j.nima.2005.08.106`. URL: `https://doi.org/10.1016%2Fj.nima.2005.08.106`.

[58]   M. Oreglia. "A Study of the Reactions $\psi' \to \gamma\gamma\psi$." Other thesis. Dec. 1980.

[59]   L. Landau. "On the conservation laws for weak interactions." In: *Nuclear Physics* 3.1 (1957), pp. 127–131. ISSN: 0029-5582. DOI: `https://doi.org/10.1016/0029-5582(57)90061-5`.

[60]   Y. Nagashima and Y. Nambu. *Elementary Particle Physics: Quantum Field Theory and Particles V1.* v. 1. Wiley, 2011. ISBN: 9783527643967.

[61]   P. Pal. *An Introductory Course of Particle Physics.* Taylor & Francis, 2014. ISBN: 9781482216981.

[62]   C. Pruneau. *Data Analysis Techniques for Physical Scientists.* Cambridge University Press, 2017. ISBN: 9781108267885. URL: `https://books.google.com.br/books?id=Yj41DwAAQBAJ`.

[63]   C. A. Meyer. *General Properties Of Three-body Decays.* June 2017. URL: `https://halldweb.jlab.org/doc-public/DocDB/ShowDocument?docid=3345`.

[64]   M. D. Needham. *Momentum scale calibration of the LHCb spectrometer.* Tech. rep. Geneva: CERN, Apr. 2017. URL: `https://cds.cern.ch/record/2260848`.

A
Appendix: Selection cuts

In this appendix we present the cuts applied in the data and MC data samples used in the analysis.

The pre-selection of events, performed by central production, is displayed in Table A.4.

Table A.1: Cuts applied at the pre-selection (central production of ntuples).

candidate	variable	cut
D_s^+	χ^2 of Impact Parameter wrt PV $<$	15
	Mass $[\,\mathrm{MeV}/c^2\,]$	$1905 - 2035$
	ProbNN_pi $>$	0.3

The criteria used in the HLT2 are shown in Table A.2.

Table A.2: HLT2 selection criteria.

Daughter cuts	
Track χ^2/ndf	< 3.0
$p_t[\,\mathrm{MeV}/c\,]$	> 250
χ^2 of Impact Parameter wrt PV	> 4.0
$\Delta \log \mathcal{L}_{K\pi}$	< 1.0
Combination cuts	
Mass $[\,\mathrm{MeV}/c^2\,]$	$1879 - 2059$
$\sum p_t \,[\,\mathrm{MeV}/c\,]$	> 3200
p_t at least one track $[\,\mathrm{MeV}/c\,]$	> 1000
p_t at least two tracks $[\,\mathrm{MeV}/c\,]$	> 400
χ^2 of Impact Parameter	> 50 (at least one tracks)
χ^2 of Impact Parameter	> 10 (at least two tracks)
Mother cuts	
Track vertex χ^2/DOF	< 6.0
Mass $[\,\mathrm{MeV}/c^2\,]$	$1889 - 2049$
lifetime [ps]	> 0.2
acos(DIRA) [mrad]	< 10
TisTosSpec	Hlt1.*Track.*Decision%TOS

In Table A.3 we show the requirements on one and two tracks before the training, and the variables used as input in the neural network.

The central production selection criteria, which include the momentum scale calibration tool[64], can be found in Table A.4.

Table A.3: HLT1 selection criteria.

Requirements used in Hlt1TrackMVA	
Track preselections	
Track χ^2/ndf	< 3.0
$p_T [\mathrm{MeV}/c]$	> 500
IP χ^2 wrt PV	> 4
Variables for NN	
p_t, χ^2 of Impact Parameter	
Requirements used in Hlt1TwoTrackMVA	
Track preselections	
Track χ^2/ndf	< 2.5
$p_t [\mathrm{MeV}/c]$	> 500
IP χ^2 wrt PV	> 4
TwoTrack preselections	
$p_t [\mathrm{GeV}/c]$	> 2
vertex χ^2/DOF	< 10
mcor $[\mathrm{GeV}]$	> 1
η	$2 - 5$
Variables for NN	
$\sum p_t$, vertex χ^2/DOF, FD χ^2, N(tracks with IP $\chi^2 < 16$)	

Table A.4: Cuts applied at the pre-selection (central production).

variable	cut
χ^2 of Impact Parameter wrt PV $<$	15
Mass $[\mathrm{MeV}/c^2]$	$1905 - 2035$
ProbNN_pi $>$	0.3

Events are generated based on a resonant model, and generator level cuts are applied, as summarised in Table A.5, as well as the trigger filter.

Generator level cuts of the MC samples for the background channel ($D_s^+ \to \eta'\pi^+$) are presented in Table A.6.

In the specific background analysis, a potential source of background is the $D^+ \to K^-\pi^+\pi^+$ feed-through, where the opposite charge daughter is misidentified as a pion in the $D_s^+ \to \pi^-\pi^+\pi^+$ decay. To mitigate this contamination, the performance of two available PID variables was evaluated. The analysis reveals that ProbNNk is more effective than PIDK in rejecting this type of background. To ensure charge symmetry, the same PID cut is applied to all three tracks. Therefore, a ProbNNk < 0.2 cut is imposed on the pions from $D_s^+ \to \pi^-\pi^+\pi^+$, which significantly reduces the $D^+ \to K^-\pi^+\pi^+$ background. This requirement results in the removal of approximately 2% of the D_s^+ signal candidates. The specifics for

Table A.5: Cuts applied at generator level for signal channel.

variable	cut
each daughter $P >$	2.0 GeV/c
each daughter $P_T >$	0.25 GeV/c
$D_s^+ P >$	14.0 GeV/c
$D_s^+ P_T >$	2.5 GeV/c
D_s^+ $c\tau >$	60 μm
Required not to come from B	

Table A.6: Cuts applied at generator level for $D_s^+ \to \eta'\pi^+$ background samples.

variable	cut
each daughter $P >$	0.6 GeV/c
each daughter $P_T >$	0.2 GeV/c
$D_s^+ P_T >$	1.0 GeV/c

this analysis and a summary of the requirements can be found in Table A.7.

L0 Trigger	
`D_L0_TIS`	Hadron, Electron, Photon, Muon, Dimuon
Fiducial	
$p_z > 3.57 \times |p_x| + 1$ GeV/c (all tracks)	
$p_z > 4.25 \times |p_y|$ (all tracks)	
$1.5 \leq \eta \leq 5.0$ (all tracks)	
$3 \leq p \leq 100$ GeV/c (all tracks)	
nSPDhits < 1000	
Cloned Tracks	
$dTX_{23} > 6 \times 10^{-5}$	
$dTY_{23} > 6 \times 10^{-5}$	
Specific Background Contributions	
isMuon	$= \mathbf{0}$ (all tracks)
ProbNNpi	$> \mathbf{0.3}$ (all tracks)
ProbNNk	$< \mathbf{0.2}$ (tracks of D^+ candidates)
D_IPCHI2	$< \mathbf{12}$
$m_{\pi^-\pi^+}(\text{low}) - 0.496$ GeV/c^2	$> \mathbf{0.012}$ GeV/c^2 (D^+)
$m_{\pi^-\pi^+}(\text{high})$	$< \mathbf{1.84}$ GeV/c^2 (D_s^+)

Table A.7: Summary of cuts applied in the offline selection.

B
Appendix: Reweighting results

The comparison between the distributions of the reweighted variables of Data and MC, before and after reweighting, as well as its associated ROC curve are shown bellow. The details of the procedure are discussed in Chapter 5.
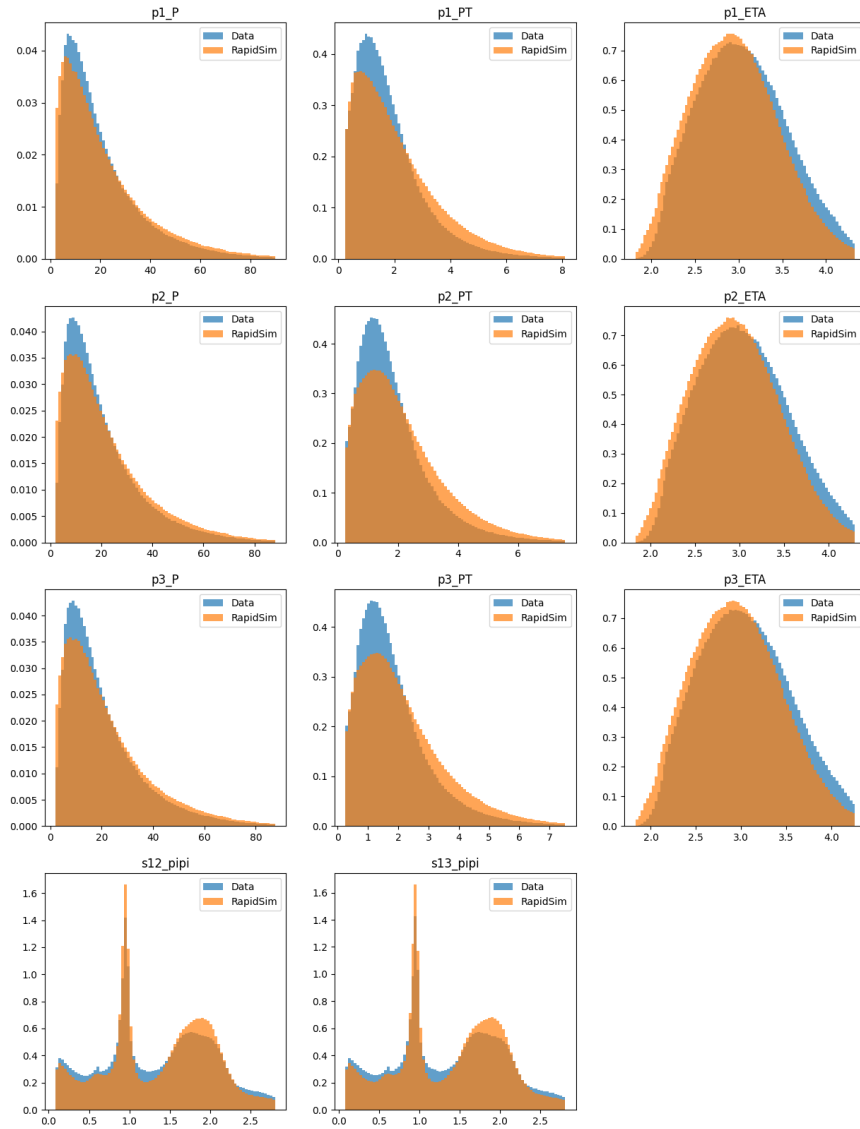
Figure B.1: Comparison between the reweighted variables distributions of the Data and MC 2016 MagUp sample, before the reweighting procedure. All variables are shown in GeV.
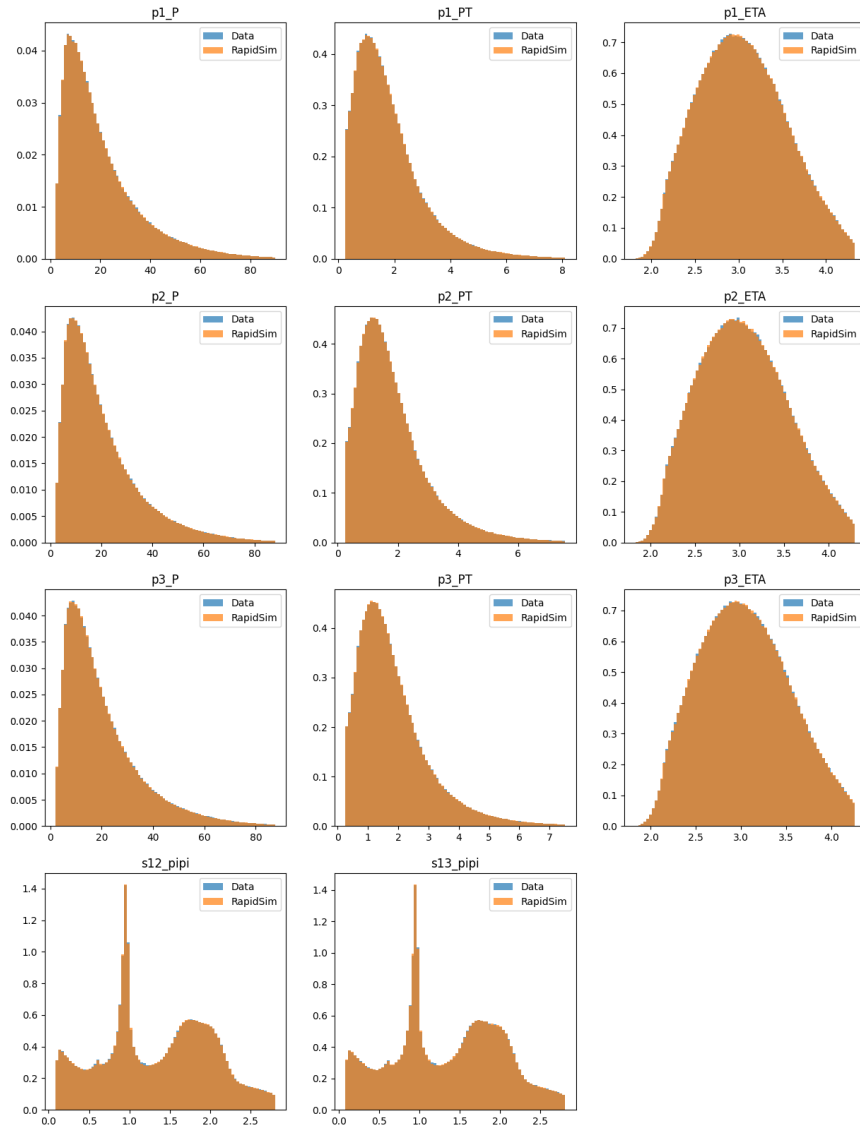
Figure B.2: Comparison between the reweighted variables distributions of the Data and MC 2016 MagUp sample, after the reweighting procedure. All variables are shown in GeV.
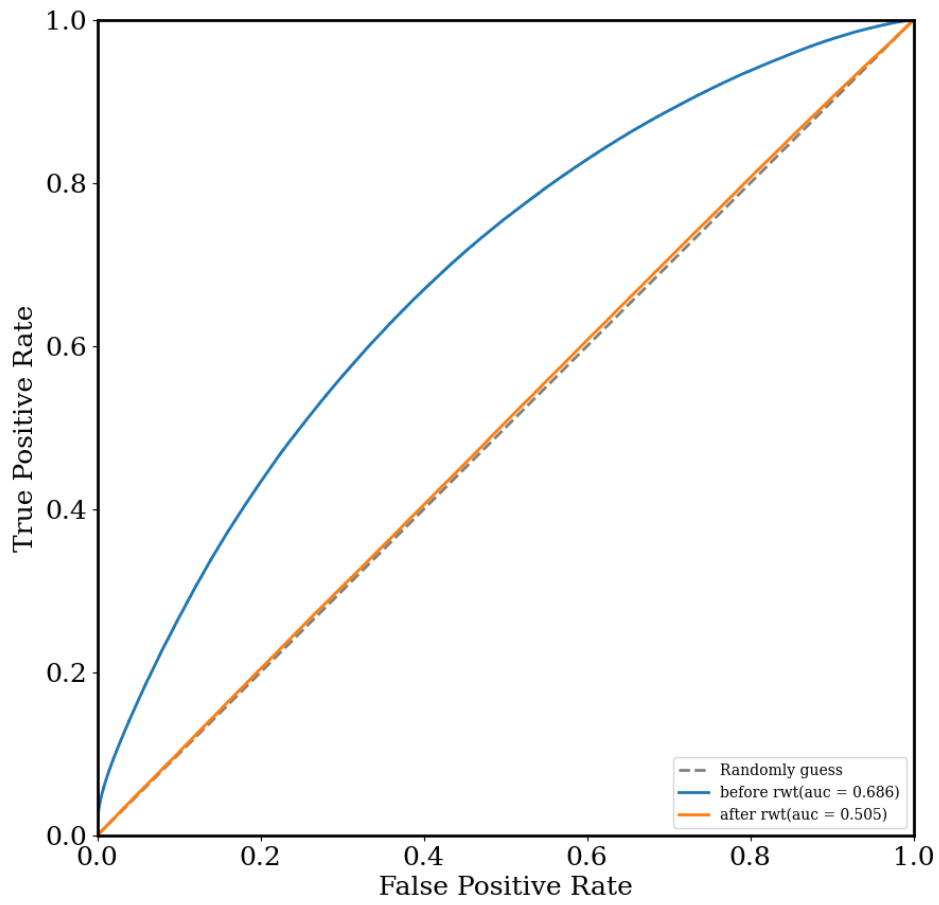
Figure B.3: ROC curve of the 2016 MagUp reweighted MC sample.
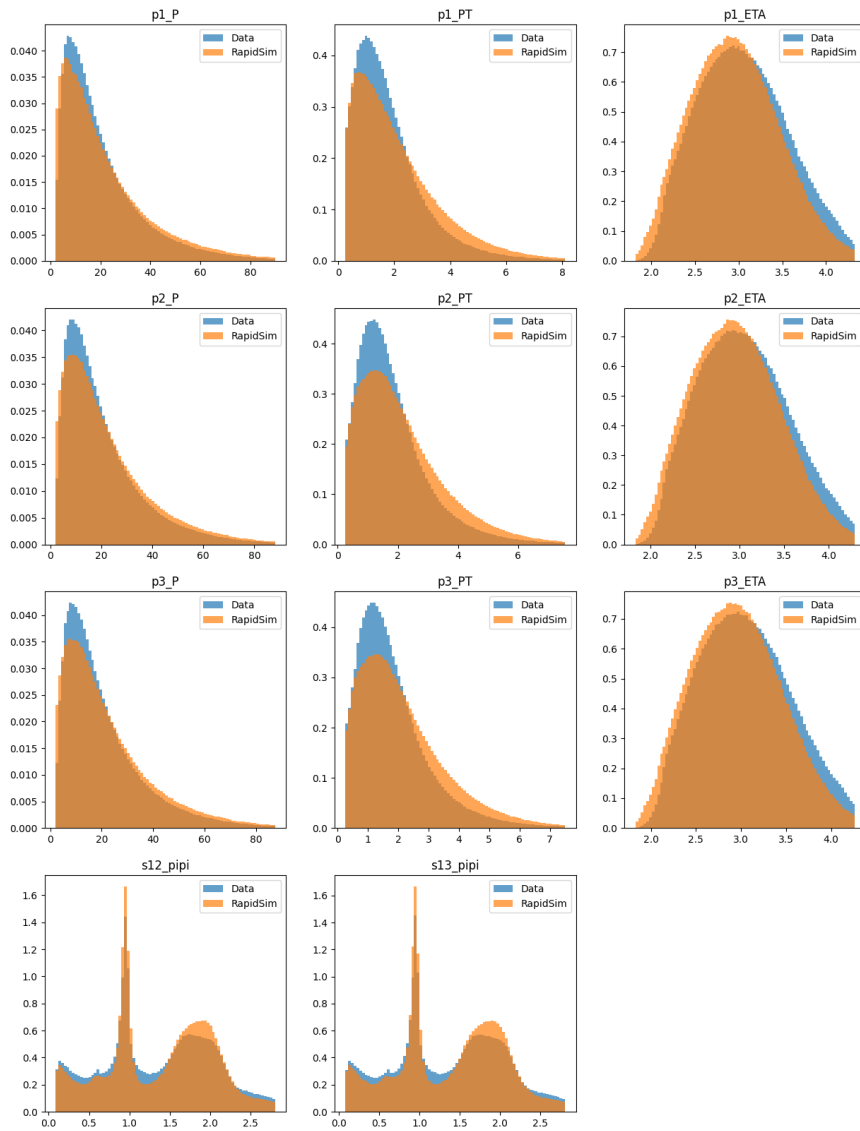
Figure B.4: Comparison between the reweighted variables distributions of the Data and MC 2017 MagDown sample, before the reweighting procedure. All variables are shown in GeV.
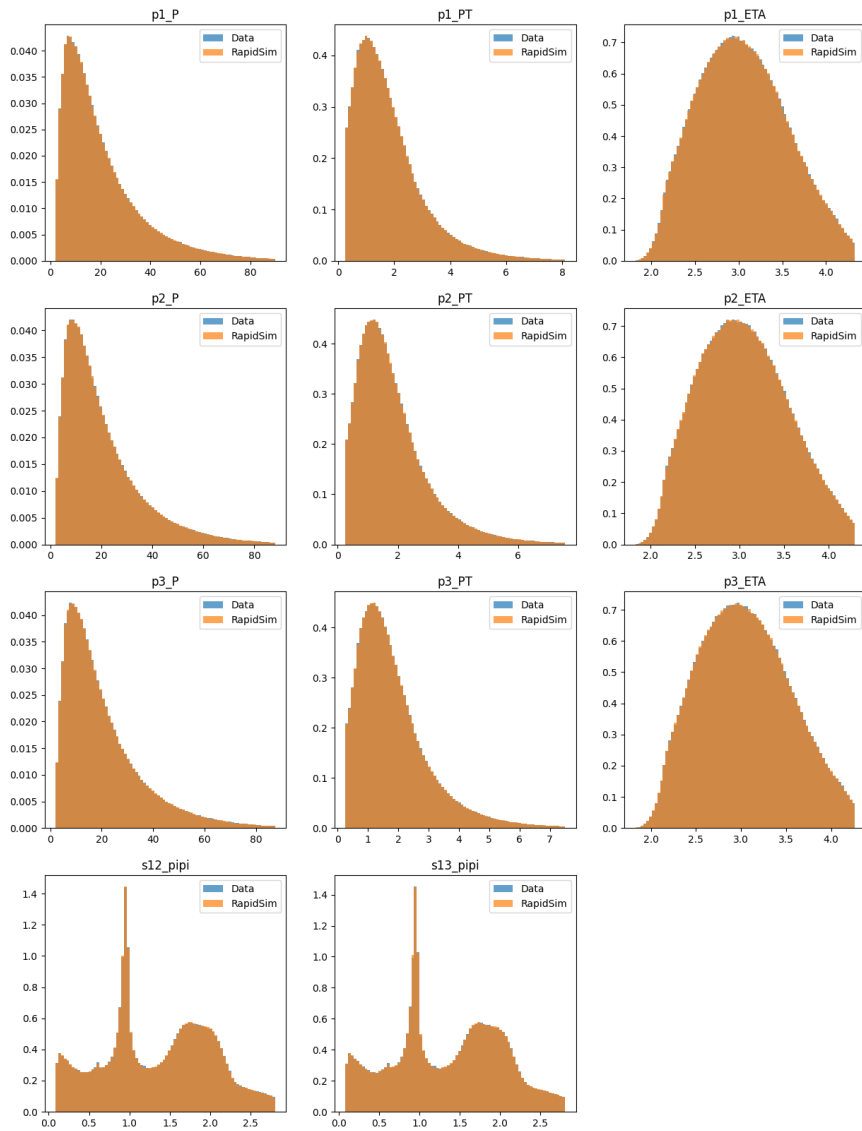
Figure B.5: Comparison between the reweighted variables distributions of the Data and MC 2017 MagDown sample, after the reweighting procedure. All variables are shown in GeV.
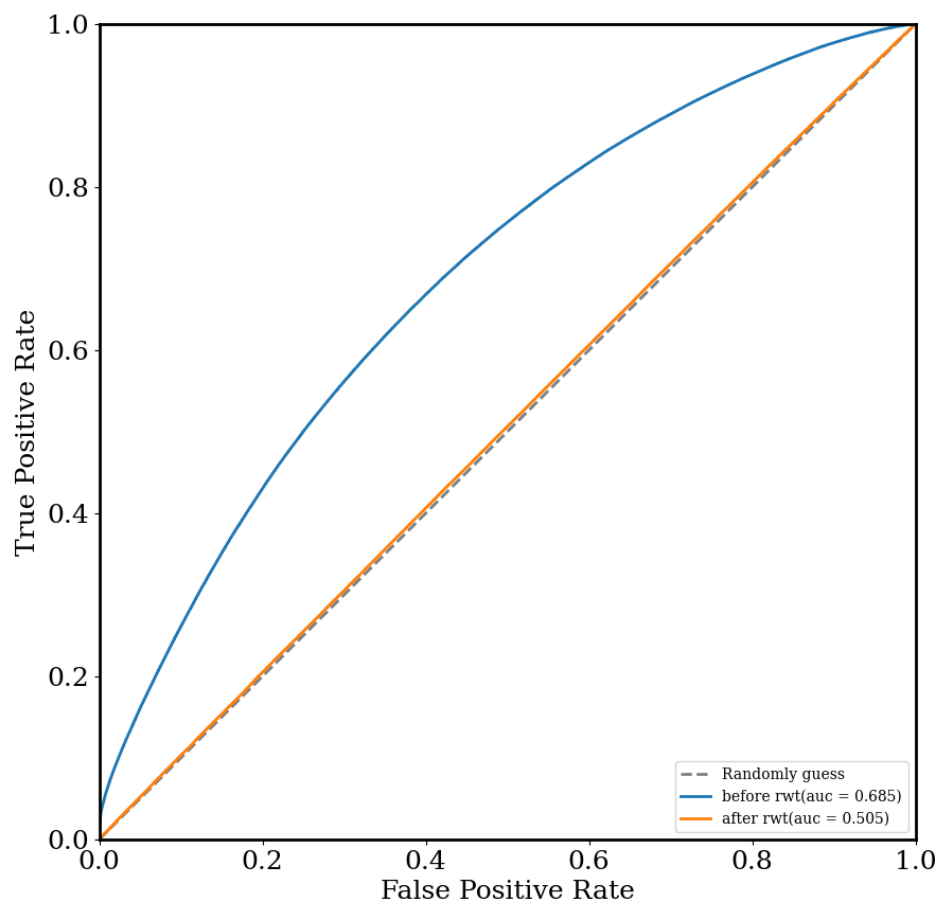
Figure B.6: ROC curve of the 2017 MagDown reweighted MC sample.
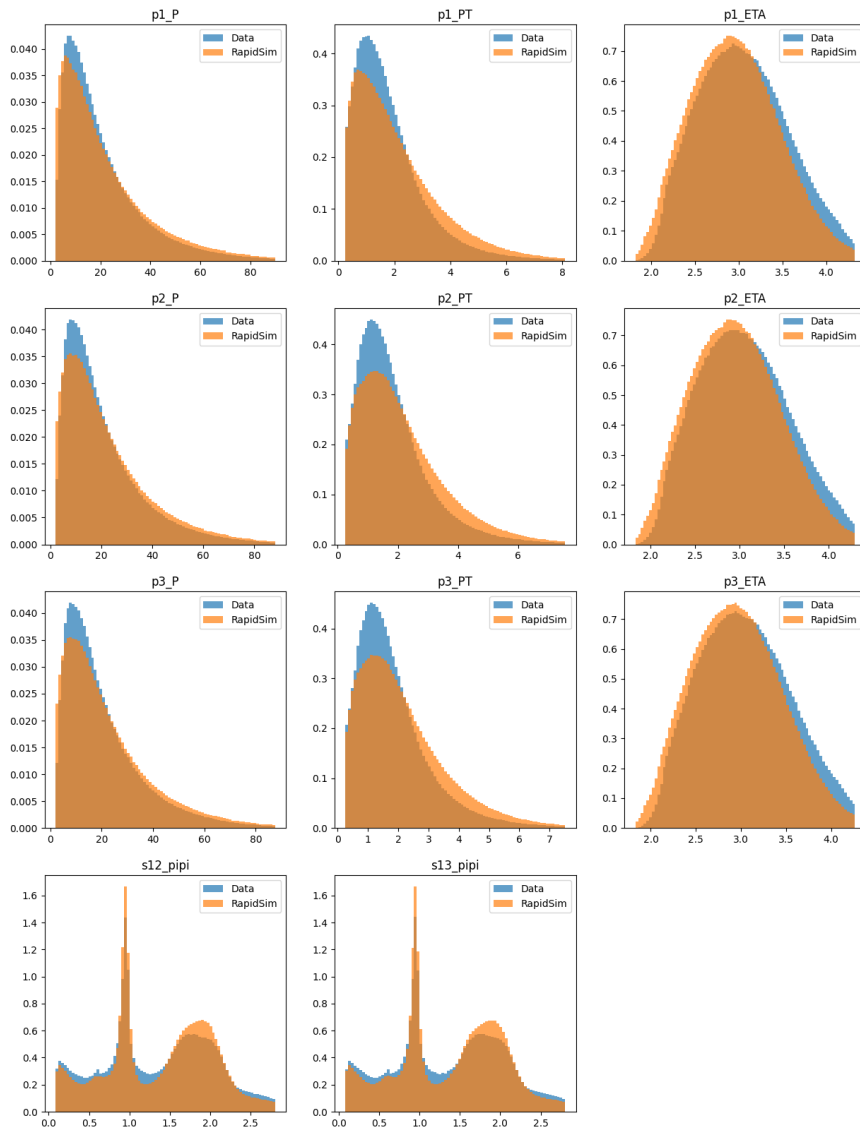
Figure B.7: Comparison between the reweighted variables distributions of the Data and MC 2017 MagUp sample, before the reweighting procedure. All variables are shown in GeV.
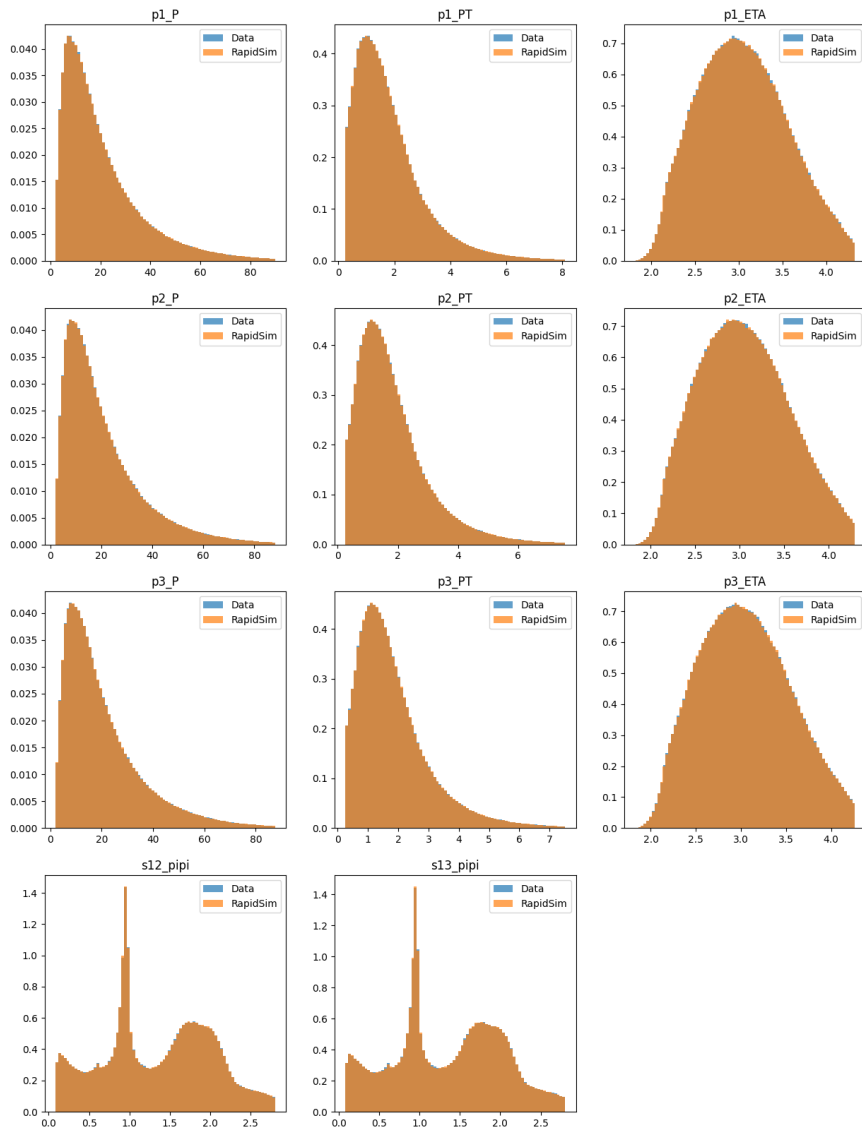
Figure B.8: Comparison between the reweighted variables distributions of the Data and MC 2017 MagUp sample, after the reweighting procedure. All variables are shown in GeV.
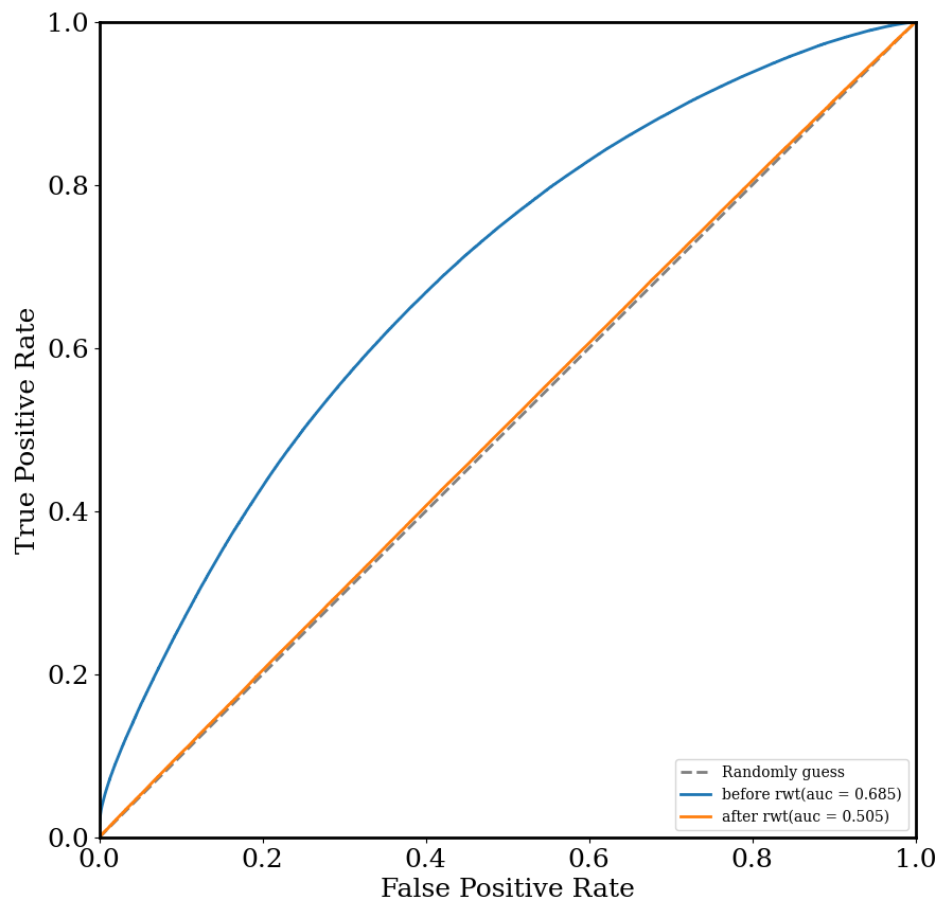
Figure B.9: ROC curve of the 2017 MagUp reweighted MC sample.
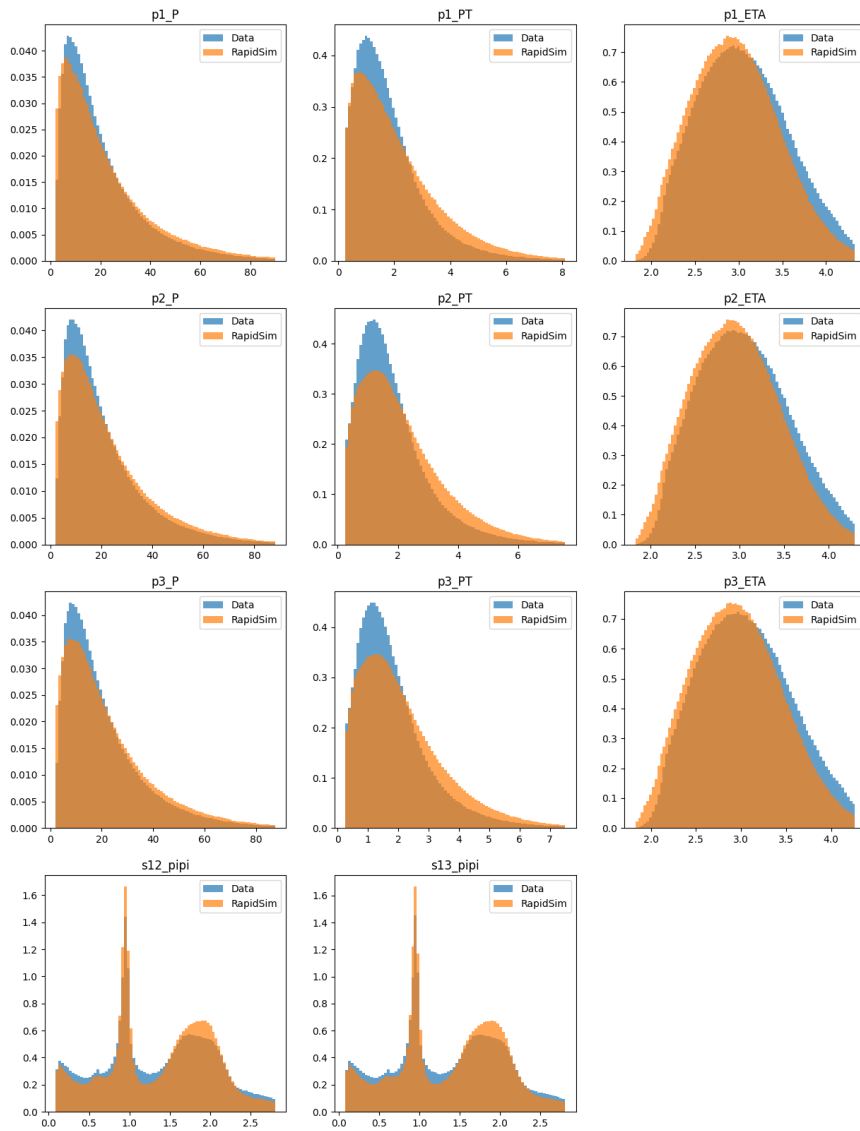
Figure B.10: Comparison between the reweighted variables distributions of the Data and MC 2017 MagDown sample, before the reweighting procedure. All variables are shown in GeV.
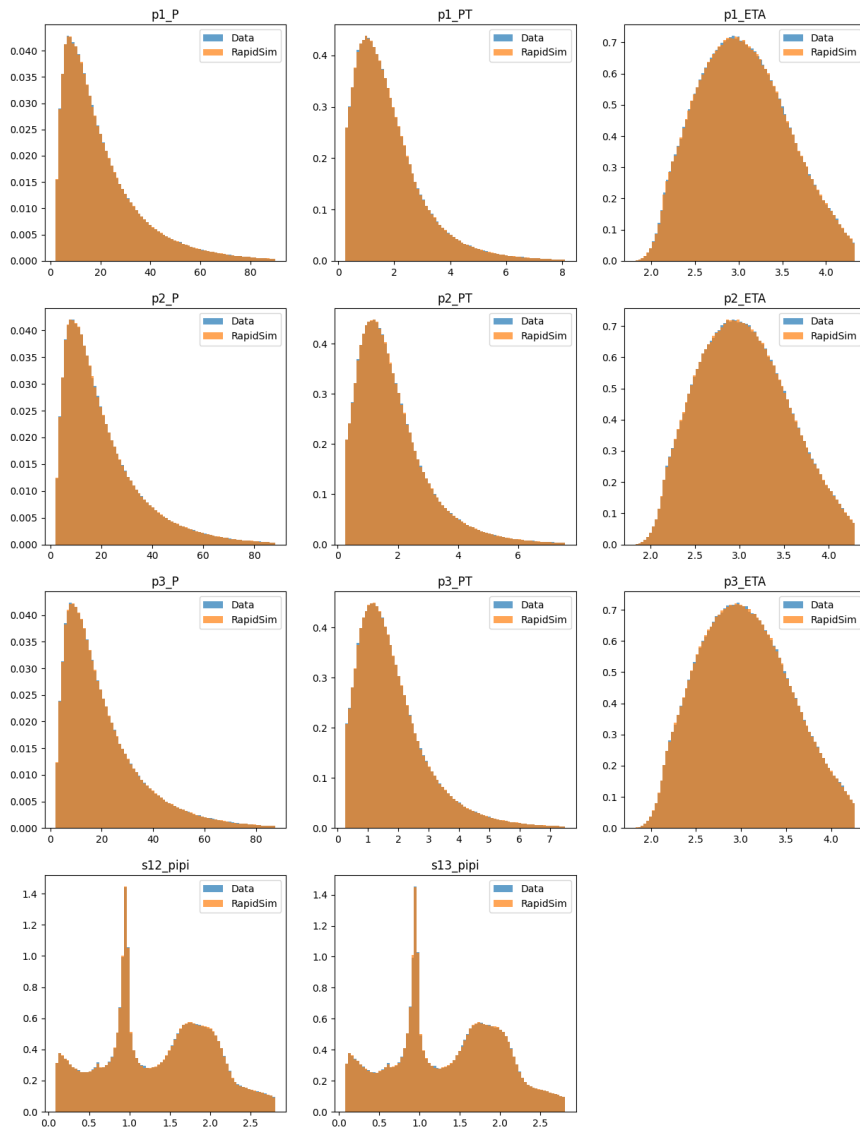
Figure B.11: Comparison between the reweighted variables distributions of the Data and MC 2017 MagDown sample, after the reweighting procedure. All variables are shown in GeV.
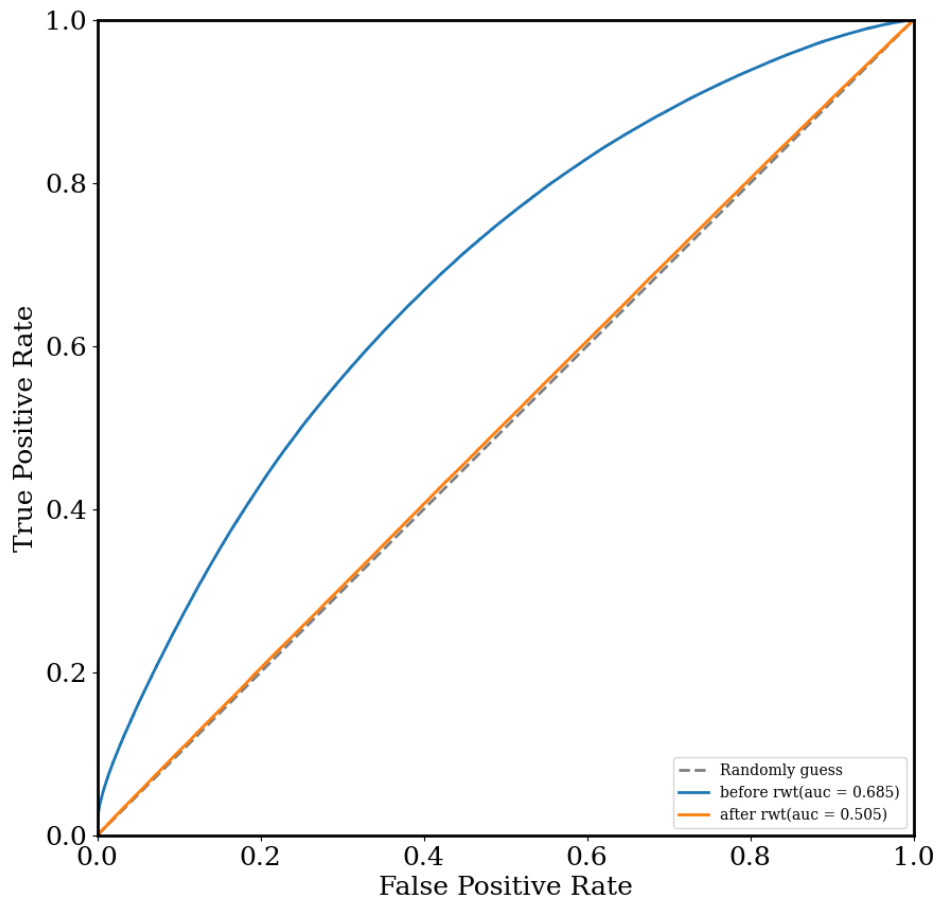
Figure B.12: ROC curve of the 2017 MagDown reweighted MC sample.
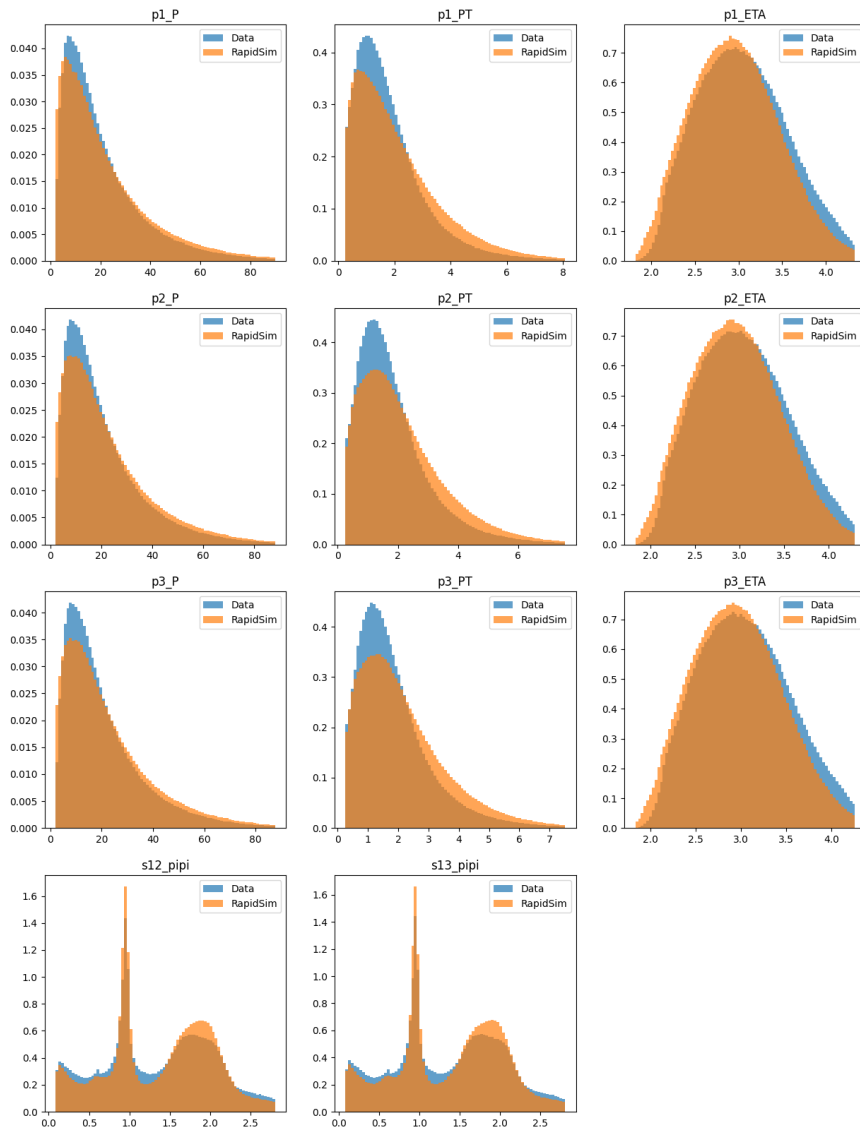
Figure B.13: Comparison between the reweighted variables distributions of the Data and MC 2018 MagUp sample, before the reweighting procedure. All variables are shown in GeV.
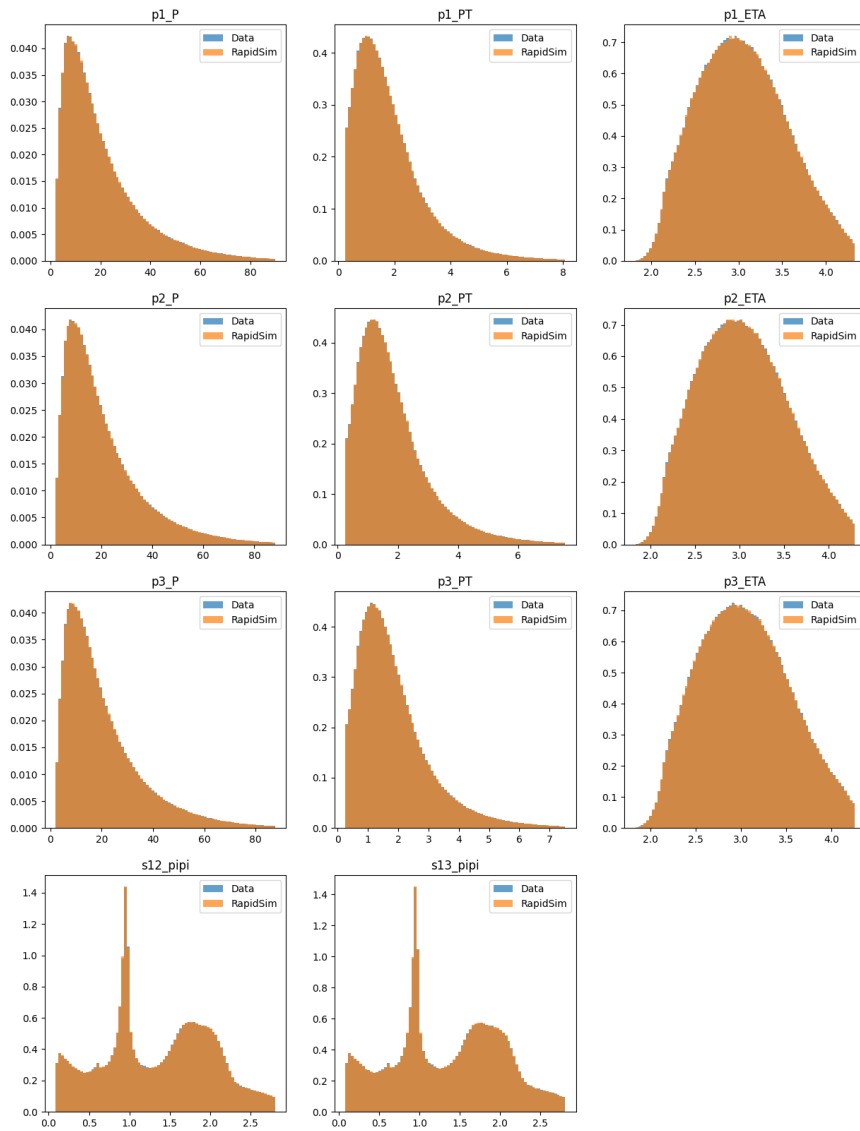
Figure B.14: Comparison between the reweighted variables distributions of the Data and MC 2018 MagUp sample, after the reweighting procedure. All variables are shown in GeV.
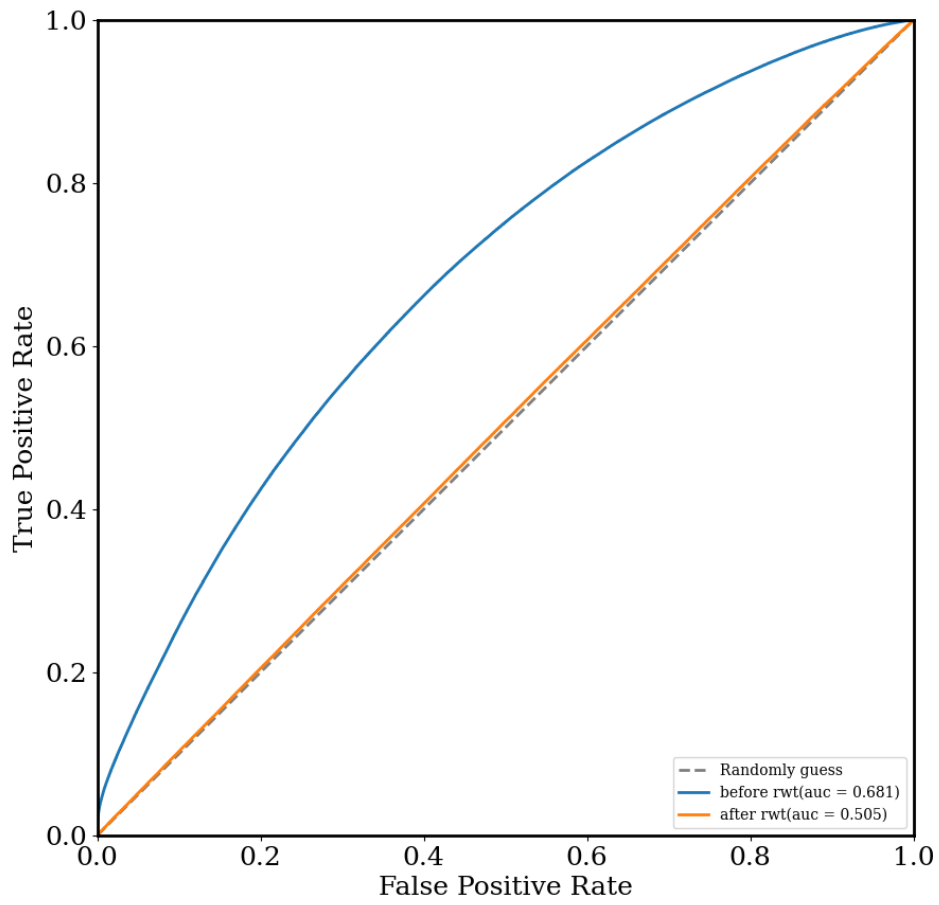
Figure B.15: ROC curve of the 2018 MagUp reweighted MC sample.

C
Appendix: Mass fits results

The plots of the mass-per-bin fits performed on the $D_s^+ \to \pi^-\pi^+\pi^+$ data sample and explained in detail in Chapter 5 are shown bellow.

Some of the fits may not have an ideal pull distribution, such as bin 16, which presented a displacement upwards in the low-mass region, for the MagDown D_s^+ components, for example. Also, in bin 11, the specific background fraction was fitted to be zero, killing the $D_s^+ \to \eta'\pi^+$ PDF.
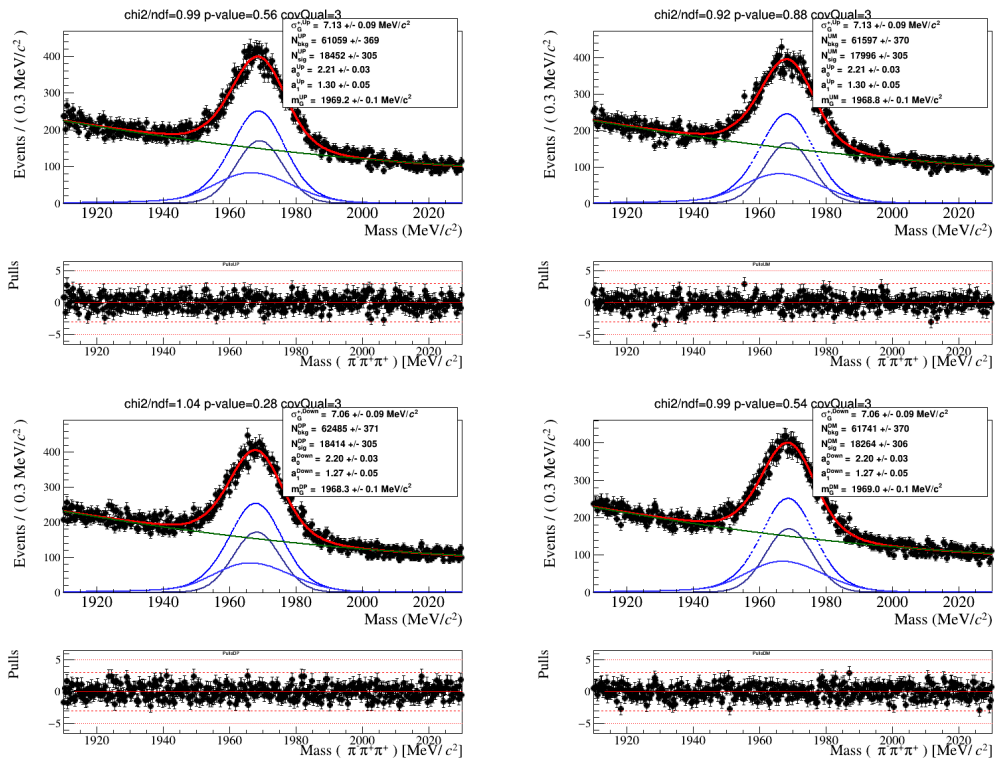
Figure C.1: Mass fit for bin 1
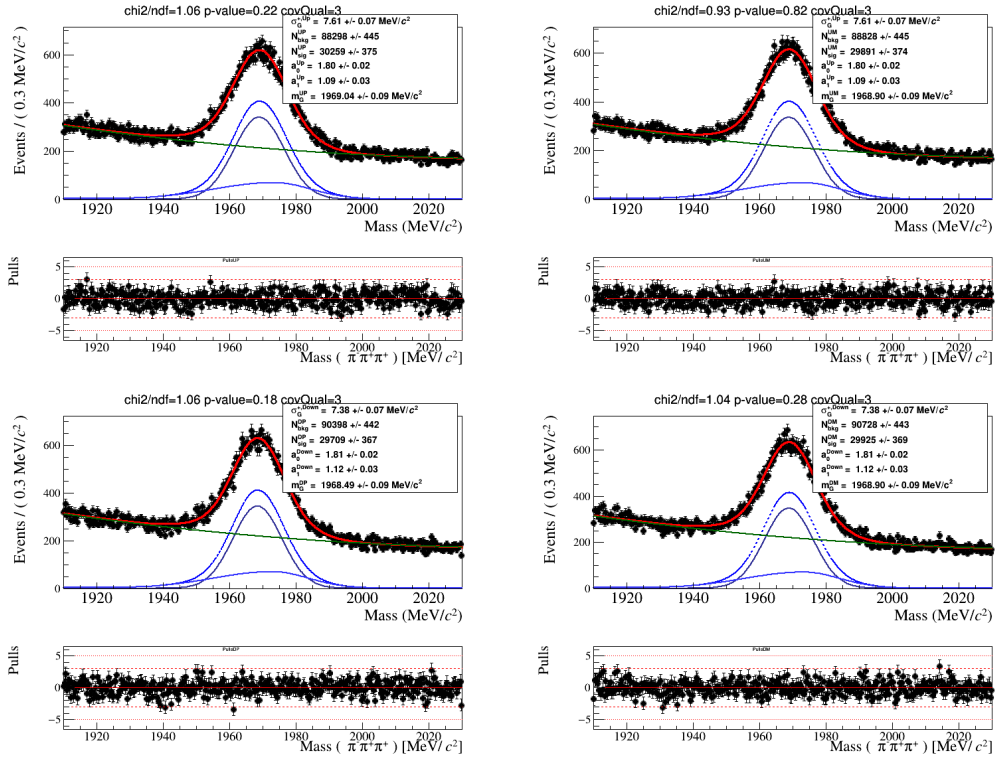
Figure C.2: Mass fit for bin 2
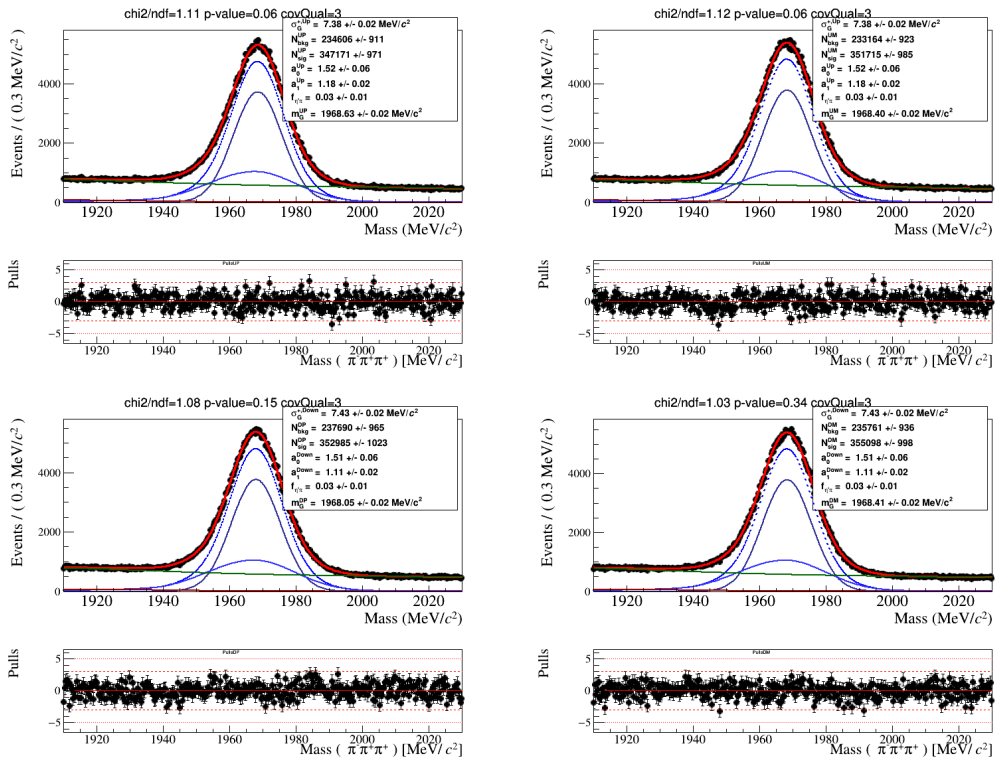
Figure C.3: Mass fit for bin 3
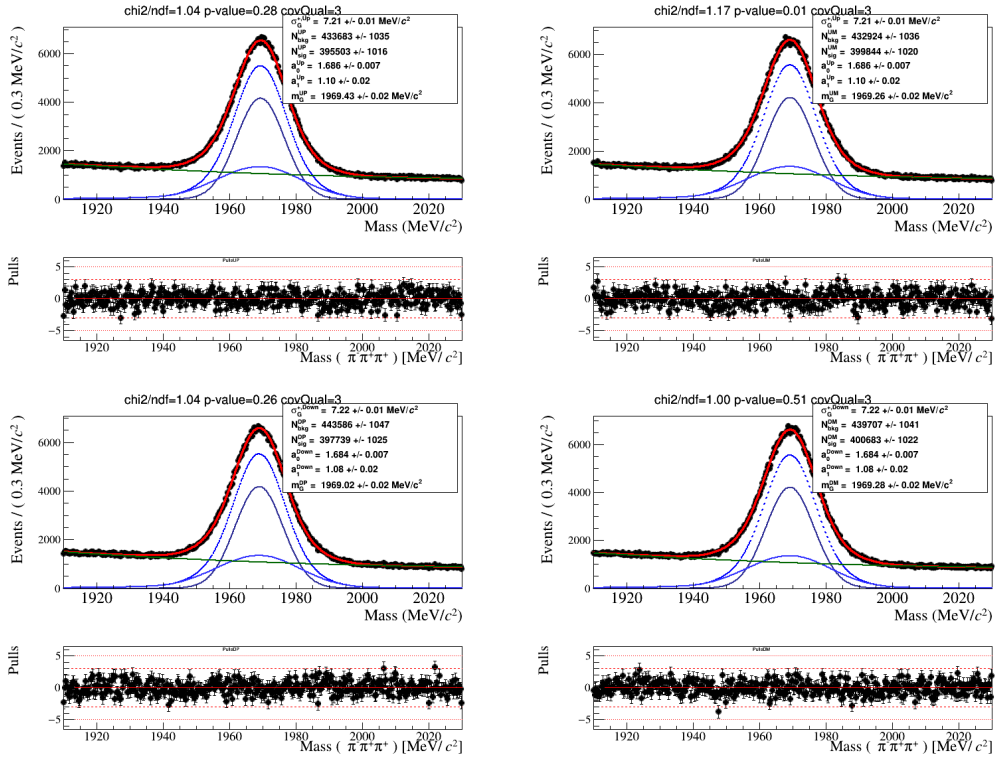
Figure C.4: Mass fit for bin 4
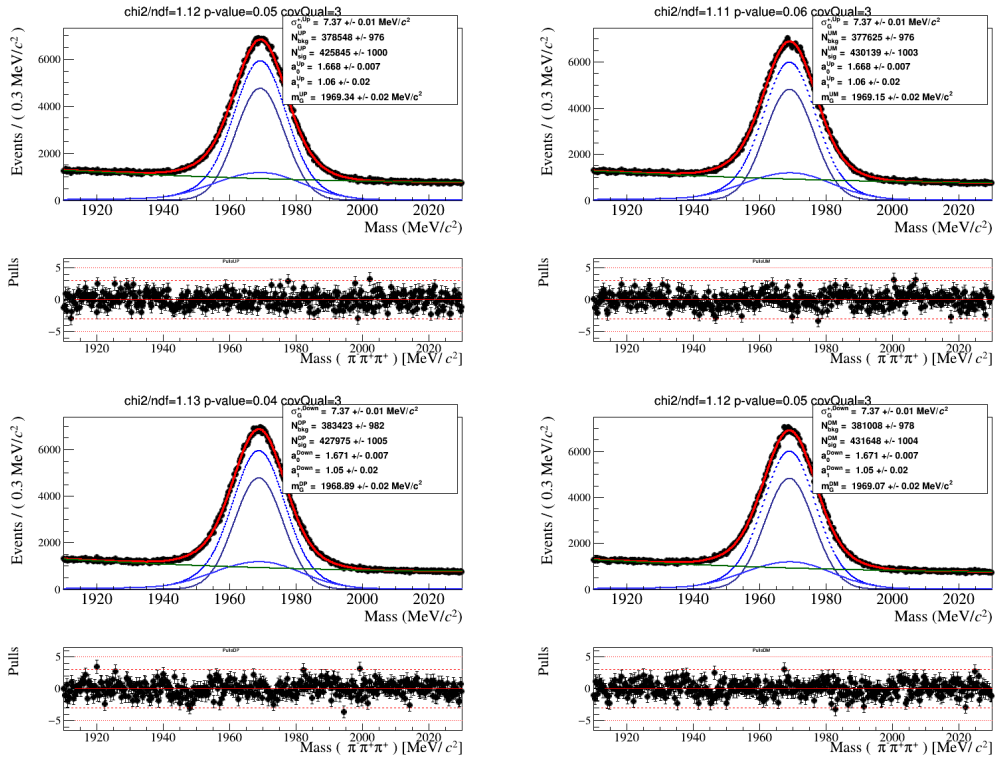
Figure C.5: Mass fit for bin 5
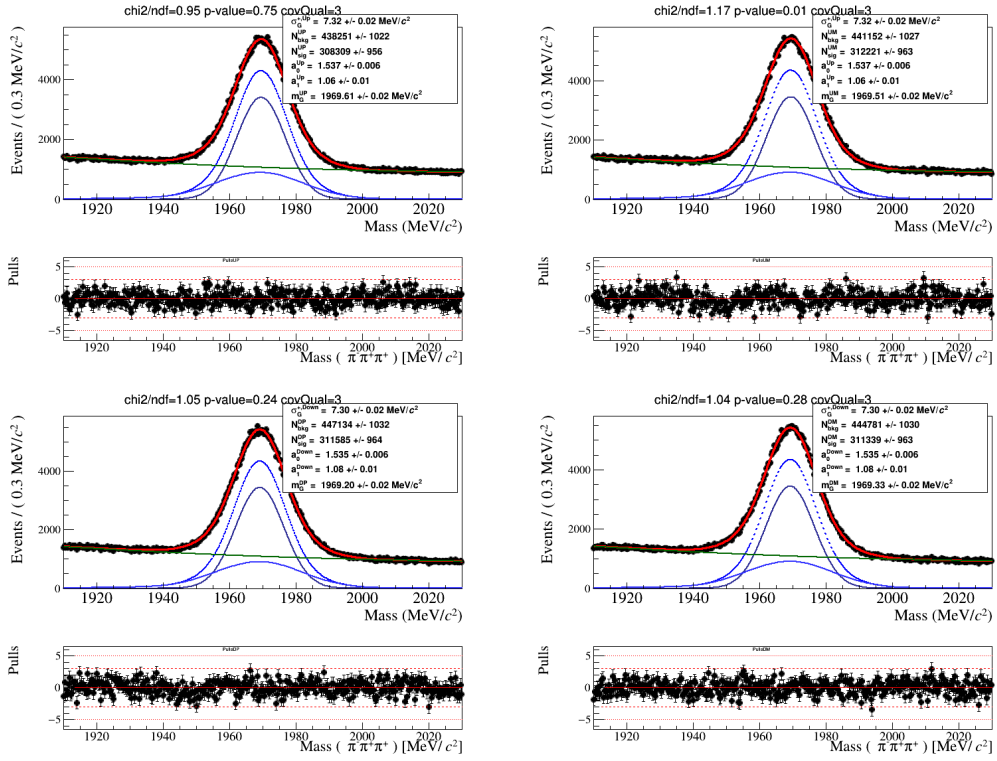
Figure C.6: Mass fit for bin 6
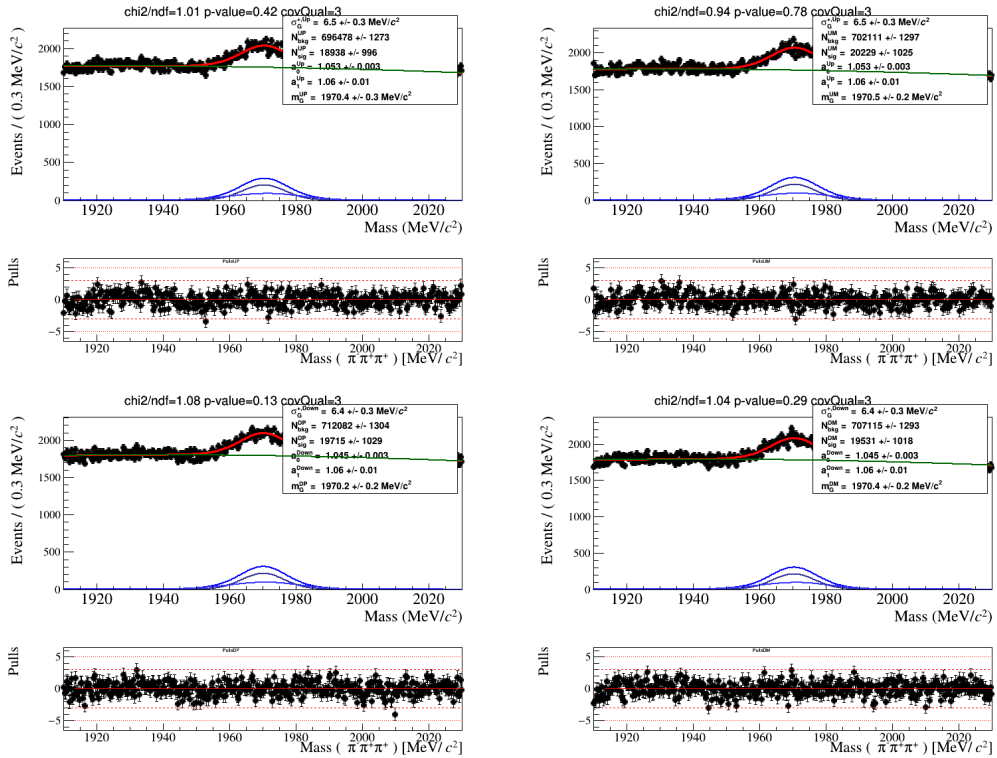
Figure C.7: Mass fit for bin 7
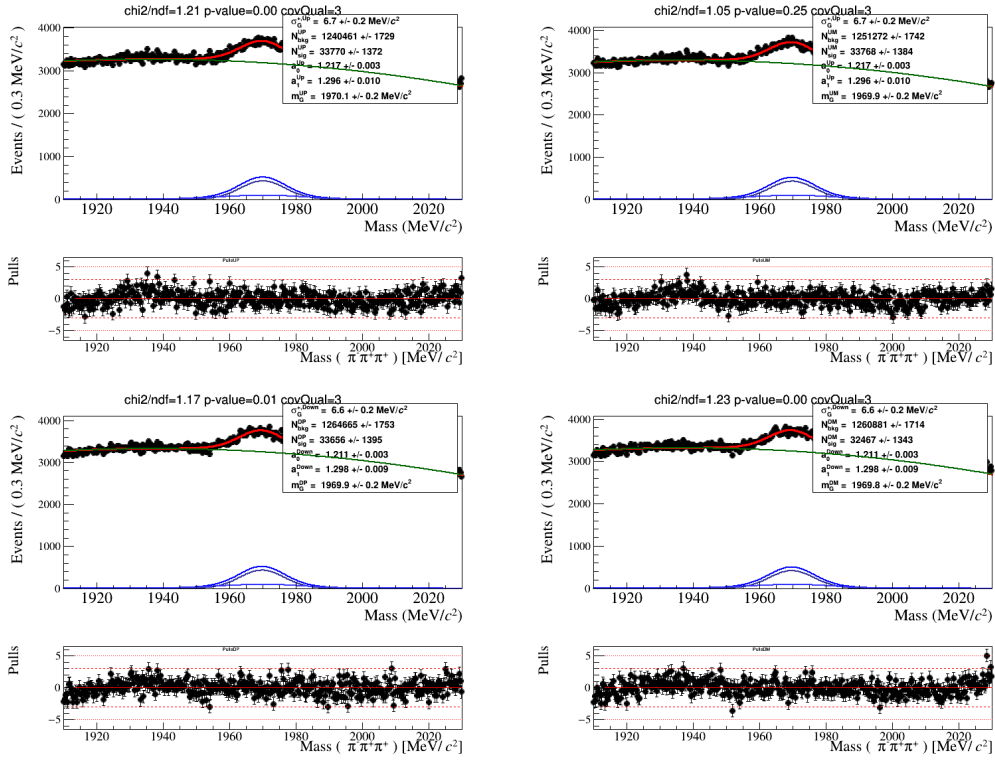
Figure C.8: Mass fit for bin 8
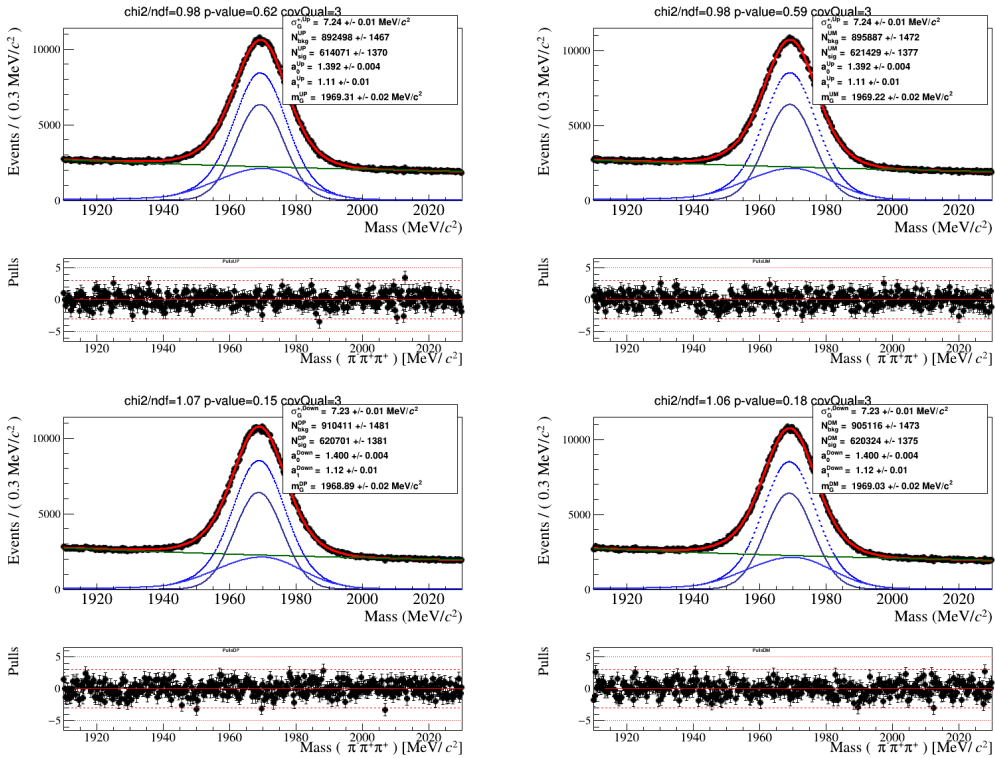
Figure C.9: Mass fit for bin 9
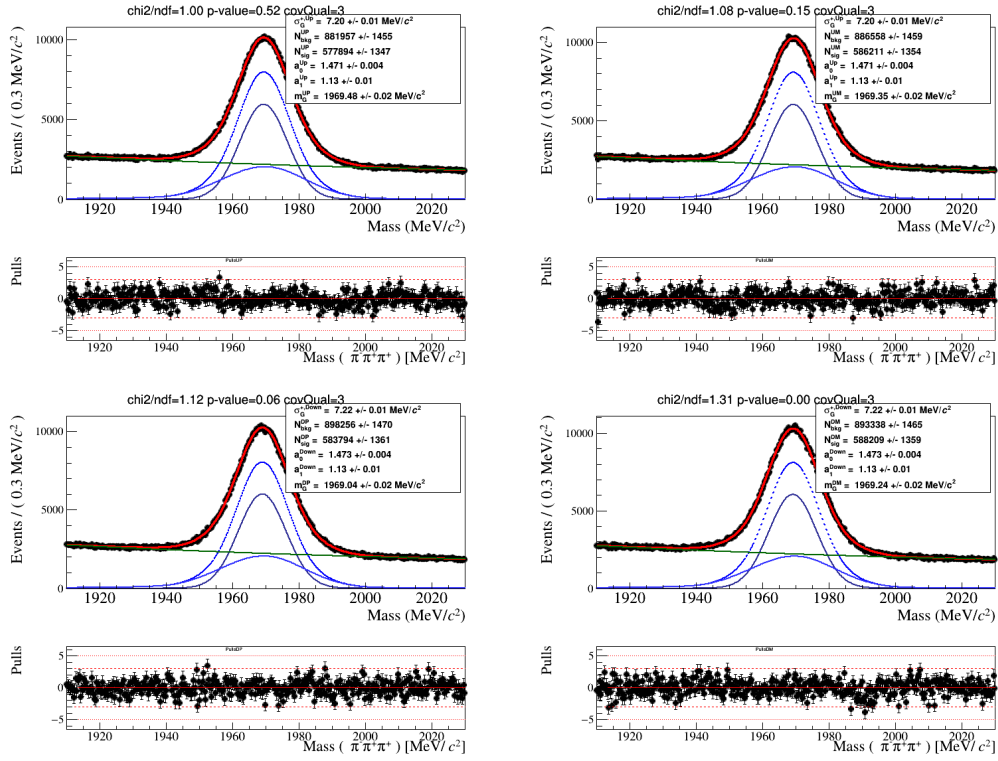
Figure C.10: Mass fit for bin 10
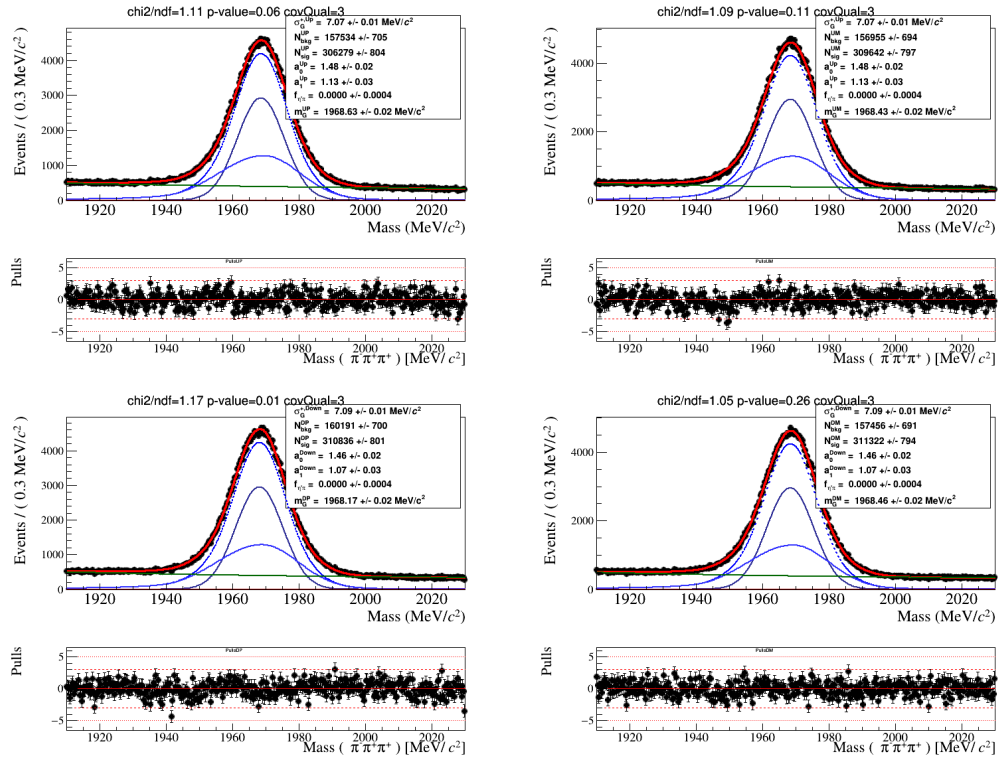
Figure C.11: Mass fit for bin 11
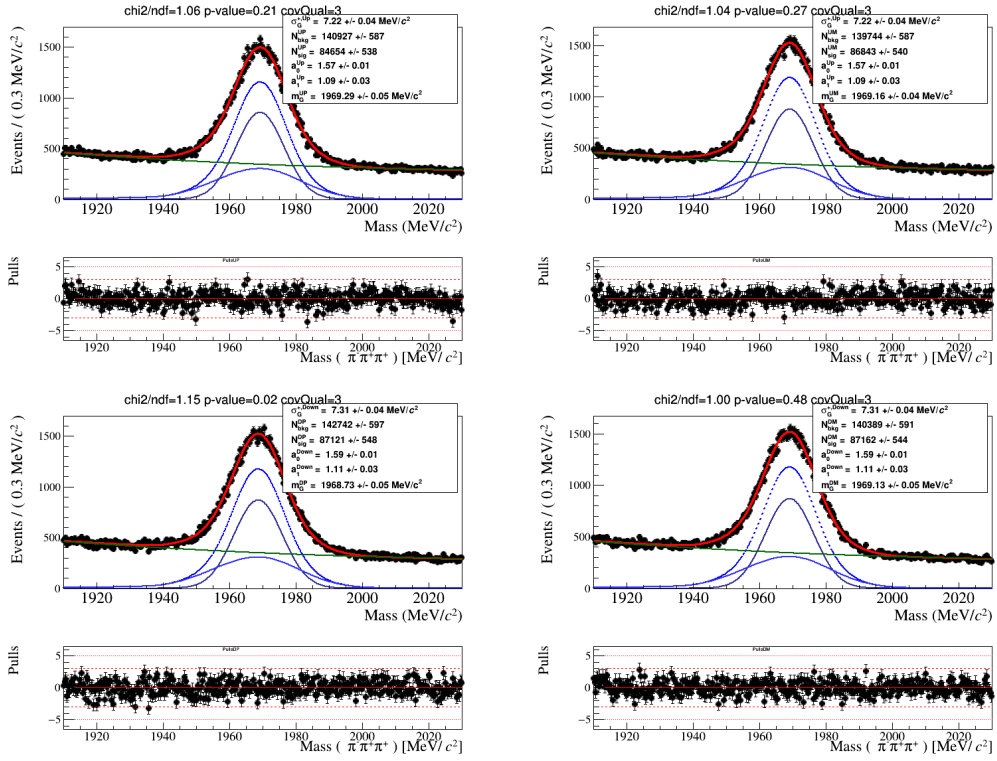
Figure C.12: Mass fit for bin 12
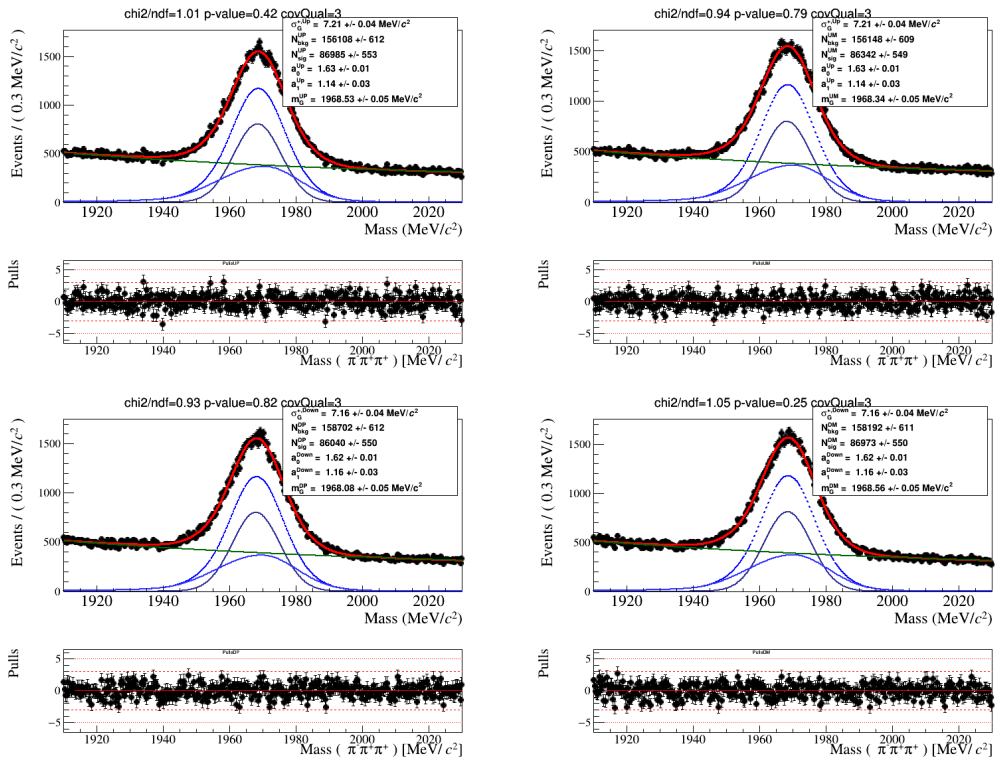
Figure C.13: Mass fit for bin 13
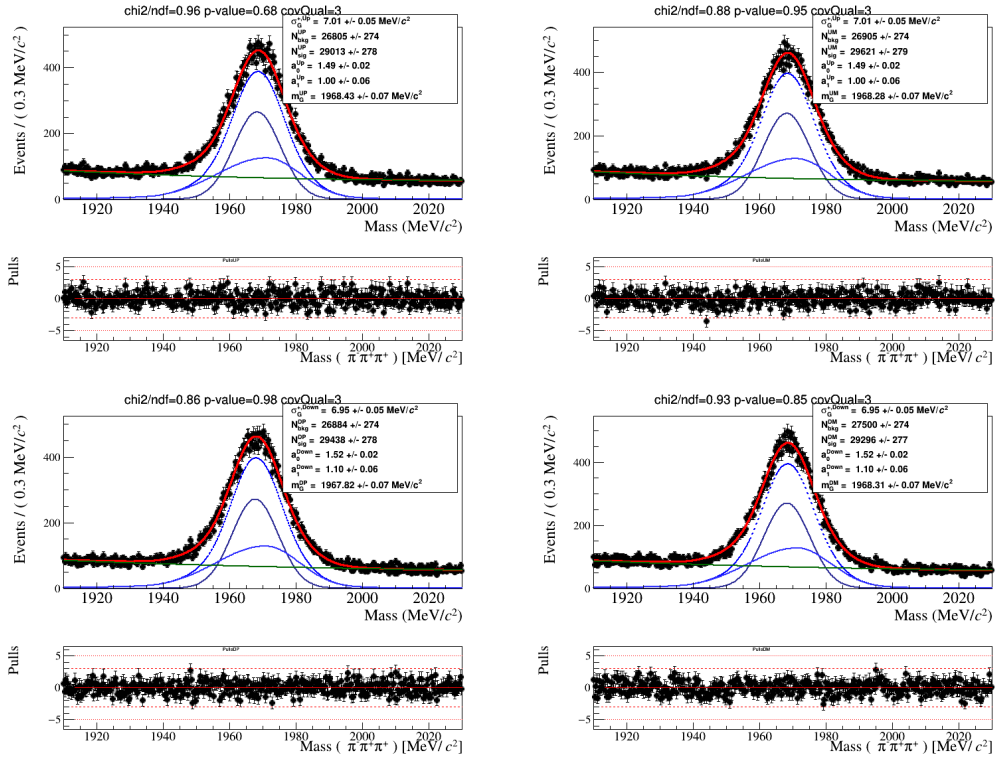
Figure C.14: Mass fit for bin 14
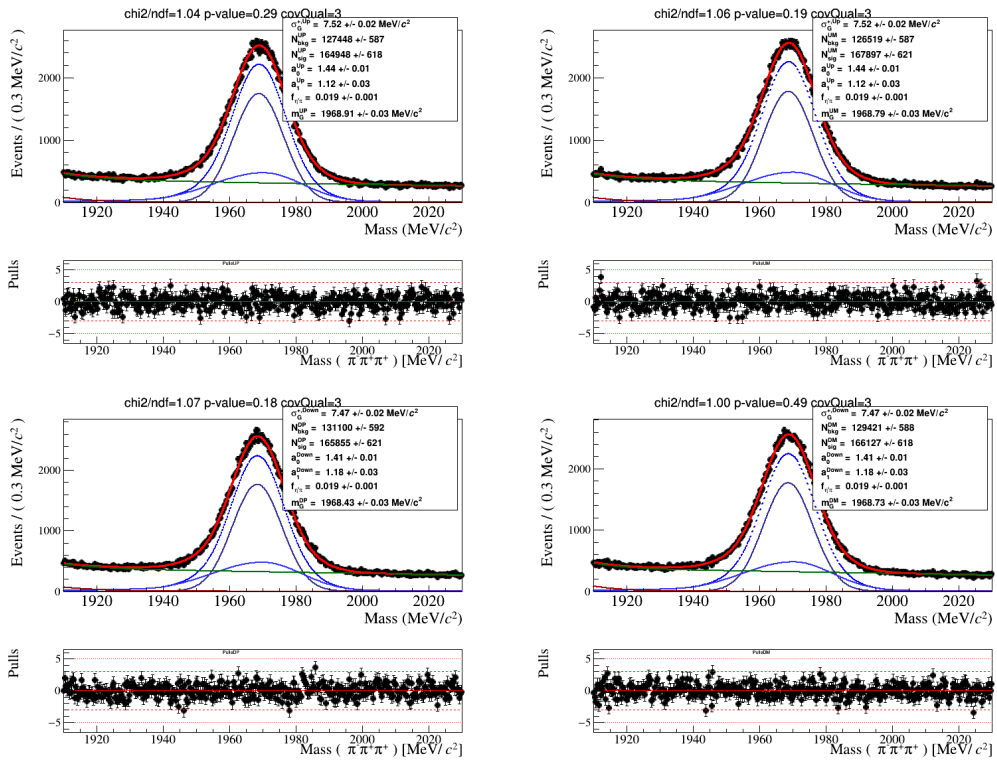
Figure C.15: Mass fit for bin 15
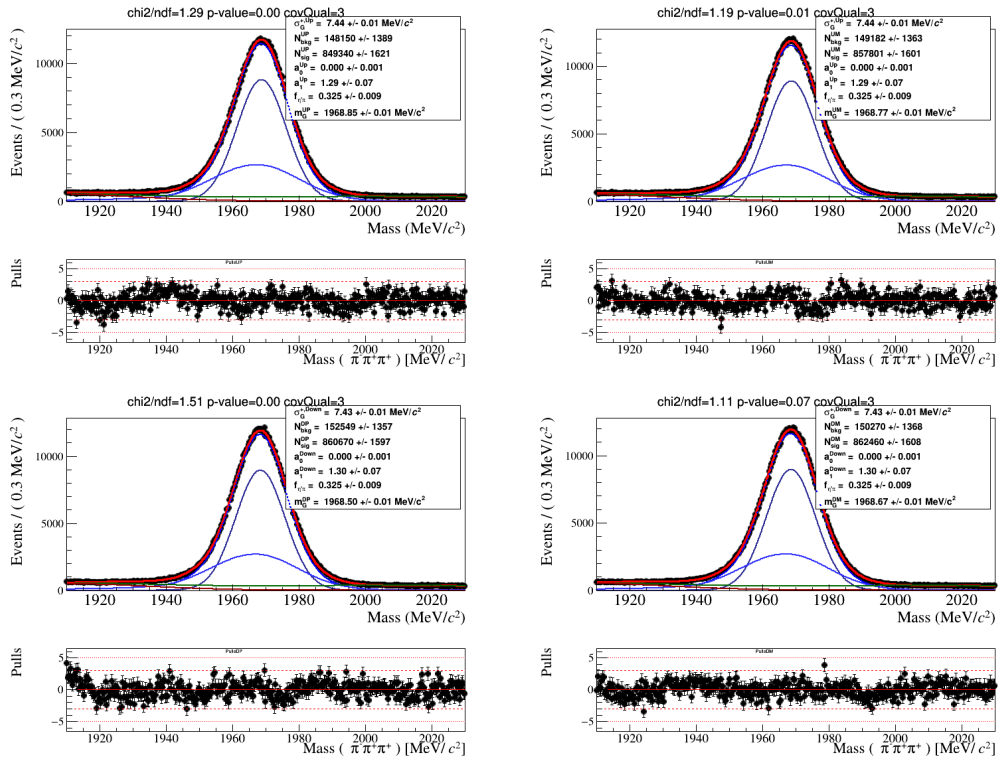
Figure C.16: Mass fit for bin 16
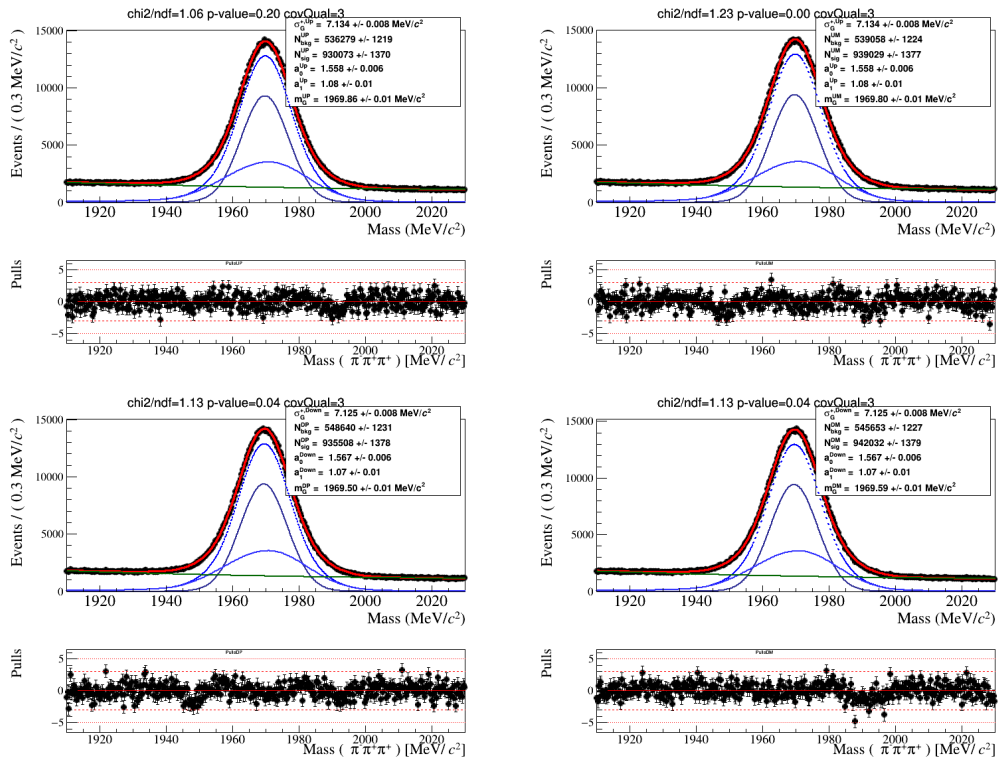
Figure C.17: Mass fit for bin 20
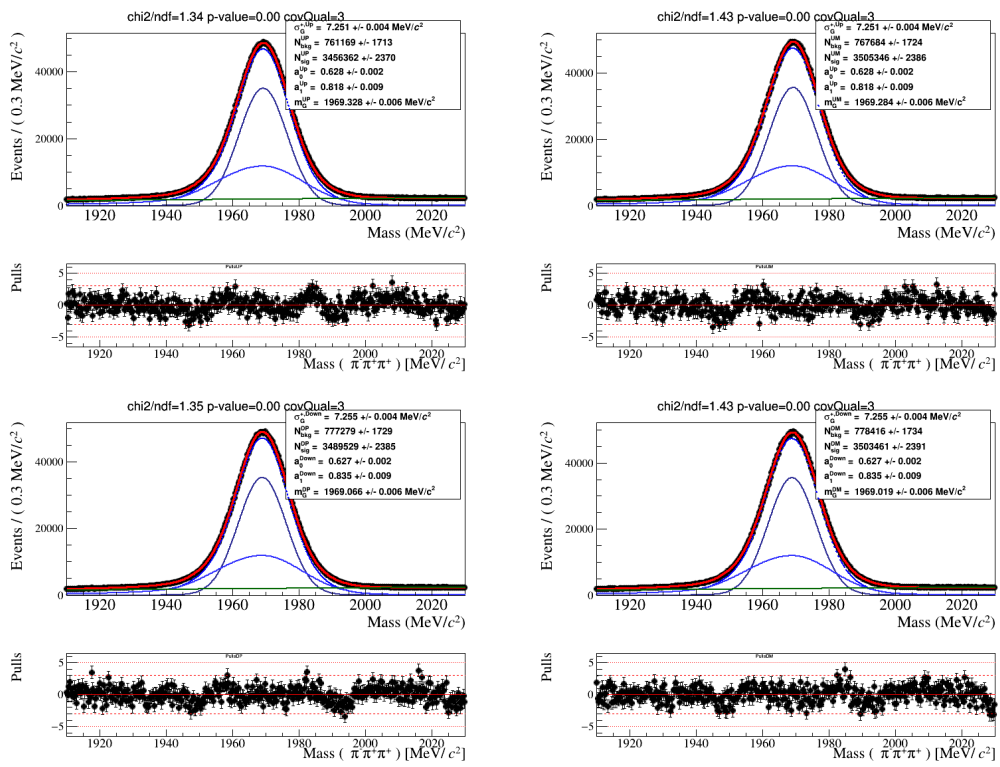
Figure C.18: Mass fit for bin 21

D
Appendix: $\mathcal{S}_{CP}$ results

Regarding the RapidSim samples, different strategies to include the asymmetries were used. For the production asymmetry, three models were employed: one based on Run1 measurements, other based on `Pythia` event generator predictions and another intensifying the asymmetry dependency with the transverse momentum[55]. A small shift of 0.5% was also tested for the detection asymmetry.

Also, the results for the RapidSim $\mathcal{S}_{CP}$ using the errors from the mass fit of the data sample are presented. It is possible to see through Equation 5-11 that a larger error in the denominator smooths out the $\mathcal{S}_{CP}$ value. As we could not achieve better errors using this strategy yet, those results are less conclusive.
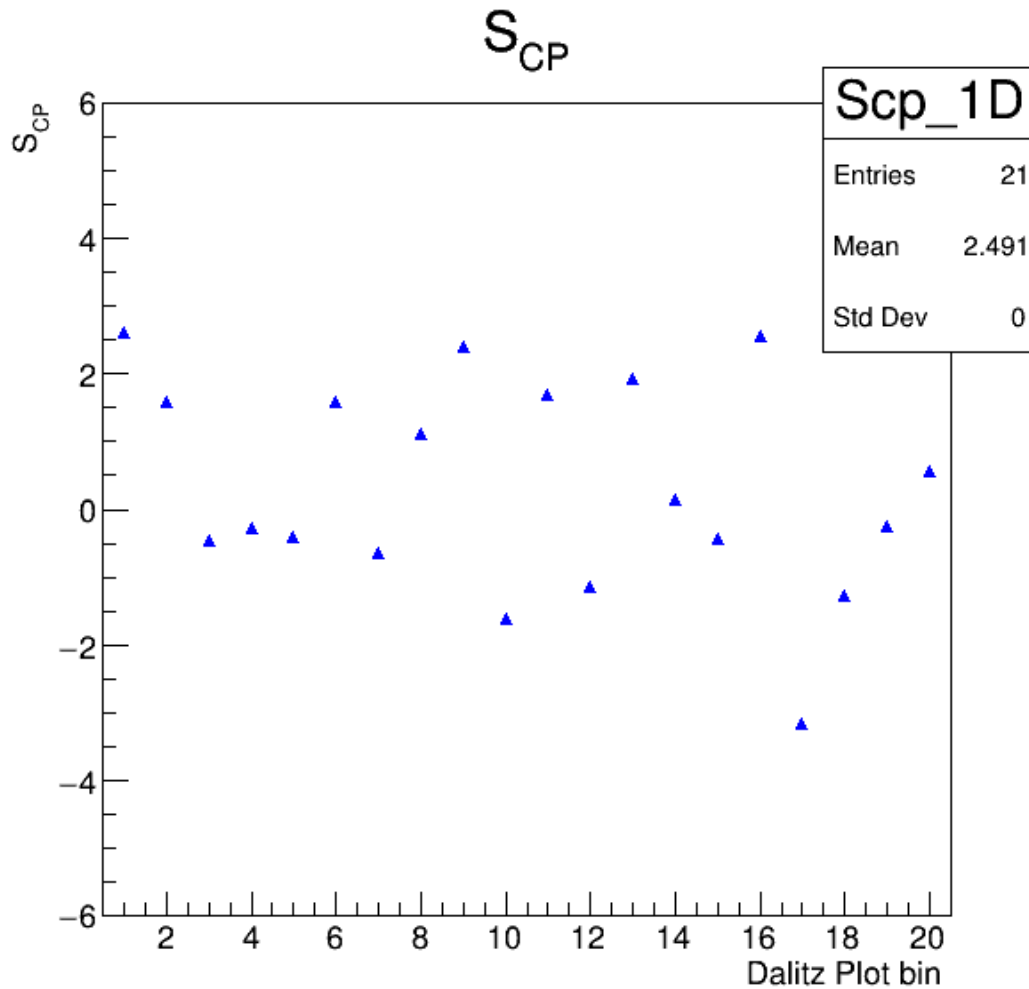
Figure D.1: 1D plot of the $\mathcal{S}_{CP}$ of the data sample
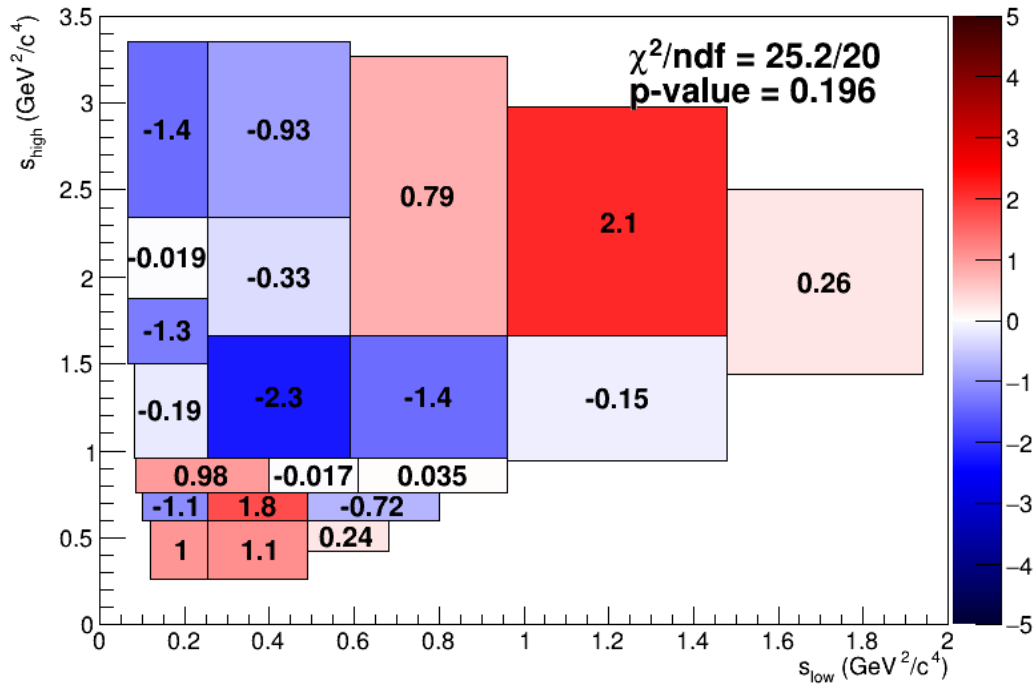
Figure D.2: $\mathcal{S}_{CP}$ result for the RapidSim sample using the Run1 model for A_{prod} and shifting A_{det}
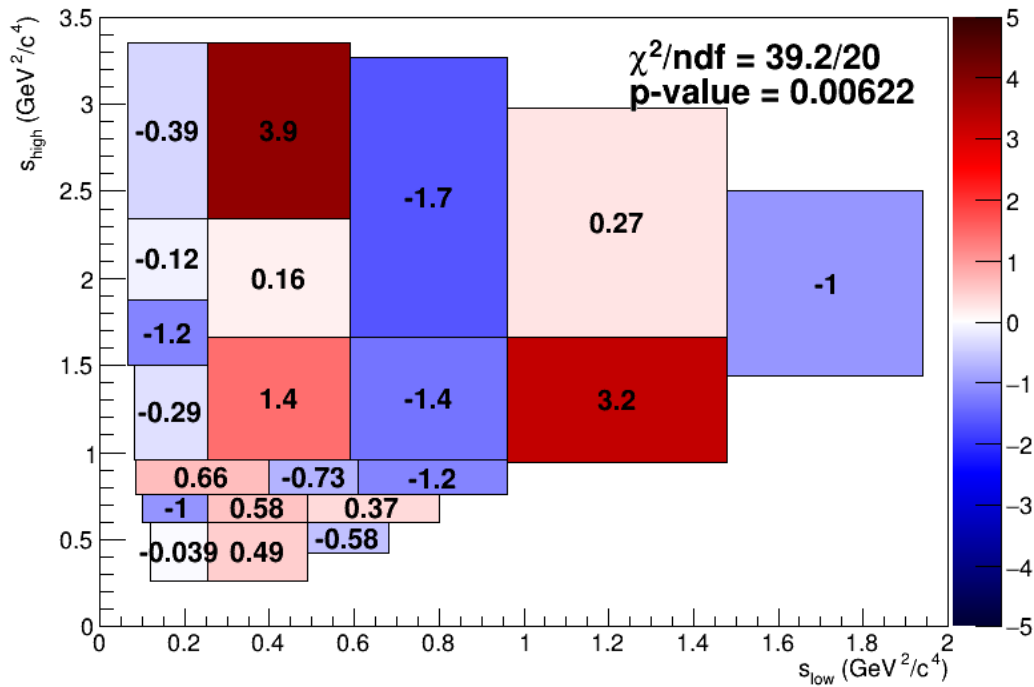
Figure D.3: $\mathcal{S}_{CP}$ result for the RapidSim sample using the Run1 model for A_{prod}
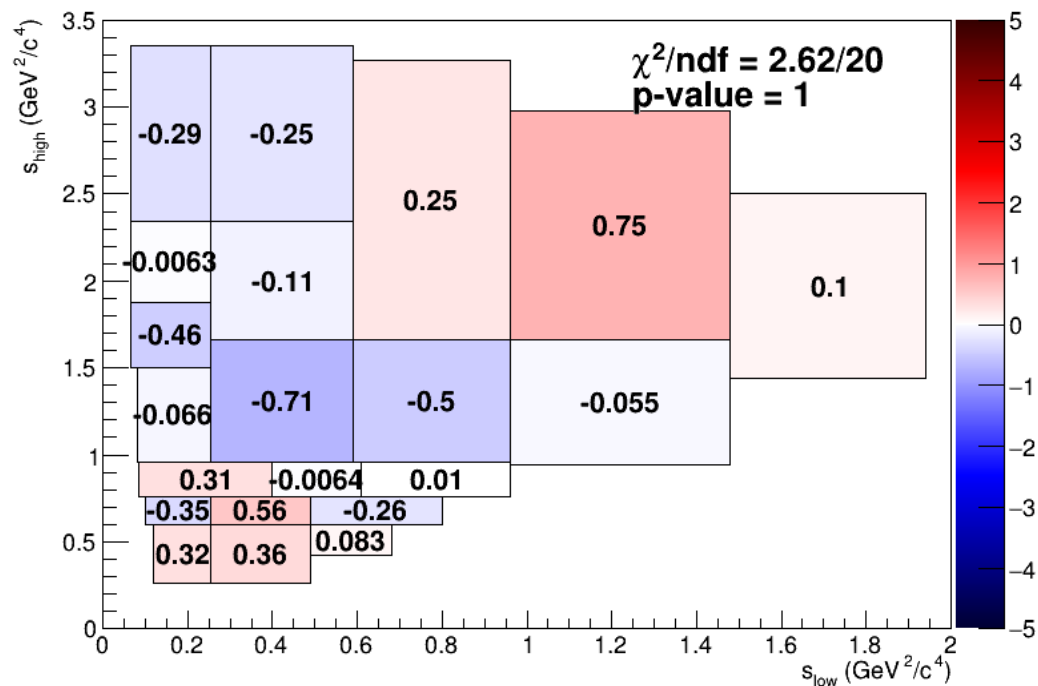
Figure D.4: $\mathcal{S}_{CP}$ result for the RapidSim sample using the Run1 model for A_{prod}, shifting A_{det} and using the errors from the mass fits
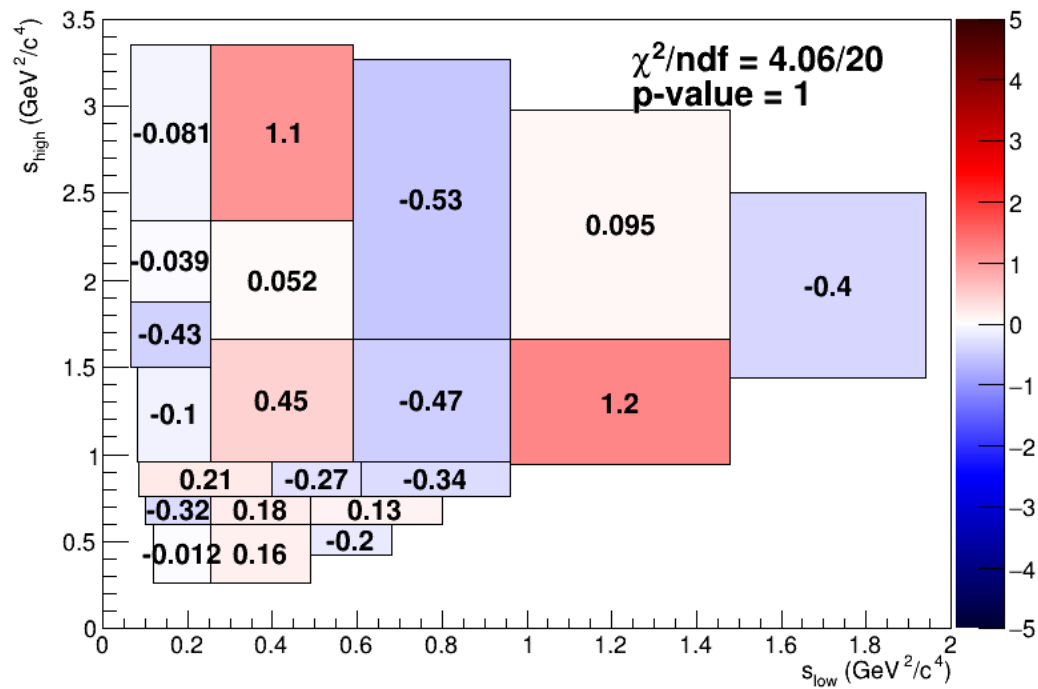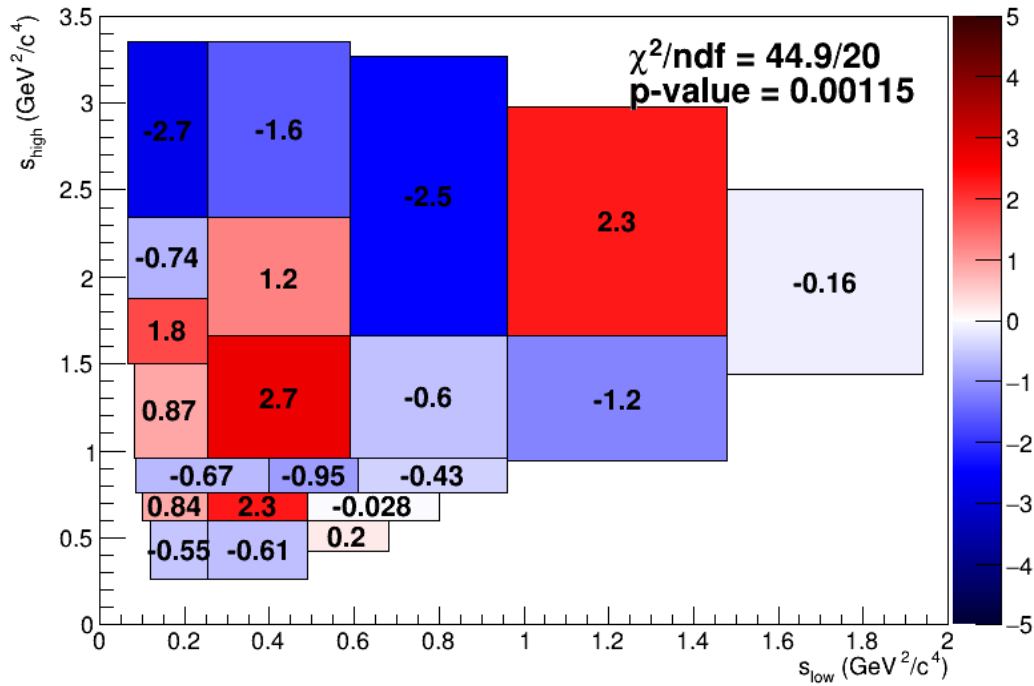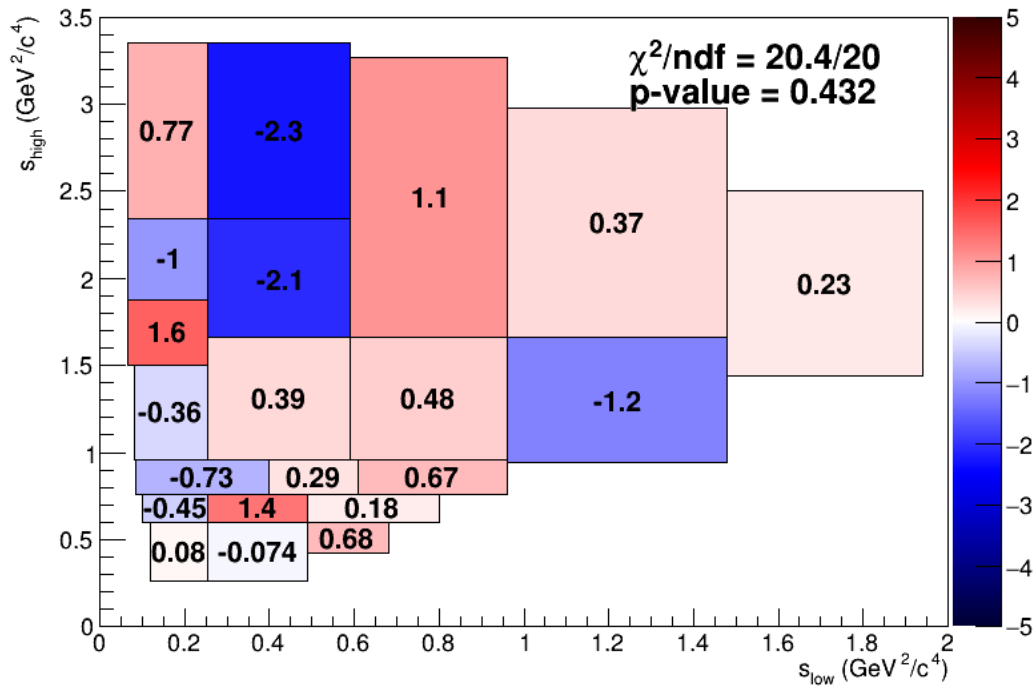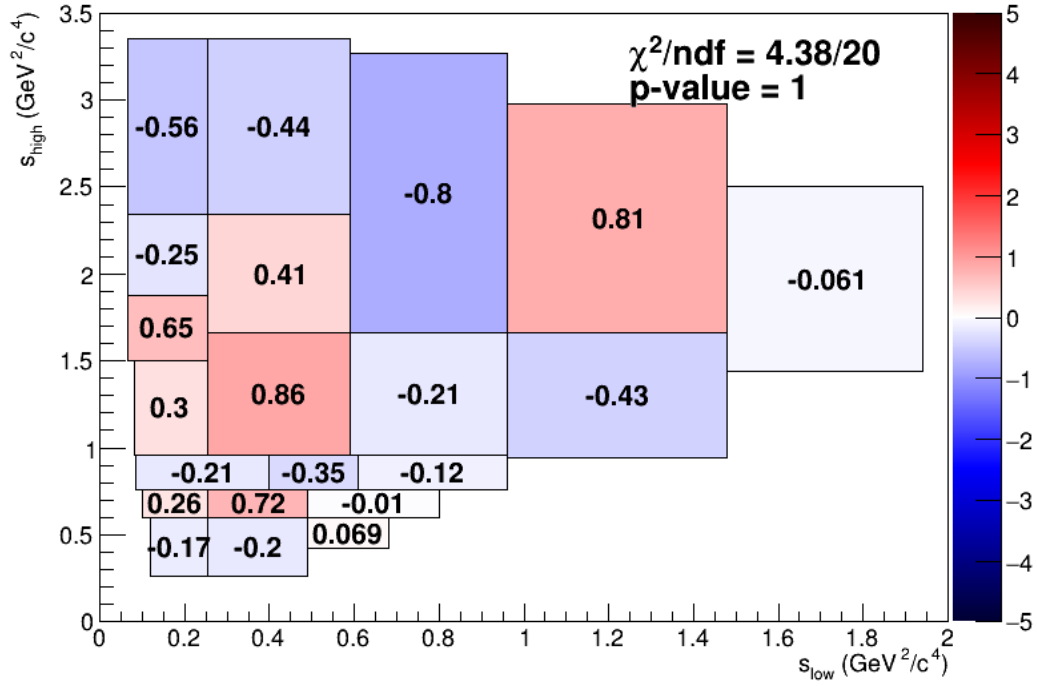

Figure D.5: $\mathcal{S}_{CP}$ result for the RapidSim sample using the Run1 model for A_{prod} and the errors from the mass fits

Figure D.6: $\mathcal{S}_{CP}$ result for the RapidSim sample using the first alternative model for A_{prod} and shifting A_{det}

Figure D.7: $\mathcal{S}_{CP}$ result for the RapidSim sample using the first alternative model for A_{prod}

Figure D.8: $\mathcal{S}_{CP}$ result for the RapidSim sample using the first alternative model for A_{prod}, shifting A_{det} and using the errors from the mass fits
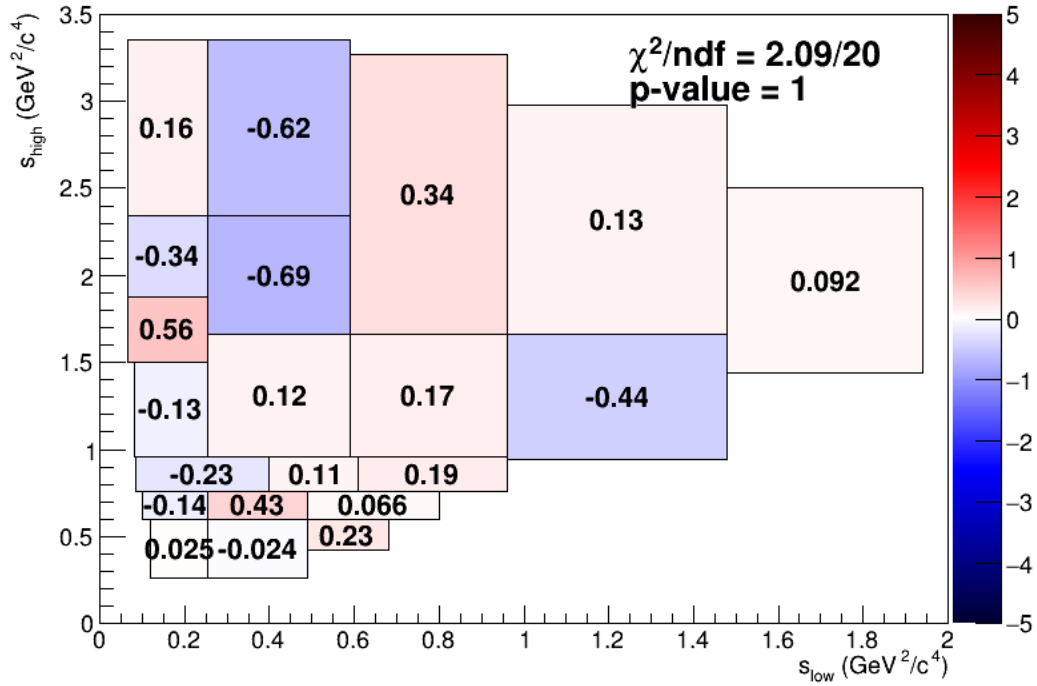

Figure D.9: $\mathcal{S}_{CP}$ result for the RapidSim sample using the first alternative model for A_{prod} and using the errors from the mass fits
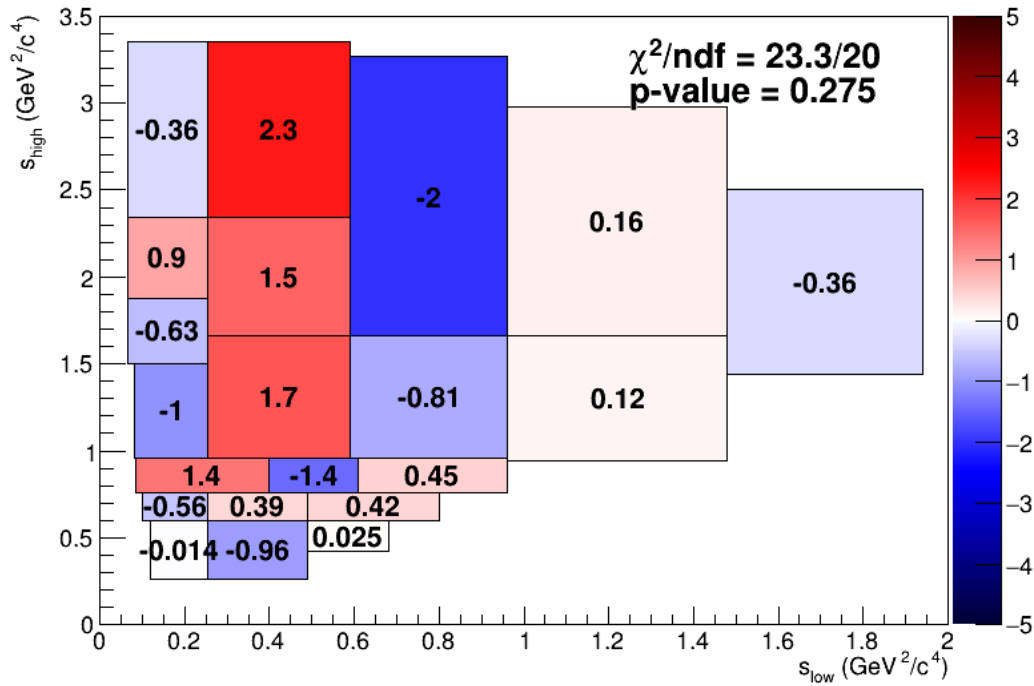
Figure D.10: $\mathcal{S}_{CP}$ result for the RapidSim sample using the second alternative model for A_{prod} and shifting A_{det}
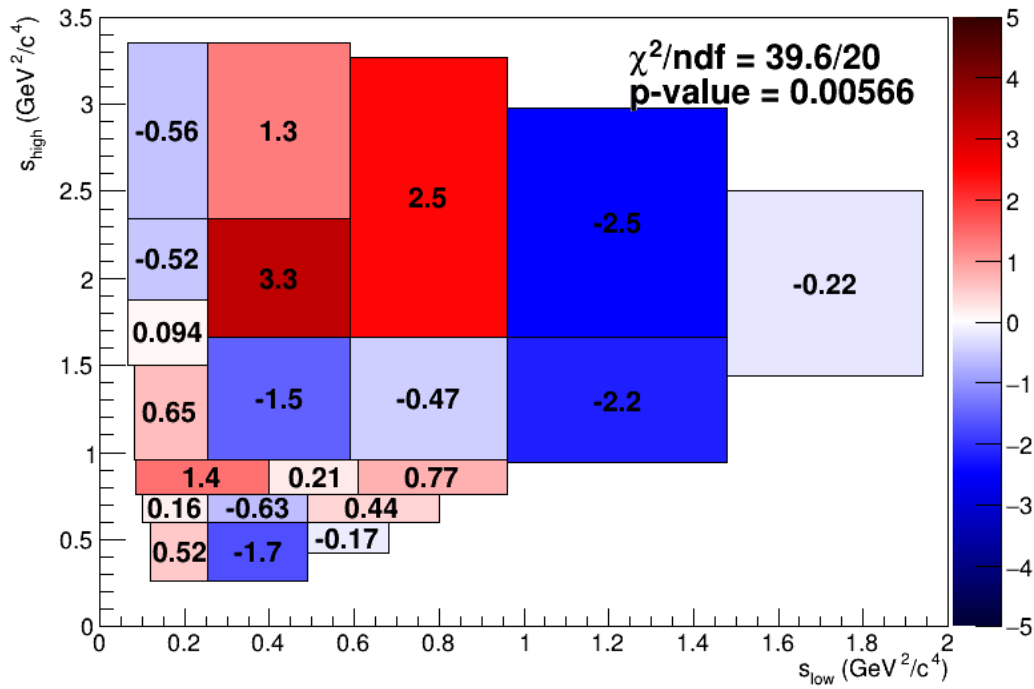

Figure D.11: $\mathcal{S}_{CP}$ result for the RapidSim sample using the second alternative model for A_{prod}
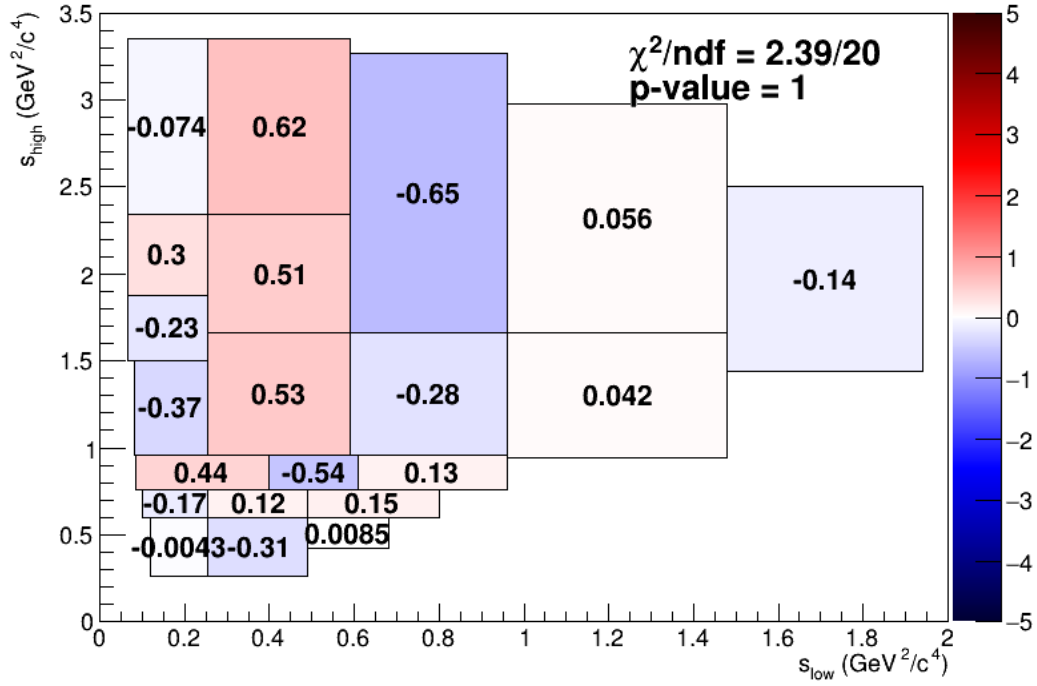
Figure D.12: $\mathcal{S}_{CP}$ result for the RapidSim sample using the second alternative model for A_{prod}, shifting A_{det} and using the errors from the mass fits
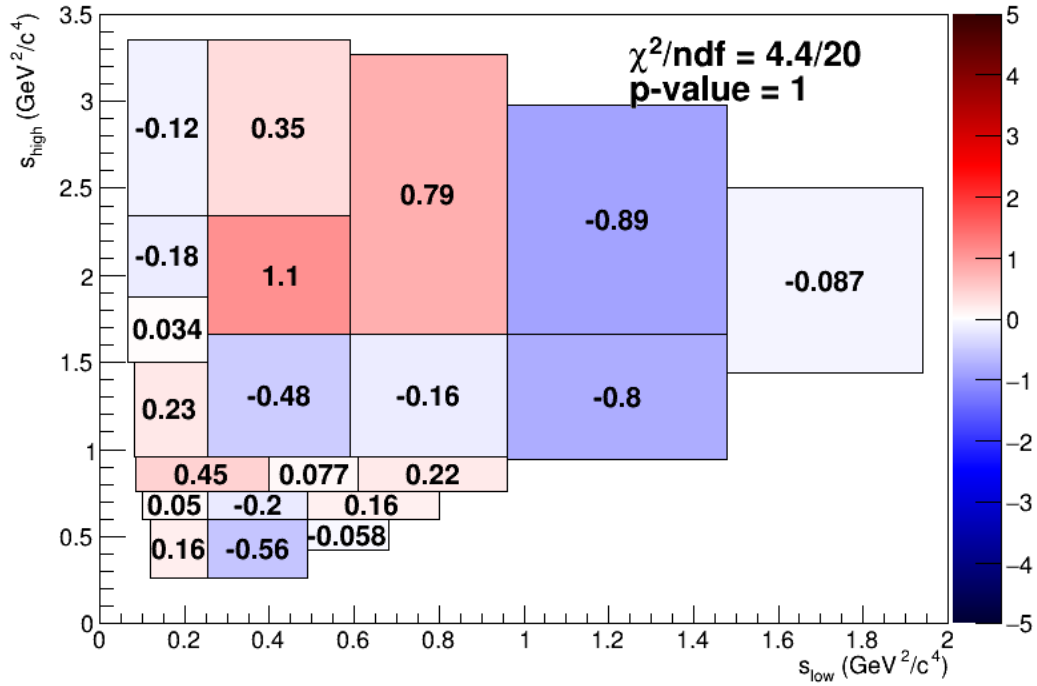

Figure D.13: $\mathcal{S}_{CP}$ result for the RapidSim sample using the second alternative model for A_{prod} and using the errors from the mass fits