CERN Summer Student Programme 2021

**Alexandre BIGOT**

# STRANGENESS TRACKING IN THE UPGRADED ALICE DETECTORS STUDY OF $\Xi_b^- \rightarrow (\Xi_c^0 \rightarrow \Xi^- \pi^+)\pi^-$

under the supervision of

David DOBRIGKEIT CHINELLATO & Matteo CONCAS

June 28th 2021 to August 20th 2021

Faculté de physique et ingénierie
Université de Strasbourg

Institut Pluridisciplinaire Hubert CURIEN STRASBOURG

# Contents

# 1 Introduction

With the recent upgrades of the ALICE (*A Large Ion Collider Experiment*) detectors and especially the one of the *Inner Tracking System* (ITS) [1], the first detection layers are closer to the primary vertex (the collision point). Before the ITS upgrade, the very first layer had a radius of 39 mm with respect to the beam axis, but now the first layer is "only" 22.4 mm away (Figure 1.2 of [1]). This is a huge improvement in secondary vertex reconstruction which is important for short-lived heavy-flavour hadrons like $\Xi_b^-$ for instance, which is at the heart of this study.

After being produced in the collision, the hadrons (which have a certain lifetime) decay via weak interaction to give birth to new particles. The measurements linked to this decay are indirect in the sense that the mother particle has a short flight distance such that it can't reach the detection system (the flight distance of $\Xi_b^-$ is about 470 $\mu$m while the first detection layer is 22.4 mm away from the primary vertex) so we have to study its descendants in order to reconstruct the tracks corresponding to a given decay channel, this is tracking.

However, in the case where we can detect the mother particle of a vertex ($\Xi^-$ in the decay $\Xi^- \rightarrow \Lambda\pi^-$ for instance) before it decays and that this mother particle is a hyperon [1] or a hypertriton [2], we do what we call *strangeness tracking*. This brand new technique is just starting to be developed. With it, the idea is to find signatures of weakly decaying hyperons (such as $\Xi^-$) or hypertritons before they decay [3], using silicon detectors (composing the new version of the ITS [2]) that are very close to the primary vertex and shall allow us to get unprecedented precision in measuring these particles during Run 3 of the LHC and beyond [4]. Indeed, this is part of a bigger effort to conceive a new detector in the future : ALICE 3 (foreseen to have three layers inside the beam pipe, and the innermost will be at 5mm) that is the foreseen geometry of the ALICE detector after LS3 [4].

This project is focused on the study of an open beauty hadron, $\Xi_b^-$(dsb), via a specific decay channel : $\Xi_b^- \rightarrow (\Xi_c^0 \rightarrow \Xi^-\pi^+)\pi^-$. This channel has been chosen for two main reasons : (i) it contains a hyperon (hence the possibility of strangeness tracking) and (ii) the particle at the origin of this channel is heavy-flavoured. Besides, this channel is not very well known (the branching ratio has not been measured), so it shall be the opportunity to learn more about it.

The desire is to create an analysis prototype using a state-of-the-art detector, the goal would be to isolate signal from background because, in the end, we want to measure the production yields of the particles and to do so, we need clean [4] invariant mass distributions. Of course the final scope of such an analysis is doing some physics, and the b quark is mostly an unexplored beast, we might learn a lot from studying how it travels through the Quark Gluon Plasma (QGP) and eventually how it gets packaged into a hadron (hadronization mechanisms as fragmentation or coalescence) as $\Xi_b^-$ for instance.

# 2 Simulation of collisions

The data studied during this project are not real data as they don't come from measurements of the ALICE detector, they exclusively come from a Monte-Carlo simulation [5] (*cf* Figure 1) that

1. Hyperons are unstable particles composed of three valence quarks whose at least one is a s quark and none is a c or b quark (*idem* for antiquarks).
2. A hypertriton is a type of hypernucleus, formed of a proton, a neutron and any hyperon.
3. We feed the tracking algorithm with these direct measurements that shall act as corrections and improve the resolution.
4. Here "clean" means that signal is well distinguished from background.
5. This allows us to go beyond the curent geometry of ALICE and study collisions with the ITS3 (the future version of the ITS that should be installed during LS3 of LHC [3], [4]) whose innermost layer is positioned at 18

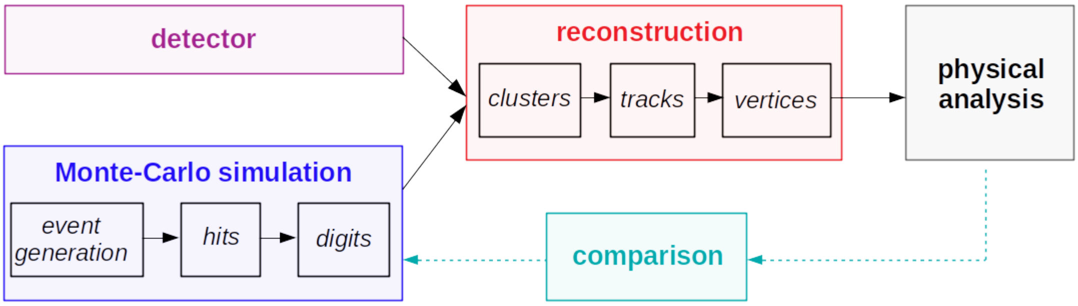



FIGURE 1 – Typical data flow for experiments in high energy physics.

can be summarized as follows : (i) a minimum bias Pythia [6] event with up to one particle on interest (it is $\Xi_b^-$ here) simulating the collision ; (ii) propagating the generated particles with GEANT3 ; (iii) simulating hits with the TRK geometry (beam pipe + ALICE 3) and the activation of the detectors by digitizing the data.
First, I was given some "training files" to get my hands on the Pythia generator. This code would simulate a minimum bias proton-proton collision at 13 TeV and allow us to play with its parameters, the type of particles produced (after collision) being one of them. My goal was to generate $\Xi_b^-$ particles and to be assured that these $\Xi_b^-$ would decay according to the wanted channel $\Xi_b^- \to \Xi_c^0(\Xi^-\pi^+)\pi^-$ [7]. So I had to understand how to translate it into instructions inside the generator by using a specific syntax [8].

Once the code is modified, one shall check if the particles have been properly generated, because one needs a data sample big enough to perform statistical analysis. To do so, we use the output files of the simulation code containing such pieces of information. For instance, for this project, among 200 simulated events [9], 200 $\Xi_b^-$ and 200 $\Xi_c^0$ particles were produced *a minima* [10], meaning Pythia generated the wanted amount of particles [11] to study the decay channel of interest.
We have gone through the "Monte-Carlo simulation" block in Figure 1 and now we are entering the next step of the process : the reconstruction.

## 3 Reconstruction of a decay channel

During a pp collision, until fifty particles can be produced (by unit of pseudo-rapidity), so associations of tracks may be wrong and constitute background noise, hence the need for a procedure minimizing contaminations from this background : topological reconstruction.

$\Xi_b^-$ can decay via many channels, leading to several daughters, grand-daughters, grand-grand-daughters, ... but here we're not interested in reconstructing the whole *family tree* underpinned by $\Xi_b^-$. Our goal shall be to reconstruct a specific *branch* of this *family tree* by going from the tip of the *branch* to the roots of the *tree*. This specific *branch* is presented Figure 2.

mm radial distance from the interaction point (even closer than the current ITS).

6. Pythia is a simulation code used to simulate pp collisions.

7. Then $\Xi^- \to \Lambda\pi^-$ with a branching ratio above 99.8% and $\Lambda$ mainly decays into $p\pi^-$ with a branching ratio of about 64%.

8. Each particle is identified by a "pdg code" [5] and specific command lines are used to declare (and possibly impose) a decay channel [6]. Information on the particles can be found in [7].

9. An event is a collision that triggered the detection apparatus. We can have either zero event (no trigger) or one event (trigger) for a given collision.

10. Some more of these particles could have been produced directly by the collision or via other decays taken into account by Pythia.

11. Note that these numbers are obtained for a single data generation. This generation code was launched thousands of times during the project, to get a data sample big enough.

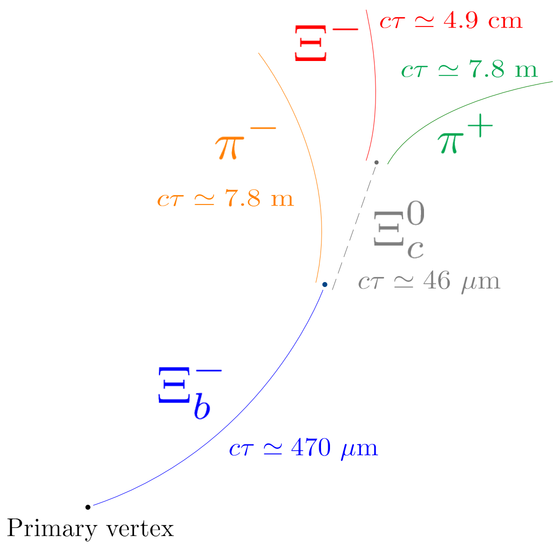



FIGURE 2 – Scheme of the decay channel $\Xi_b^- \to (\Xi_c^0 \to \Xi^-\pi^+)\pi^-$. The flight distances are indicated.

Via this channel, one aims at studying $\Xi_b^-$ disintegrating according to $\Xi_b^-(\mathsf{dsb}) \to \Xi_c^0(\mathsf{dsc})\pi^-(\bar{\mathsf{u}}\mathsf{d})$. Then the $\Xi_c^0$ decays in two particles with opposite electric charges such as $\Xi_c^0(\mathsf{dsc}) \to \Xi^-(\mathsf{dss})\pi^+(\mathsf{u}\bar{\mathsf{d}})$. As $\Xi_b^-$ leads to a succession of V shaped decays, it is called a cascade. In such a cascade, one could consider $\Xi_b^-$ as a mother giving birth to her daughters $\Xi_c^0$ and $\pi^-$. At this stage, only $\pi^-$ is detected because of its non null electric charge (contrary to $\Xi_c^0$ that is electrically neutral, hence undetected [a]) : it is called the *bachelor* particle. Then it is the turn of $\Xi_c^0$ to become a mother [b], giving birth to $\Xi^-$ and $\pi^+$ respectively called the *negative* and *positive* particles according to their electric charge. The distinction as a function of the sign of the electric charge is important because this sign affects the curvature of the particles tracks, an observable accessible through the measurements and used during the topological reconstruction procedure.

*a.* The detection relies on the ionization of the medium composing the detection layers which is only possible if the particle passing through is electrically charged.
*b.* $\Xi_b^-$ is then promoted to the status of grand-mother.

Topological reconstruction aims at associating the tracks of the *negative* ($\Xi^-$) and *positive* ($\pi^+$) particles to reconstruct the undetected $\Xi_c^0$, whose track shall itself be associated to a track of a *bachelor* particle ($\pi^+$) and eventually constitute a candidate to $\Xi_b^-$. Let's not forget that $\Xi^-$ can directly be detected so it allows strangeness tracking hence improved track reconstruction. This improvement shall feed through to the study of $\Xi_c^0$ and then $\Xi_b^-$.
We are not going to detail this procedure in this report. Yet we can mention that the main goal of topological reconstruction is to find the point in space where the tracks of two potential daughters are the closest (this point becomes a vertex candidate) and then use momentum conservation and mass invariance at this vertex candidate to extrapolate pieces of information about the particle that supposedly decayed at this very point.

Once we reconstructed $\Xi_c^0$, we repeat the procedure presented previously but instead of associating *positive* and *negative* particles, we associate a *bachelor* with the freshly reconstructed $\Xi_c^0$ to eventually reconstruct the track of $\Xi_b^-$. This whole procedure is done by a ROOT macro running in the O2 environment [12] and about 2 000 lines long. I was given access to such a macro that was functional but dedicated to another particle and another decay channel, so I had to understand its philosophy and adapt it to the reconstruction of the decay channel of $\Xi_b^-$ we are studying here. The output files of this macro contain lots of information on the reconstructed particles and the associated tracks from which we can extract some physics.

## 4 Data analysis and main results

Data analysis is statistical analysis, meaning we must have enough data to study. I had to produce a sufficient amount of data and to do so, I resorted to parallel computing to get more

12. O2, standing for Online-Offline, is the ALICE software and allows data treatment online (while the detection system is operating) as well as offline (after the recording of the data).

data and faster. I used two computers that ran codes for about a week (in total) so I could generate more than two millions events, so more than two millions $\Xi_b^-$ generated. I also ran codes to simulate the background by generating $\Xi_c^0$ instead of $\Xi_b^-$ and not imposing any decay channel. The same number of events was generated for background and signal data.

The main question is : how much data can we distinguish between signal and background ? When we study a specific decay channel and/or a specific particle, we want to reduce the background noise and to do so, we have to know how topological observables such as the decay radius [13] or the DCA [14] behave with and without background. These topological observables can be used to apply *topological cuts* on the data to constrain the reconstruction process and decrease contaminations from the background.

### 4.1 Decay radius

The decay radius of a particle gives us information on how far away from the center of the detector (that is close to the primary vertex) this particle is at the moment of its decay, so it indirectly gives us information on the background.
Pions are a huge part of the background (around 90% of hadrons produced in a pp collision). Near the primary vertex there are a lot of random pions or not well reconstructed pions and as we go away from the primary vertex, the number of these "parasite" pions diminishes (the background becomes less noisy).
Let's take the example of $\Xi_c^0$ (whose flight distance is about 46 $\mu$m) that we try to associate with a $\pi^-$ to reconstruct the decay vertex of a $\Xi_b^-$. If a $\Xi_c^0$ is primary (i.e. coming from the primary vertex, with a highly noisy background) its decay radius should be shorter than if it is secondary (produced via the decay of a $\Xi_b^-$ in our case, and with less noisy background). Careful, this last statement is only true if the primary and secondary particles have the same $p_T$. Indeed, the decay radius depends on the $p_T$ range we're looking at :

- If a primary particle has a low $p_T$ (compared to its mass, in natural units), a high decay radius (significantly above its flight distance) has a much higher probability to come from the background than from the signal ;
- If a primary particle has a high $p_T$ (compared to its mass, in natural units), a small decay radius (significantly below its flight distance) has a much higher probability to come from the background than from the signal ;
- For secondary particles this is more difficult because the decay radius of the mother particle has to be taken into account (the decay radius of the daughter depends on the decay radius of the mother).

Accessing the values of the decay radius for several ranges of $p_T$ might give us intel on the behaviour of the reconstruction procedure with respect to the background noise and allow us to select intervals of decay radius (for each $p_T$ range) to reduce this noise.

### 4.2 Distance of Closest Approach

There are two kinds of DCA :

13. The decay radius of a particle is the distance seperating the disintegration vertex of this particle and the center of the ALICE detector [8], in the transverse plane.

14. The DCA is the algebraic *Distance of Closest Approach*. It can be evaluated between two particles or between a particle and a vertex. $|\text{DCA}| = \sqrt{\text{DCA}_{xy}^2 + \text{DCA}_z^2}$ where $\text{DCA}_{xy}$ (resp. $\text{DCA}_z$) is the DCA in the transverse plane (resp. the beam direction).

- DCA to primary vertex that must be small for primary particles (we want to find an upper bound of |DCA| above which we consider the particle is not primary) and not too small for secondary particles (we want to find a lower bound of |DCA| below which we consider the particle is not secondary) ;
- DCA between two daughters (we want to find an upper bound of |DCA| above which we consider that two particles don't come from the same vertex).

Plotting the DCA with and without background for the particles we reconstructed shall help us determine these critical values [15] that will constrain the definition domains of primary particles, secondary particles and daughters pairs among our reconstruction process.

### 4.3 Main results

The efficiency of the reconstruction process [16] for $\Xi_b^-$ and $\Xi_c^0$ (coming from $\Xi_b^-$) is shown Figure 3. Note that, at $p_T \to 0$, the efficiency for $\Xi_b^-$ does not tend to zero as the one for $\Xi_c^0$ does. This could be explained by the huge mass of $\Xi_b^-$ (about 5.78 GeV/$c^2$), actually so big that even if the momentum of $\Xi_b^-$ is very low (tending to zero), its decay releases enough energy so that the daughters are not only created (they possess an energy at least equal to their rest mass) but animated with a momentum high enough to detect them.

We can check that the reconstructed particles have coherent mass (i.e. the invariant mass distributions obtained via the reconstruction process are well centered around the theoretical mass of the mother particle) by drawing the Figure 4.

We also study the decay radii of $\Xi_b^-$ and $\Xi_c^0$ distributions for different $p_T$ intervals, Figure 5 :

- From the plot Figure 5a, we notice that we can define a lower bound of the decay radius of $\Xi_b^-$ to a value of about 100 $\mu$m at high $p_T$ ($p_T \in [3; 20]$ GeV/c) as the background is two orders of magnitude higher than the signal below a decay radius of around 100 $\mu$m. This cut should be very good at killing the background, we would still lose signal but we have to make a compromise. At low $p_T$ ($p_T \in [0; 3]$ GeV/c), the difference between signal and background (for a decay radius below 100 $\mu$m) is not as significant but we still can define the same lower bound even if it means losing some signal [17]. We will see that because of the correlations between the DCA to primary vertex and the decay radius, we will choose a cut at a decay radius of 80 $\mu$m for $\Xi_b^-$ ;
- With the same philosophy, thanks to the Figure 5b, we can define a cut for the decay radius of $\Xi_c^0$ around 10 $\mu$m.

Now we are going to study different DCA intervening in the studied decay channel :

- Figure 6 : The distribution of DCA between $\Xi_c^0$ daughters is almost identical for signal and background, this is exepected as the background is produced by generating primary $\Xi_c^0$ that shall undergo the same reconstruction procedure as the secondary $\Xi_c^0$ coming from $\Xi_b^-$. Concerning Figure 6a, the difference between the two distributions is not significant enough to determine a critical value of DCA, if we cut we could lose background but also a great fraction of signal so this is not a good variable to cut on in our case ;
- Figure 7 : Concerning $\Xi_c^0$, it is expected to have a better DCA resolution with background because this background was obtained by generating primary $\Xi_c^0$. These plots show us a

15. The |DCA| to primary vertex upper bound (resp. lower bound) for primary (resp. secondary) particles and the |DCA| between two daughters upper bound.

16. The reconstruction efficiency, for a given type of particle, is the number of reconstructed particles divided by the number of generated particles.

17. A future development might be to make this cut vary with momentum. See the discussion about invariant mass plots for different $p_T$ ranges in 5.

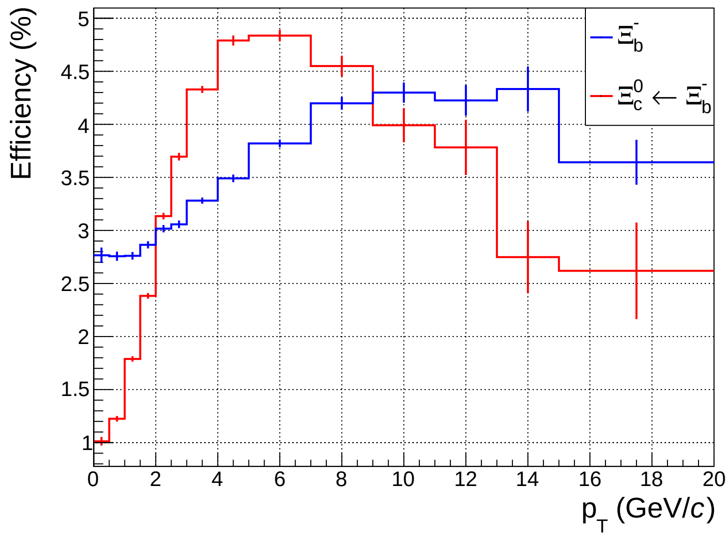



FIGURE 3 – Representation of the reconstruction efficiency as a function of the transverse momentum for $\Xi_b^-$ (in blue) and $\Xi_c^0$ (in red).

clear difference between signal and background and we shall select the $\Xi_b^-$ for $|\text{DCA}_{xy}|$, $|\text{DCA}_z|< 80$ $\mu$m and the $\Xi_c^0$ for $|\text{DCA}_{xy}|$, $|\text{DCA}_z|> 10$ $\mu$m. Note the two peaks around $\text{DCA}_{xy} = \pm 80$ $\mu$m (7a), this shows the correlation between the decay radius (defined in the transverse plane and whose minimum value has been chosen to 80 $\mu$m for this plot) and the DCA in the transverse plane.

The study of the DCA (especially the one to the primary vertex) is quite fruitful because it gives us some critical values that shall help us reducing the background noise.

Other than comparing signal and background data, we can compare the reconstruction process with and without using strangeness tracking (Figure 8). It is the opportunity to highlight the better resolution offered by strangeness tracking with respect to the "standard" method, hence proving the legitimacy of this new technique.

Finally, we arrive to the most important results of this project presented Figure 9. These plots show the consequences of the topological cuts applied to the data.
The Figure 9a shows that the background is highly reduced and the signal is isolated to get a cleaner sample, the cuts killed (almost) all of the background.
There is a strong reduction in the $\Xi_c^0$ mass peak Figure 9b, which is due to the fact that not all of the $\Xi_c^0$ signal comes from $\Xi_b^-$ signal. This is because the background we have considered in this study is comprised of events with a primary rather than a secondary $\Xi_c^0$. When showing these sources separately, we can see that the majority of the $\Xi_c^0$ peak loss is from primary $\Xi_c^0$ loss (79% of loss) and not $\Xi_c^0$ from $\Xi_b^-$ (with a loss of 34%), indicating that we can reduce the background thanks to the cuts applied to the data.
Our main focus shall be on 9a that is revealing all the potential of our analysis.

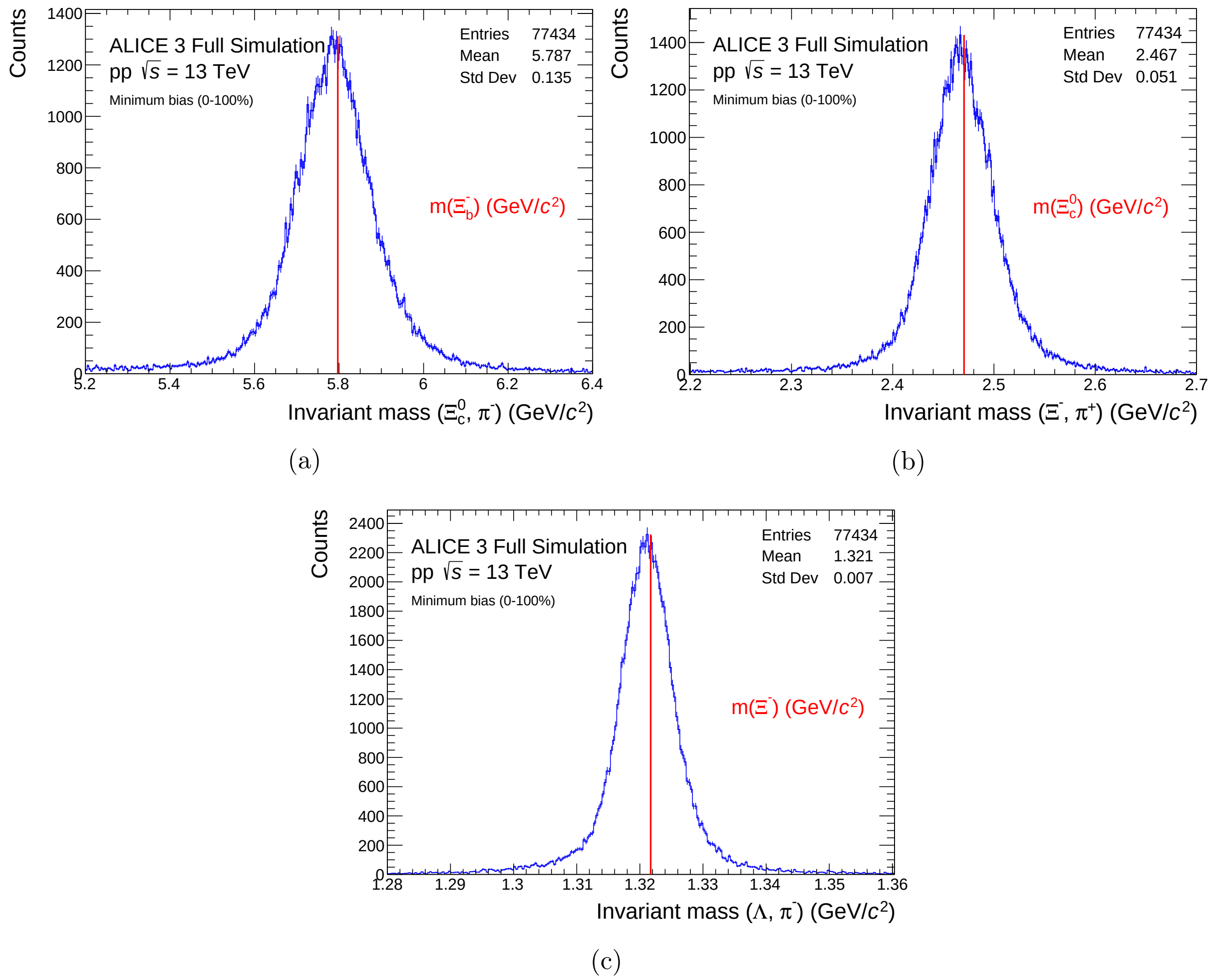



FIGURE 4 – Invariant mass distributions for the reconstruction of $\Xi_b^-$ (4a), $\Xi_c^0$ (4b) and $\Xi^-$ (4c) for signal data. The masses indicated in red are drawn from [7].

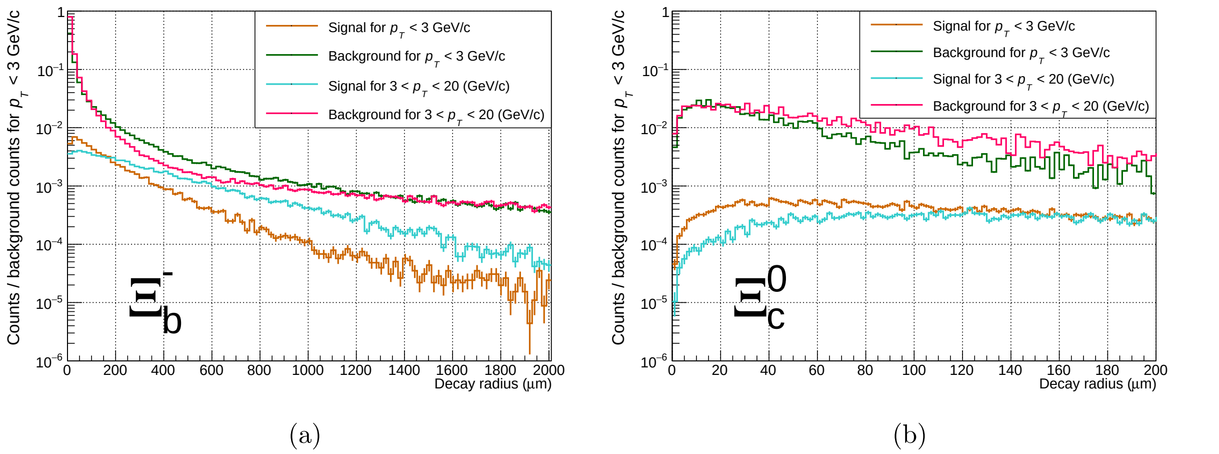
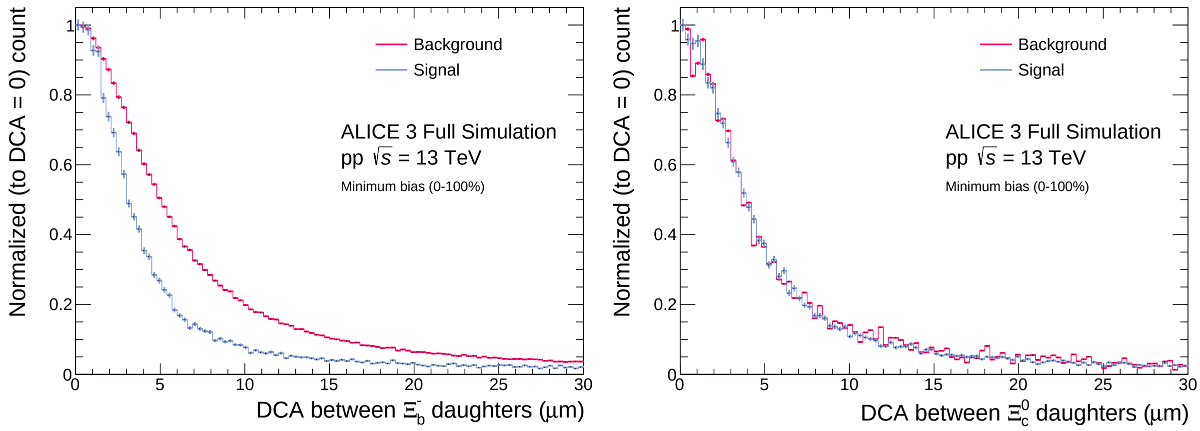
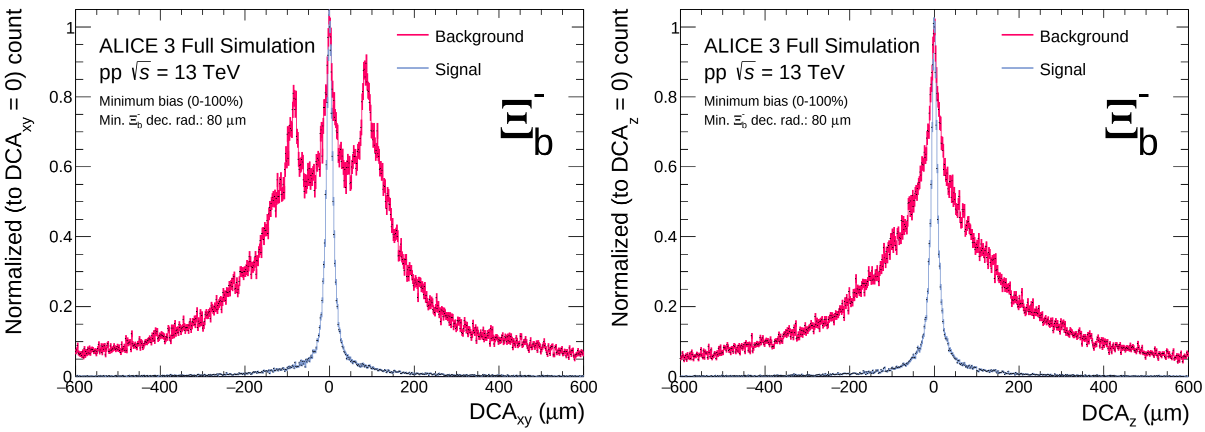



FIGURE 5 – Decay radius of $\Xi_b^-$ (5a) and $\Xi_c^0$ (5b) for signal and background in two different $p_T$ intervals. These $p_T$ intervals are chosen as $[0; 3]$ GeV/c and $[3; 20]$ GeV/c because the $p_T$ distributions of $\Xi_b^-$ and $\Xi_c^0$ (from 0 to 20 GeV/c) are peaked at 3 GeV/c.

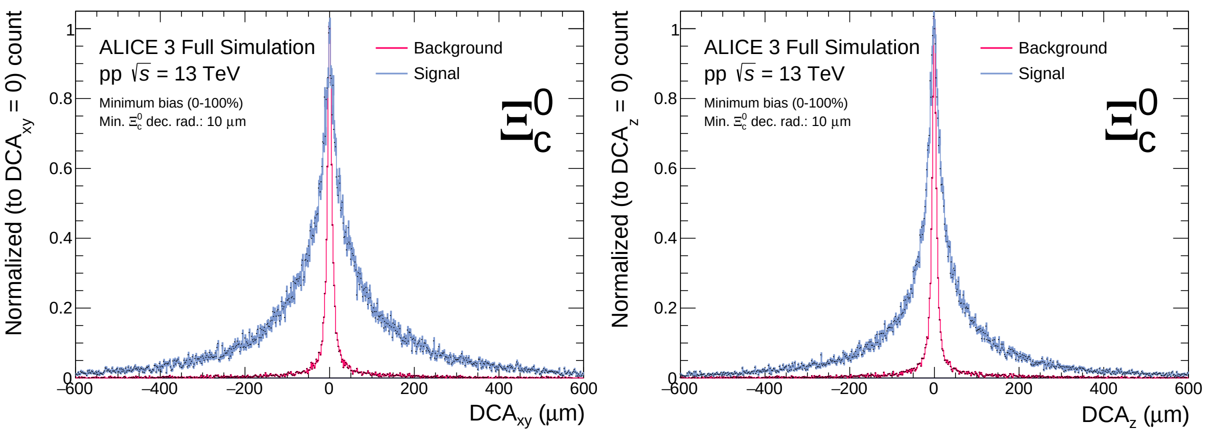



(a) (b)

FIGURE 6 – DCA between the daughters of $\Xi_b^-$ (6a) and $\Xi_c^0$ (6b) for the signal (in blue) and the background (in red).

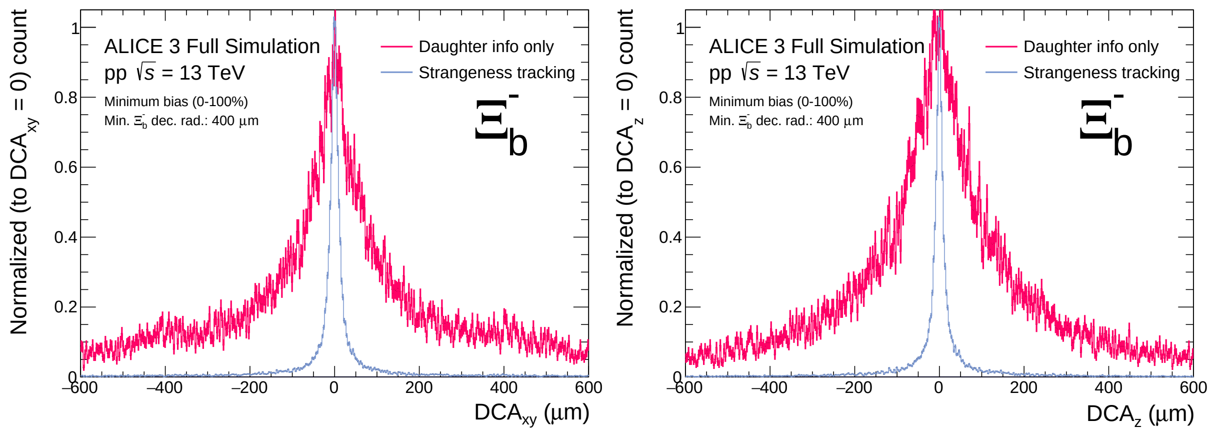



(a) (b)



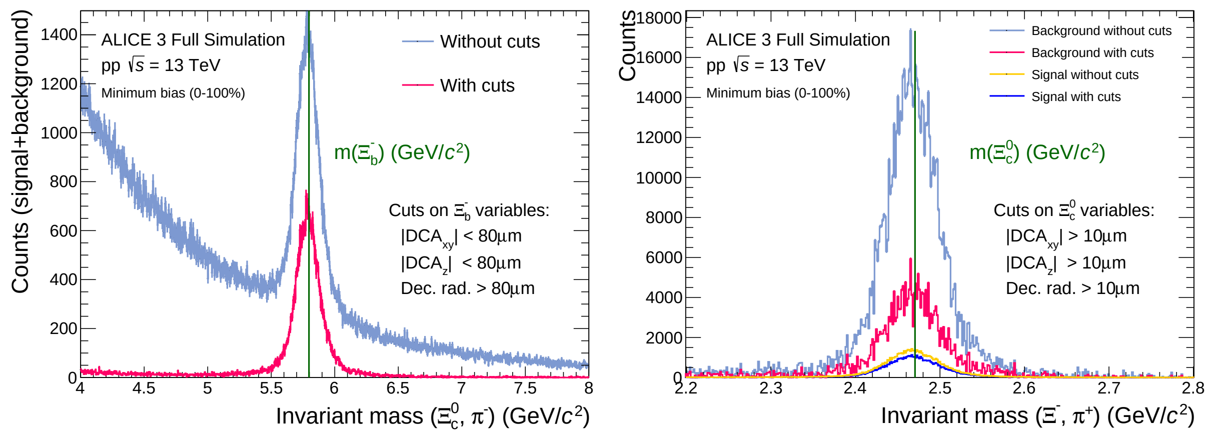

(c) (d)

FIGURE 7 – DCA to the primary vertex of $\Xi_b^-$ (7a and 7b) and $\Xi_c^0$ (7c and 7d) for the signal (in blue) and the background (in red).




(a) (b)

FIGURE 8 – Comparison between topological reconstruction with (in blue) and without (in red) strangeness tracking in the distribution of DCA to primary vertex for $\Xi_b^-$.




(a) (b)

FIGURE 9 – Invariant mass distributions for the reconstruction of $\Xi_b^-$ (9a) and $\Xi_c^0$ (9b).

## 5 Conclusions

I was able to simulate the generation of beauty hadrons ($\Xi_b^-$) in pp collisions at 13 TeV thanks to the Pythia generator, and to force their decay into $\Xi_b^- \to (\Xi_c^0 \to \Xi^-\pi^+)\pi^-$. Then I reconstructed this decay channel using topological reconstruction as well as strangeness tracking. By repeating this generation-reconstruction procedure thousands of times (with 200 events per procedure), I got enough data to perform statistical analysis and study the physics of the particles involved in the decay channel of interest. The Figure 9 (especially 9a) is the final result of this project, it shows that our analysis prototype (simulating a state-of-the-art detector and using the brand new strangeness tracking method) is functional and gives coherent results because we manage to isolate the signal data by working with the mix of signal and background data. Note that the Figure 9a was obtained without any restriction on the transverse momentum $p_T$ of $\Xi_b^-$. The next step would be to simulate even more data and cut the whole $p_T$ range (going from 0 to 20 GeV/c here) in several intervals to deduce the production yield of $\Xi_b^-$ for each one of these intervals. By repeating this whole procedure for $\Xi_b^-$ in pp and Pb-Pb collisions, we should be able to derive the nuclear modification factor of $\Xi_b^-$, that can be defined as follows

$$R_{\text{Pb-Pb}} = \text{normalization factor} \times \frac{\text{Particle yield in Pb-Pb}}{\text{Particle yield in pp}}$$

and that highlights the difference between pp collisions (where no QGP is expected) and Pb-Pb (where QGP appears). This observable could eventually reveal some information about the behaviour of `b` quarks evolving inside the Quark Gluon Plasma.

This project taught me a lot from the coding point of view (hands on Pythia, reconstruction code, O2 environment, ...) but also from the physics point of view (study of the products of high energy collisions, use of momentum conservation and mass invariance, topological observables such as the decay radius of a particle or the DCA, ...) and allowed me to put myself in a researcher's shoes starting to study a problem, overcoming the obstacles encountered to finally being able to extract the physics out of the produced data.

Finally, I would like to thank my supervisors, David DOBRIGKEIT CHINELLATO and Matteo CONCAS, for the many helpful (and cheerful !) discussions we had (wether about physics or informatics) and also for their availability that highly improved the conditions of this fully remote programme and facilitated the development of this project.

## References


[1] B. Abelev *et al* [ALICE Collaboration], J. Phys. G41, 087002 (2014), Technical Design Report for the Upgrade of the ALICE Inner Tracking System

[2] G. Aglieri Rinella [ALICE Collaboration], Nucl. Instrum. Meth. A 845 (2017) 583-587, The ALPIDE pixel sensor chip for the upgrade of the ALICE Inner Tracking System

[3] ALICE Collaboration, Expression of Interest for an ALICE ITS Upgrade in LS3

[4] Longer term LHC schedule

[5] P.A. Zyla *et al.* (Particle Data Group), Prog. Theor. Exp. Phys. 2020, 083C01 (2020) and 2021 update, *Monte Carlo Particle Numbering Scheme*

[6] University of Arizona, *The Particle Data Scheme*

[7] P.A. Zyla *et al.* (Particle Data Group), Prog. Theor. Exp. Phys. 2020, 083C01 (2020) and 2021 update, PDG Live

[8] ALICE Collaboration, *Definition of the ALICE Coordinate System and Basic Rules for Sub-detector Components Numbering*