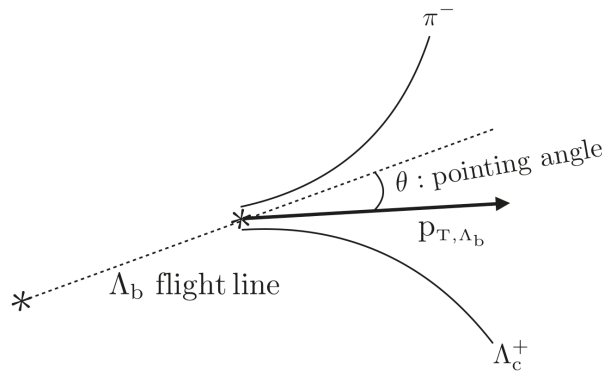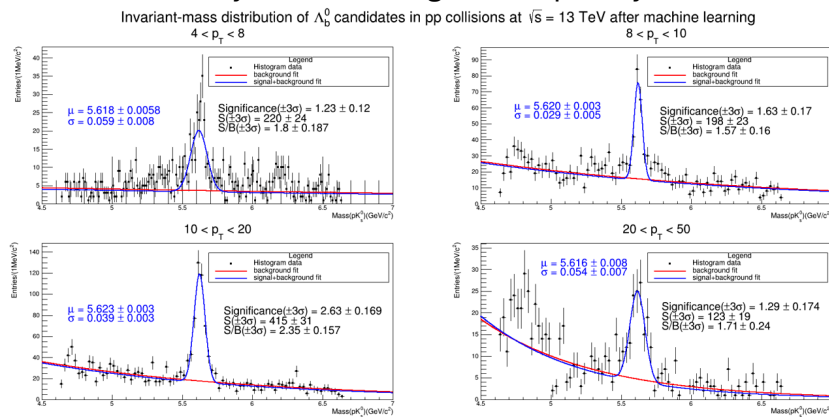## Λb baryon reconstruction

The challenge

Beauty quark is one of the heavy quarks of the Standard Model of particle physics. Its study has important implications in the study of QCD in extreme conditions, as well as in searches for new physics. It is considered an excellent probe for the hot and dense matter produced in heavy-ion collisions, the Quark Gluon Plasma (QGP) [1]. QGP is believed to have been the state of matter that permeated our Early universe a few microseconds after the Big Bang. Due to confinement, in our world we cannot observe a free beauty quark, therefore, to study its properties we need to reconstruct particles that contains it. Among others the Λ_b baryon in a very suitable candidate for charm studies.
In addition to the QGP studies Λ_b baryon, and particularly its ratio with respect the B^0 meson (Baryon-to-meson ratio) are fundamental to fully understand hadronization (i.e how quarks come together to form the particles). The Λ_b baryon is reconstructed by means of invariant mass analysis in the decay channel Λ_b->$\Lambda_c\pi$ (Lambdac + pion) with the Lc then decaying in K^0_s + proton and finally K^0_s decaying in two pions. See left panel of the figure.

Since Λ_c and K^0_s have very short lifetime, the detected particles in this case are the three pions and one proton. However, the Λ_b baryon is rare, we can expect one per several thousands of collisions. On the other side hundreds of pions and protons are produced in each collision. This fact brings to the problem of having an enormous combinatorial background originating from picking uncorrelated protons and pions for the invariant mass calculation that effectively makes the signal completely invisible.

To reduce the combinatorial background selections are made in the topology of the decay (See right panel of the figure and find the detailed explanation of the analysis in **[2]**, chapter 4). Those selections create a 17th dimensional feature space of continuous variables that in addition are partially correlated among each other.

The project

In this project you have the simulated data from the ALICE experiment at CERN LHC. All the topological selections are included both for signal (both Lb and Lc) and background in addition to the invariant mass calculation. At a first investigation of the data, you realize that the invariant mass plot contains only an exponential background without any sign of Lb signal. The scope of the project is to build a learning strategy that allows for efficient classification of signal and background. Note: you might need all the selection variable or only a subsample of them, this is a question you should answer: do we need all the 17th or we can select as efficiently with a lower number? Which one are the most important? A successful project should end with a fit of the invariant mass with a clear gaussian signal on top of the remaining combinatorial background. A plus for the project would be to test two (or more) learning algorithms and compare the outcome in terms of performances and reflect on which one is performing better and why. Note: this project is at the cutting edge of the present research. As you read there are several ALICE groups around the world attempting a successful classification and subsequently detection on real Pb-Pb data.

Bibliography

[1] https://www.ncp.edu.pk/assets/docs/slp-2009/mehnaz-hydrodynamics-001.pdf
[2] https://cds.cern.ch/record/2802084/files/CERN-THESIS-2021-283.pdf