CERN Summer Student Report 2014
June 30th to August 22nd 2014

STUDENT: Nada Ezzelarab*
Egyption Center for Theoretical Physics (ECTP) and ASRT-ENHEP, EGYPT†

SUPERVISOR: Dr. Mateusz Ploskon
CERN - European Organization for Nuclear Research,
ALICE Experiment, 1211 Geneva - SWITZERLAND
(Dated: August 22, 2014)

During my stay at CERN, I have attended lectures mornings and worked on my research project under orienting guidance of my supervisors afternoons. The lectures were informative and pedagogically well-prepared and presented. Their contents was an excellent combination of theoretical and experimental topics in high-energy physics. Furthermore, I was privileged to visit the ALICE, CMS and LHCb detectors and the LINIAC accelerator. I have participated in workshop on "MadGraph software". I was furnished with excellent experiences and cultural exchanges with good colleagues from different countries. I got opportunities to know what the other students have done, in which projects they were involved and how they performed their scientific researches, especially regarding LHC data analysis. For my own project, I have to prove excellent experience with C++ and of course LINUX, ROOT and AliROOT. Tools such as Histograms, Graphs, Fitting, trees and many others were very essential. Furthermore, I am very proud getting a fruitful training in GIANT 3 but I exclusively use PYTHIA8 in order to generate pp collisions at different energies. we have to initialize fitting parameters for thee individual Gaussian fits for $\pi^{\pm}$, $K^{\pm}$ and p(anti-p), Then we integrate over each component in order to extract the number of counts. The results are very convincing. They are nicely comparable with the published measurements.

I. ALICE EXPERIMENT

ALICE (A Large Ion Collider Experiment) is a general-purpose, heavy-ion detector at the CERN LHC which focuses on Quantum Chromodynamics (QCD), the strong-interaction sector of the Standard Model. It is designed to address the physics of strongly interacting matter and the quark-gluon plasma at extreme values of energy density and temperature in nucleus-nucleus collisions. Besides running with Pb ions, the physics programme includes collisions with lighter ions, lower energy running and dedicated proton-nucleus runs. ALICE also takes data with proton beams at the top LHC energy to collect reference data for the heavy-ion programme and to address several QCD topics for which ALICE is complementary to the other LHC detectors. The ALICE detector has been built by a collaboration including currently over 1550 physicists and engineers from 151 Institutes in 37 countries (including my country, institute, colleagues and myself).

II. COMPUTING METHODS IN HIGH-ENERGY PHYSICS

I was privileged to have excellent practice on Unix Shells, File Handling, Pipes, Shell Programming and bash Script. Studying the mechanisms for using `bash script` and the importance for using `bash script` to run the PYTHIA8 simulation code at different energies. In the sections that follow, I will describe the computations tools, especially the programming language C++ from the textbook *Thinking in C++*, volume 1, 2nd Edition. At last, I ll come to the main task; compiling and running main program containing all header files for each class.

III. DATA ANALYSIS IN THE ALICE EXPERIMENT

The process of data analysis would include extracting, compiling, and modelling raw data for purposes of obtaining constructive information that can be applied to formulating conclusions, predicting outcomes or supporting scientific conclusion. The data analysis in the ALICE experiment is possible by ROOT, which is today a tool for their daily work. ROOT is a very appropriate tool for use by actuaries and other insurance analysts who do *ad hoc* data analysis

* A Member of the ALICE Collaboration (ASRT-ENHEP institution)
†Electronic address: `Nada.Ezzelarab@cern.ch`

and predictive modelling type work. It is a framework that is specifically designed for large scale data analysis and stores data in a very efficient way in a hierarchical object-oriented database. This database is machine independent and highly compressed. If one loads a 1 GB text file into a ROOT file, it will take up much less disk space than the original text file. ROOT also has tools to interact with data in a very efficient way. It has built in tools to do multi-dimensional histograms, curve fitting, modelling and simulation. All these tools are designed to handle large volumes of data. Conversely, relational databases (databases where the data is organized as tables and rows) were originally designed for transactional systems and not for data analysis. Thus a relational database is very good for use in a policy administration system, which looks at one policy at a time, or claim administration system, which looks at one claim at a time. But, when one is interested in segmenting the data across all the policies or across all the claims, a relational solution falls apart. In order to make the relational solution work for large scale data analysis, we use the brute force method. A typical brute force method will involve adding considerable computing power, adding sophisticated I/O capabilities such as cache, etc., adding numerous indices to tables, creating additional summaries of the data (like OLAP cubes), and other similar techniques.

A. ROOT/AliROOT Analysis in TPC (two weeks)

My task was distinguishing TPC (Time Projection Chamber) signals for $\pi^{\pm}$, $k^{\pm}$ and p(anti-p) in a given momentum range by using ROOT/AliROOT. In reality they are three Gaussian curves combined together. The concrete task is then to initialize fitting parameters for the three Gaussian fits for $\pi^{\pm}$, $K^{\pm}$ and p(anti-p), individually. Then I have to integrate over each component in order to extract the number of counts.

1. Procedure

We wanted to disentangle the three components from the total distribution and get the yield of each of them, Fig. 1. The pp results are presented as production yields in inelastic collisions. The experimental data are - first - normalized to the number of events passing the event selection. This is converted to inelastic events using the ratio of event-after-event-selection and all-inelastic-events computed in the diffraction analysis. The uncertainty in this correction factor is propagated to the final spectra. In order to perform particle identification, the measured $\delta E/\delta x$ signal in the ALICE TPC must be compared to the expected value of a given particle at a given momentum. This is done via a comparison with a parametrization of the Bethe-Bloch curve. The parametrization for the ALICE-TPC is based on a parametrization which was used for the ALEPH-TPC.

- Step 1: initialize the fitting parameters with three individual Gaussian fits in the corresponding peak region By define the function for each Gaussian
- Step 2: apply the combined fit with the initialized parameters from step 1 and obtain the Gaussian parameters of each individual component
- Step 3, integral over each component (the Gaussian function) and extract the number of counts

2. π, k and p(anti-p) Spectra

The analysis of $\pi^{\pm}$, $k^{\pm}$ and p(anti-p) is based on 5 independent analyses, which focus on individual PID detectors allowing to cover a large range in transverse momentum. The detectors are used with two main techniques referred to as n σ and ($unfolding$). In the n σ approach, a particle is identified as a pion, kaon or proton according to the Bethe-Bloch (or time of flight, depending on the detector used) curve to which the concrete particle is close. In the unfolding method, the response of the detector is fitted with a superposition of Gaussian-like shapes (one for each particle) in bins of p_T. The former approach allows for an easier extraction of the signal, provides track-by-track identification and it is easily automatized, but it is only applicable in regions of clear separation. The latter approach is less sensitive to the description of the response function in the Monte Carlo (whose details are adjusted in the fitting procedure) and allows to extend the identification p_T reach, making it possible on a statistical basis even when the separation is only $\sim 2\sigma$. This method, however, requires a careful tuning of the response function of the detectors and of the fits in the individual p_T bins.

3. The Results

The results for the parameter fitting and the yield of particles are summarized in Tab. I and Tab. II, respectively. Our results seem to point out that

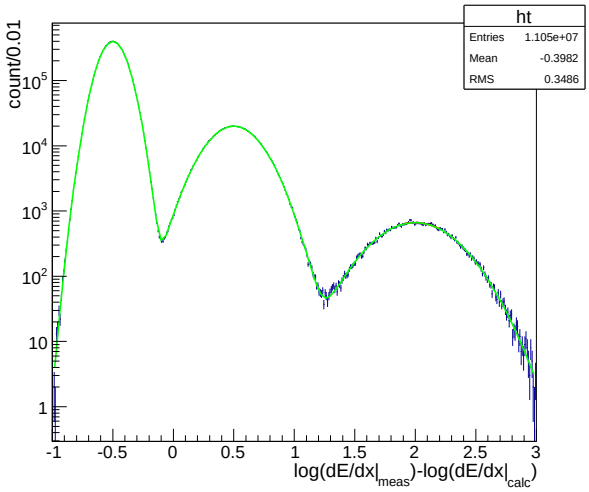


Figure 1: $\pi^{\pm}$, $k^{\pm}$ and p(anti-p) data spectra without the combining fit.

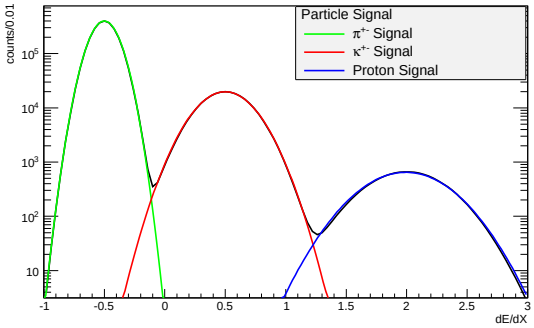


Figure 2: With the combining fit, the green, red and blue curves stand for $\pi^{\pm}$, $k^{\pm}$ and p(anti-p) data spectra, respectively.

- for all elements, the relative errors without the combined fitting (only initialization fitting) are larger than relative errors with the combined fitting with respect to inputs meaning that the combined fitting serviced to improve the experience .
- the relative errors without the combined fitting for $\pi^{\pm} < k^{\pm} < p$(anti-p). This means that the overlap area has an effect on the element with large bin-width.



Figure 3: The red curve stands for the relative error without combined fitting. The blue curve shows the results with for relative error and the combined fitting. The yellow curves represents for relative error between the initializing fitting and the combined fitting.

	Pion		Kaon		Proton	
	Mean	Sigma	Mean	Sigma	Mean	Sigma
The initialization fitting	-5.00042×10^{-1}	1.00023×10^{-1}	5.00660×10^{-1}	2.03569×10^{-1}	1.99515	3.10340×10^{-1}
The combined fitting	-4.99996×10^{-1}	1.00045×10^{-1}	5.00300×10^{-1}	1.99938×10^{-1}	1.99989	2.99482×10^{-1}

Table I: The fitting parameters for the initialization fit to each Gaussian and for the combined fitting or multi fit to the three Gaussian.

	Pion	Kaon	Proton
input yield	10^7	10^6	5×10^4
yield from inetialization fitting step 1	9.9867×10^6	1.01468×10^6	51278.4
yield from combined fitting step 2	9.9912×10^6	1.00061×10^6	50158.8
relative error: step 1 vs input	0.00133	0.01468	0.02557
relative error: step 2 vs input	0.00087	0.00061	0.00317
relative error: step 1 vs step 2	0.00045	0.01406	0.02232

Table II: the yields results for $\pi^{\pm}$, $k^{\pm}$ and p(anti-p) and comparing with the real yield input and extracting some relative error between the result and real input value.

B. PYTHIA8 Simulations for *pp* Collisions (two weeks)

PYTHIA is a Monte Carlo software project that has been developed from high-energy collider physics data in 1978, It is hard-coded to implement specific processes. But it can also be used in conjunction with other event generators to do initial and final state parton showers, multiple interactions, fragmentation and decay. PYTHIA is designed to generate multiparticle production in collisions between elementary particles, e^+e^-, pp and ep, is possible at very high energies. The bulk of PYTHIA multiplicities is formed in jets, i.e. in collimated bunches of hadrons or resonances decaying into further hadrons produced by the hadronization of partons. The relative proportion of strange particles is as expected smaller in comparison with nonstrange hadrons. PYTHIA is capable of simulating for different processes including hard and soft interactions, parton distributions, initial/final-state parton showers, multiple interactions, fragmentation and decay. The measured particle production is conjectured as an indicator for the formation of QGP, especially in heavy-ion collisions. In pp collisions, the spatial and time evolution of the system is too short to assure initial conditions required to drive hadronic matter into partonic QGP. The PYTHIA program is frequently used for event generation in high energy physics. The emphasis is on multi-particle production in collision between elementary particles. Most of the components of the program represent original research. The PYTHIA program can be used to generate high-energy-physics events, i.e. sets of outgoing particles produced in the interactions between two incoming particles. The objective is to provide as accurate as possible a representation of event properties in a wide range of reactions, within and beyond the Standard Model, with emphasis on those where strong interactions play a role, directly or indirectly, and therefore multihadronic final states are produced.

1. My Project

The goals of the first part of my project are as follows.

- Using PYTHIA8 to generate MORE events (something between 10000 and 100000).find the result at different energies (900, 2760, 7000 and 8000 GeV).
- Apply a few particle selection cuts for the two cases:
  - Select only particles which are in our acceptance (within ± 0.8) and
  - Select only charged particles.
- Drawing all histograms in one TLegend.
- The plot should be normalised by the number of events.

C. Results from PYTHIA8

The results from PYTHIA simulation are presented in Fig. 4. It is obvious that:

- The normalization with respect to the number of events improves the results.

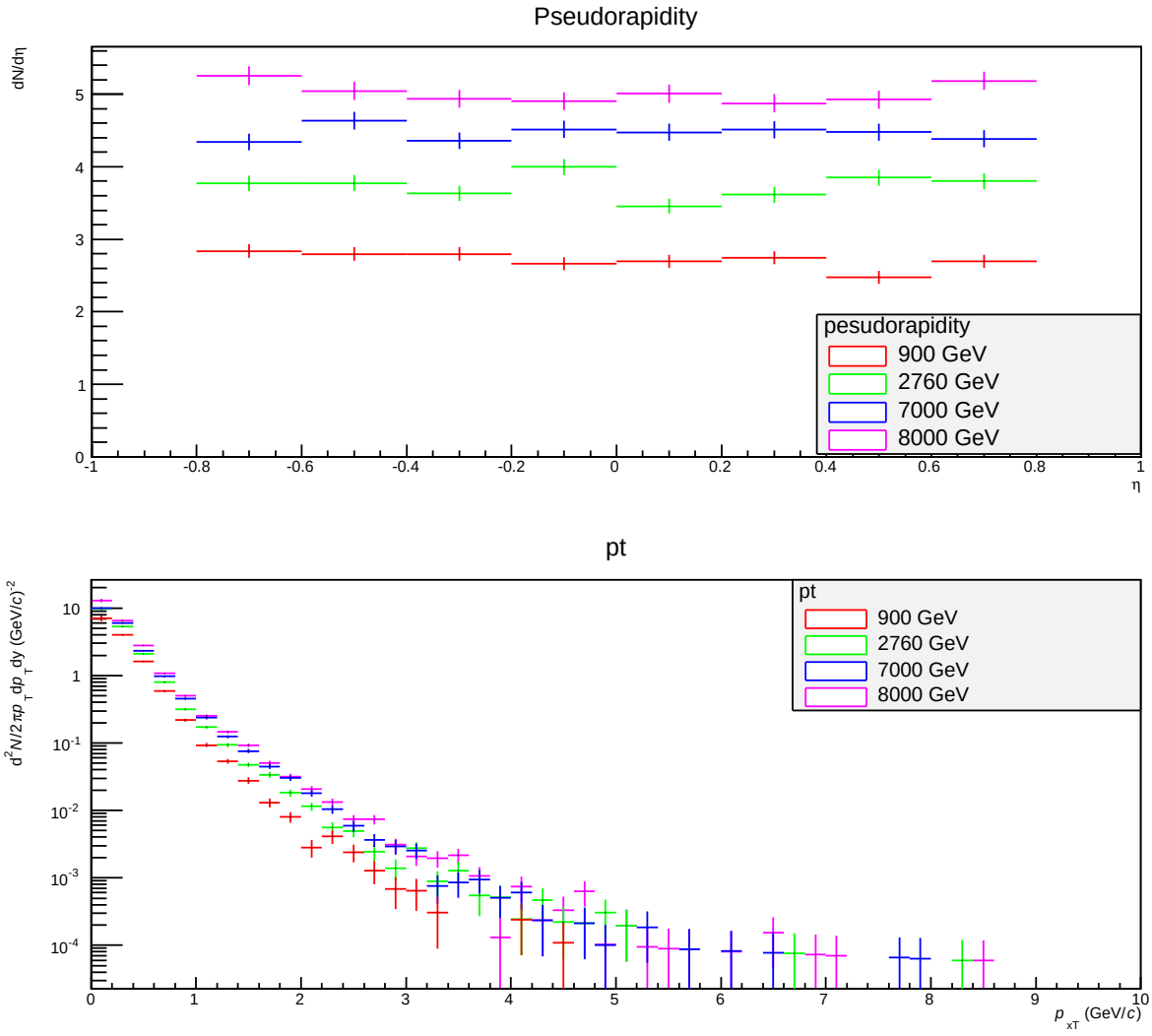


Figure 4: PYTHIA8 simulations for particle multiplicity in pseudo-rapidity and p_T ranges comparable with the ALICE experiment.

- Also the particles momentum should be normalized by the rapidity.
- To compare the result with the experiment results one just needs to normalize with respect to the cross-section of the reaction of interest.
- In PYTHIA, we can use softQCD or hardQCD processes which are different from the experimental processes. They are likely absent in the experimental measurement.
- In light of all these aspects, one should be very careful with such corrections.

IV. OUTLOOK

It is obvious that the present project certainly needs more time. The relatively short summer training did not allow to deeply study various related tasks. I just kicked off the ball and hope to resume in my home institution. Nevertheless, the tools, methods and skills which I gained through this intensive training assure facilitating the long-term engagement with the ALICE experiment.

V. ACKNOWLEDGEMENT

I would like to thanks CERN-NMS and the organizers of the Summer Student Training 2014. My special thank goes to the ALICE experiments for the generous support and excellent facilities. I am very pleasant to express my

sincere grateful to my CERN supervisors Dr. Mateusz Ploskon, Dr. Xiaoming Zhung and Dr. Markus Fasel for their engagement, patience and pedagogical training. Of course, I am very delighted to thank my Egyptian supervisor Professor Abdel Nasser Tawfik for his excellent guidance in the particle physics and offering me great opportunities to exchange with world-class experts.