Introducing HEP to high-school and university students through ATLAS event analysis tools

Dimitris Fassouliotis[1], Christine Kourkoumelis[1], Stylianos Vourakis[1]

[1] Physics Faculty and Institute of Accelerating Systems and Applications, University of Athens Panepistimioupoli, Ilissia 15771, Greece

Abstract. Several EU outreach projects have been running for a few years now and have created a large number of inquiry based educational resources for high-school teachers and students. Their goal is the promotion of science education in schools though new methods built on the inquiry based education techniques, involving large consortia of European partners and implementation of large-scale pilots in a very large number of European schools. Until recently there has been a shortage of educational scenaria addressed to university students to be implemented in the framework of the laboratory courses. Two such scenaria were introduced recently at the National and Kapodistrian University undergraduate labs and are described below.

1 Introduction

The European Union (EU) has long recognized the decreasing interest of young people in science subjects and has been very supportive towards promoting outreach projects which would bridge the gap between school science education and active research.

The authors of this contribution have been involved in several EU outreach projects for more than a decade and have been developing several resources and good practices for use in high schools.

All resources were developed according to the Inquiry Based pedagogical framework, involving the phases of hypothesis, experimentation and conclusions. The authors have been promoting the scenaria in a large number of Greek high schools during half day mini-masterclasses. Up to very recently, there was a lack of more advanced scenaria for the university undergraduate labs.

As a result, the university students were not getting educated in an effective and efficient way in line with the most advanced HEP research and the state-of-the-art experimental and technological achievements. Consequently, the authors launched an effort to develop new laboratory exercises which would permit the undergraduate physics students to interact with the data taken at the world's most powerful accelerator the LHC. The intervention for university students has been designed in the context of the HYPATIA tool created by the authors.

2 Description of the "HYPATIA" data analysis tool

HY.P.A.T.I.A (Hybrid Pupil's Analysis Tool for Interactions in ATLAS) [1], is an event visualization tool that uses real data from the ATLAS experiment, namely offers a graphical representation of the products of proton collisions. The user can interactively analyse the LHC data collected almost in real-time. In addition, the tool has the flexibility which permits the user to perform different analysis tasks: from simple event browsing, to searches for the Higgs boson. The analysis paths are mainly based on virtual inspection of the events and the primary target audience is senior high school students. Two versions of HYPATIA exist and are both widely used. The off-line version [2] which has been used by IPPOG's International Masterclasses since 2008 in the so called "Z-path" [3] by thousands of high-school students in all continents each year. It uses java and has to be downloaded and run locally. The on-line version of HYPATIA [4] was first launched in 2010 by the University of Athens, in an effort to address all devices, including mobile phones and tablets; it is simpler to use, yet all the different searches can be performed. It includes four different parts each with an increasing level of complexity and difficulty. This version has been continuously developed in the context of the EU outreach projects [5-10] and has been used in about 100 Greek schools across the country during the last five years. In 2016 the HYPATIA online version was awarded the GOLC award [11] for the "best visualized experiment".

3 The HYPATIA tool for a university laboratory

There is a general lack of resources for university laboratory exercises which introduce students to state-of-the art technological achievements as well as to the "Big Science" experimental effort pursuit at CERN. The Higgs boson was discovered more than four years ago after a fifty year-long effort and although it attracted huge media coverage and generated great public interest, in most universities worldwide there is not a single experimental exercise to allow the students to experience the discovery themselves.

For the above reasons, the authors decided to introduce activities where the university students would directly connect themselves with the most advanced research done at CERN. These activities are designed in the context of HYPATIA but the university students are asked to perform more complex tasks. Since 2010 the HYPATIA tool has been used in a lab course at the National and Kapodistrian University of Athens (NKUA) for third year undergraduate physics students, majoring in HEP [12]. Each year about 50 students take part in the two three-hour long courses performing all the steps of event analysis mimicking the work of the ATLAS researchers (Figure 1). The first of the three hour lab guides the students to become familiar with the HYPATIA tool, visually identify the signatures which the different leptons leave in the detector and combine decay products to reconstruct the parent particles which in this case are the Z and Higgs bosons. In addition, after the end of the first part of the lab course, they have to hand-in a homework paper where they try to answer advanced questions which are related to the decays modes of the bosons, their widths, masses and experimental resolution and use statistical tools to investigate the significance of discoveries.

The second part of the lab course which has been recently implemented involves the students' analysis of large samples of events both simulated and real ones collected by ATLAS in the 2011-2012 period. ATLAS has released for public use 1fb^{-1} [13] of the Run I data and these are used by the students in the so called "batch" mode [14] of the HYPATIA tool. The main philosophy behind the "batch" analysis is to allow the students to realistically repeat the exact analysis which the ATLAS researches do by optimizing a number of event selection criteria. The goal of this process is to improve the signal to background rejection and eventually make discoveries.

Figure 1. Third year physics students at the NKUA working on the HYPATIA exercises and analyzing ATLAS events.

There are essentially three different analysis paths possible:
a) The optimization of criteria for the Z to two lepton decay in order to enhance the signal over background ratio.
b) The optimization of criteria for the Higgs to four lepton decay in order to make a "discovery" by suppressing the reducible background.
c) Simple analysis and studies of kinematical distributions of real events.

For the first two analysis paths three kinds of samples are required: simulated signal, simulated background and real events. For the third path -which is the simpler- mainly real events should be used.
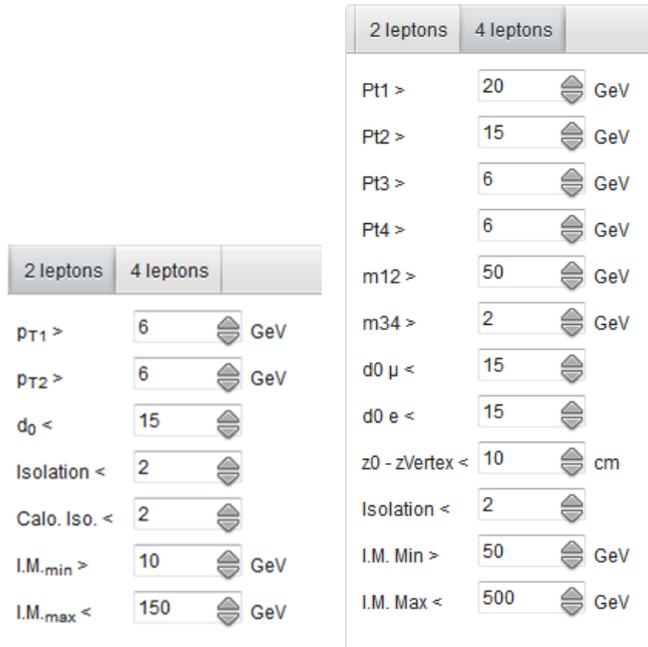
Figure 2. The sets of available optimization criteria. Left for the Z→two lepton decay and right for the Higgs→four lepton decay.

Below only the first two paths will be described in detail. In both of these a group of criteria is given using a simple GUI (Figure 2) and the students can optimize each one separately by inspecting three plots provided for each criterion: the simulated signal distribution as a function of the value of the criterion, the simulated background distribution and the value of the significance as a function of the criterion value. The student is guided to choose as an optimization value, the value where the

significance $\sqrt{2*\left((S+B)*\ln\left(1+\dfrac{S}{B}\right)-S\right)}$ is maximized and then proceed to the next criterion.

The procedure should be repeated a few times for each criterion. The starting point of each iteration should be the most recent values of all criteria while optimizing one at a time. Figure 3 shows the optimization of the isolation criterion for the four lepton sample. The "isolation" provides an indication of whether or not there are tracks (or energy for "calo isolation") accompanying the leptons of the event. Leptons originating from the Z or Higgs events tend to be isolated whereas leptons from $b\acute{b}$ semileptonic decays -which are part of the reducible background- are as they are typically parts of a particle jet.
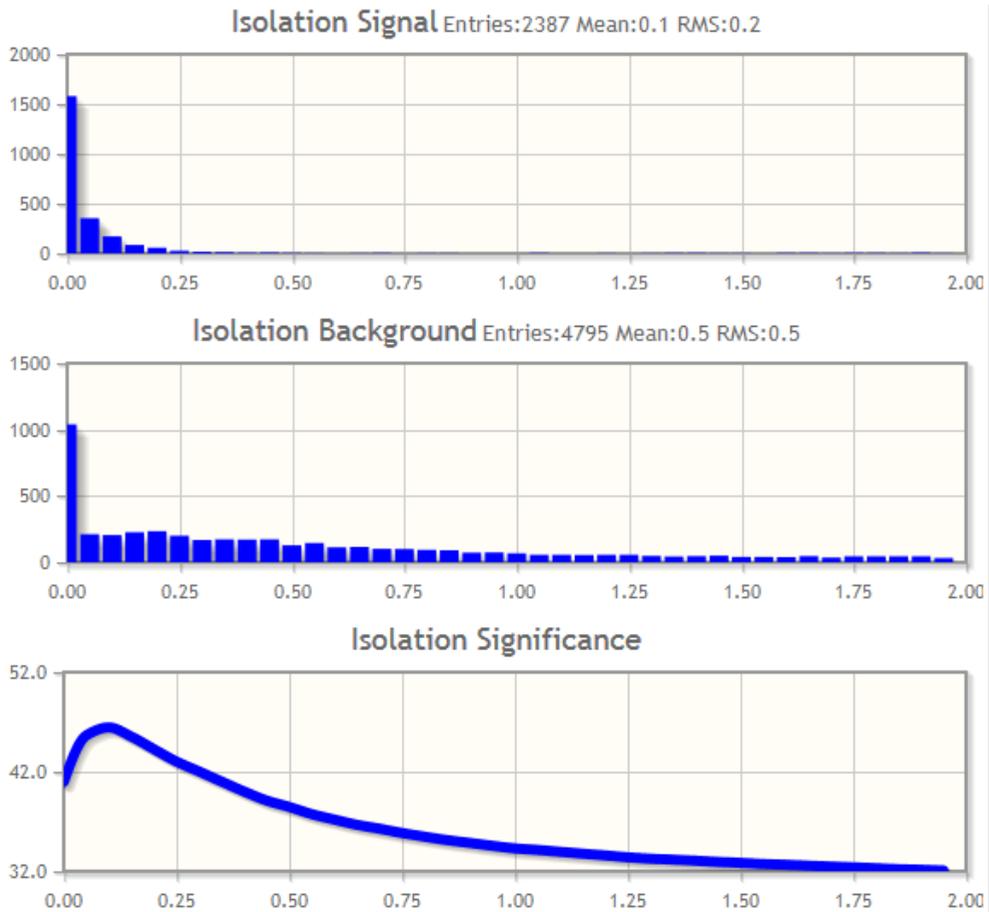
Figure 3. Histogram of the isolation criterion (for definition see above) .The top plot is the simulated signal, in the middle the simulated background and at the bottom the significance (for definition see above) as a function of the criterion value.

The criteria which are given to the students to optimize were chosen as close as possible to the ones which the ATLAS researches use in the analysis. These are (as shown on Figure 2)

a) For the two lepton analysis: the momenta of the pre-selected two leptons $P_{T1,2}$, the largest impact parameter of the lepton in the event (divided by its error) and the largest isolation parameter of the leptons in the event (both for accompanying tracks and energy). The impact parameter gives an indication of how far a secondary vertex is (if there is one) from the primary vertex of the event. Background events -especially in the case of jets from b quarks- tend to have well separated secondary vertexes.

b) For the four lepton analysis: The momenta of all pre-selected four leptons $P_{T1,2,3,4}$, the mass of the lepton pair (opposite sign, same flavor) closest in value to the Z mass (m_{12}), the mass of the next lepton pair (opposite sign, same flavor) closest in value to the Z mass (m_{34}), the largest impact parameter of the leptons in the event divided by its error (separately for muons and electrons) and the largest isolation parameter of the leptons in the event.

During the optimization iterations, the students can compare histograms of the invariant mass of the signal they are after -two leptons or four leptons- with the background before and after the application of the criteria. They can also compare the combined simulated signal plus background with the real data. Their ultimate goal -after they define the optimum set of cuts- is to get the simulated combined

data after the application of the optimized criteria to look as close as possible to the histograms of real data after the application of the criteria which ATLAS researchers defined after their in depth analysis. For the Z analysis, the data released by the ATLAS experiment are enough for this comparison. In the Higgs case, since the Higgs published signal for Run I contains only a few tens of events [15] and the data released correspond to 1/25 of the whole sample, this is not possible (Figure 4). Instead the students are asked to compare their combined simulated four lepton histogram of Figure 4 with the ATLAS publication [15].
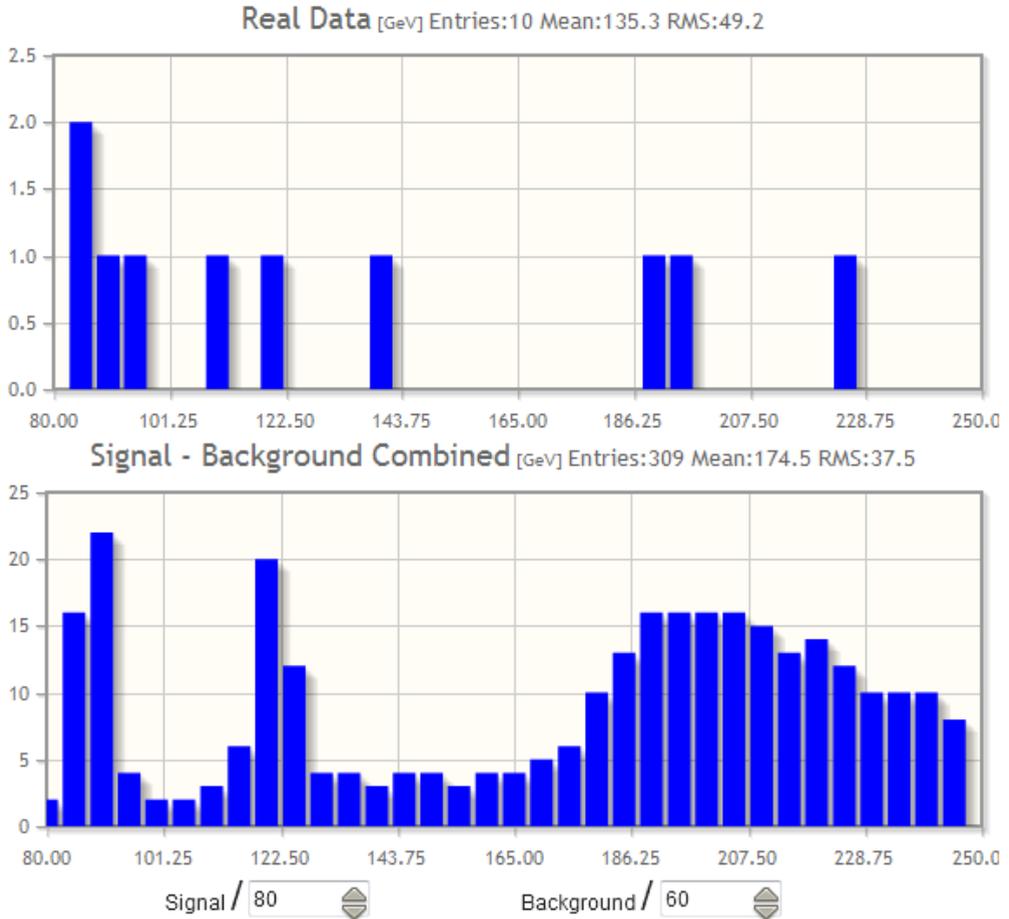
Figure 4. Histogram of four lepton invariant mass after application of the optimum criteria. Top: real data corresponding to 1 fb^{-1}. Bottom: signal plus background both simulated and combined according to their cross sections.

During the academic years 2015-2016 and 2016-2017 about 100 students in total have followed the recently designed lab course with great success. The part which includes the real ATLAS data was added only a few months ago and has been tried out by both the NKUA as well as very recently by the University of Birmingham students as the hands-on exercise of their 4th year Experimental Particle Physics Techniques course.

4 Conclusions

A two part university laboratory course has been developed which involves the analysis and selection criteria optimization of ATLAS events in a way which mimics the experiment's analysis. This way the undergraduate students are exposed to top-level research and grow a real enthusiasm for it.

The exercise is based on the HYPATIA event visualization tool and detailed instructions are available online [16]. It is also listed as an external resource [17] of the ATLAS opendata portal. The exercises have been presented in multiple outreach meetings and have attracted the interest of several universities.

5 Acknowledgements

6 References

1. Kourkoumelis, C., & Vourakis, S. (2014). Hypatia-An online tool for ATLAS event visualization. Physics Education, 49, 21-32
2. [Online] http://hypatia.phys.uoa.gr/
3. [Online] http://atlas.physicsmasterclasses.org/en/zpath.htm
4. [Online] http://hypatia.iasa.gr/
5. [Online] http://www.learningwithATLAS-portal.eu.
6. [Online] http://www.pathway-project.eu/
7. [Online] http://www.discoverthecosmos.eu/
8. [Online] http://www.go-lab-project.eu/
9. [Online] http://inspiringscience.eu/
10. [Online] http://creations-project.eu/
11. Global Online Laboratory Consortium: [Online] http://online-engineering.org/GOLC_about.php
12. D. Fassouliotis, C. Kourkoumelis, S.Vourakis, The Inspiring Science Education project and the resources for HEP analysis by university students, 4th ICNFP Kolympari 2015,EPJ Web of Conferences 126, 02017 (2016),http://dx.doi.org/10.1051/epjconf/201612602017
13. [Online] http://opendata.atlas.cern/
14. [Online] http://hypatia.iasa.gr/en/batch.html
15. G.Aad et al (ATLAS collaboration) http://dx.doi.org/10.1016/j.physletb.2013.08.010
16. [Online] http://hypatia.iasa.gr/en/Exercise_6_eng.pdf
17. [Online] http://opendata.atlas.cern/externals/