ATLAS software tutorial: *evolving to accelerate integration*

Unleashing the Power of Integration: A Poster Showcasing Collaborative Efforts that Accelerated the Development of Software Tutorial for a Full LeptoQuark Analysis from Scratch

Abstract: The ATLAS software tutorial [1] is a centrally organized suite of educational materials that helps to prepare newcomers for work in ATLAS. The broad objective is to familiarize participants with the basic skills needed to accomplish data analysis tasks, with a strong focus on software tools. In this poster, we will outline the recent changes to the ATLAS software tutorial to follow a project-based pedagogy. The tutorial walks through all the key stages of a published ATLAS analysis, including topics from Monte Carlo generation, analysis optimization techniques, usage of the computing grid and statistics. We will also discuss some of the preliminary feedback from the pilot program and early trials of the new tutorial format.

Introduction

- The ATLAS collaboration has offered specialized training for various aspects of the available software and workflows since 2004, with a special focus on the tools for offline physics analyses
- Regular centralized training events are organized since 2008 for early Ph.D students and for senior individuals who are new to the collaboration
- Since its inception, the training curriculum has evolved to be **increasingly pedagogical with a focus on the practical applications for physics analyses**
- This poster highlights the latest developments in the training tutorial with an emphasis on "How to do an end-to-end ATLAS physics analysis"

How to do ATLAS analysis from start to finish – a hands on tutorial.

10–12 Oct 2022
SLAC
US/Pacific timezone

Enter your search term

Overview
Scientific Programme
Timetable
Registration
Code of conduct
Travel logistics
Pre-tutorial account setup
Introductory HEP material
Arrival at SLAC

Ben Carlson
bcarlson@cern.ch

The workshop will aim to teach basic ATLAS analysis tools by reproducing the second-generation leptoquark search. Introductory material for physics principles will be given, followed by hands-on activities to implement ATLAS methods yourself.

After the first day, we will break into small groups of about 5 participants per group, each with a tutor.

Pre-exercises: before the workshop we will require that all participants complete some setup work. Details will be given in September, and include items such as setup of lxplus, SLAC computing account, obtain a grid certificate and more.

Target audience: primarily we are designing this tutorial for graduate students transitioning from classes to research. We welcome applications from all ATLAS researchers, including new postdocs, and even faculty who want to learn how to use ATLAS tools.

Can the tutorial be followed virtually?

While we strongly encourage in-person attendance, we will make some of the introductory lectures available by Zoom. The hands-on notebooks will be possible to follow remotely. However, a large part of the tutorial will be in breakout sessions, which will be difficult to follow remotely.

New Structure

- Interactive, hands-on and project-based structure with the aim of conducting end-to-end physics analysis
- Months of collective expertise and effort went into designing this new structure, with a strong focus on re-using existing materials
- First test of new format in 3 days at SLAC (October 10-12, 2022)
- Adopted the **2nd generation LeptoQuark (LQ) analysis** due to its importance and simplicity
- Students worked in small groups (less than 5), with a tutor assigned to each group
- Relied on many experts in the Bay area

Learning Objectives

- Generate signal events using MadGraph
- Retrieve trigger decisions, electrons, muons, jets and E_T^{miss} from DAOD-format samples and submit grid jobs
- Create a histogram of the Z-peak from the di-electron events
- Perform a cut-based optimization for the LQ analysis
- Perform Data/MC comparisons in control samples
- Train a boosted-decision-tree (BDT) to classify signal and background events
- Understand the systematic uncertainties and compute them using a minimized example
- Statistically interpret the results as a limit and as a discovery (with injected signal)

Team Building

- Assigned small groups of 4-6 people to work through the tutorial together
- Included participants from different institutions
- Different backgrounds: undergraduate, graduate, postdocs
- Worked together for the entire tutorial

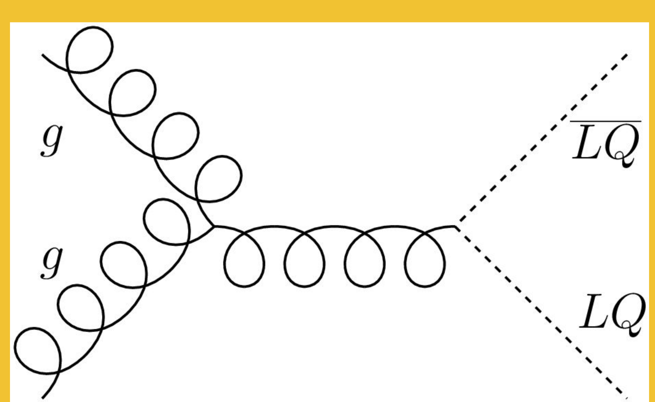
Hands-On Component With Jupyter Notebooks

- Guided philosophy that participants should be evaluating each component of the analysis on their own rather than just listening
- Some components (e.g. ATLAS software) involved C++ developments within ATLAS
- Wherever possible, Jupyter notebooks were used to illustrate the **flow** of doing a physics analysis

```
In [13]:
# You should only see the Electron object we defined earlier in the schema.
# This object lets us easily access information about the electrons without
# worrying about all of the individual branches we saw earlier.

for i in range(10): # Loop through the first 10 events
    if len(events['Electron'][i])>0: # Check if there are electrons in the event
        print("Event {}:".format(i))
        # Print the electron transverse momentum, mass, and charge for each event
        print("  Electron pT (in GeV): ", events['Electron'][i].pt/GeV)
        print("  Electron mass (in MeV): ", events['Electron'][i].mass)
        print("  Electron charge: ", events['Electron'][i].charge)

Out [13]:
Event 2:
  Electron pT (in GeV):  [57.9, 53.8]
  Electron mass (in MeV):  [0.511, 0.511]
  Electron charge:  [1, -1]
Event 3:
  Electron pT (in GeV):  [47.7, 50.7]
  Electron mass (in MeV):  [0.511, 0.511]
  Electron charge:  [-1, 1]
```



Feynman diagram for LQ pair production [3]

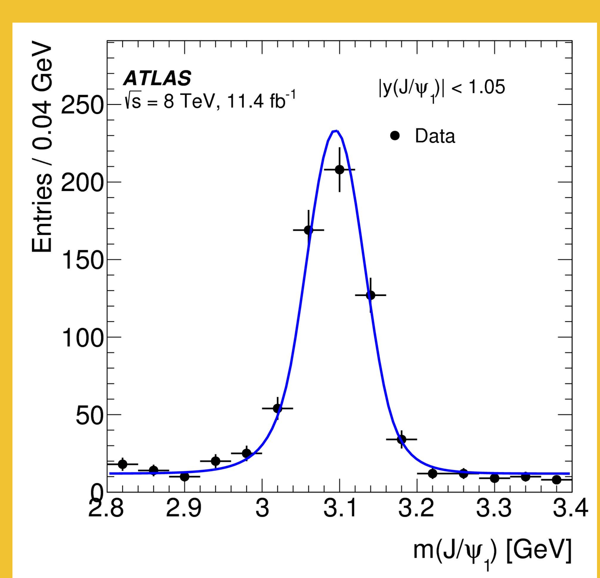


Illustration of reproduced invariant mass distribution taken from published ATLAS work [2]

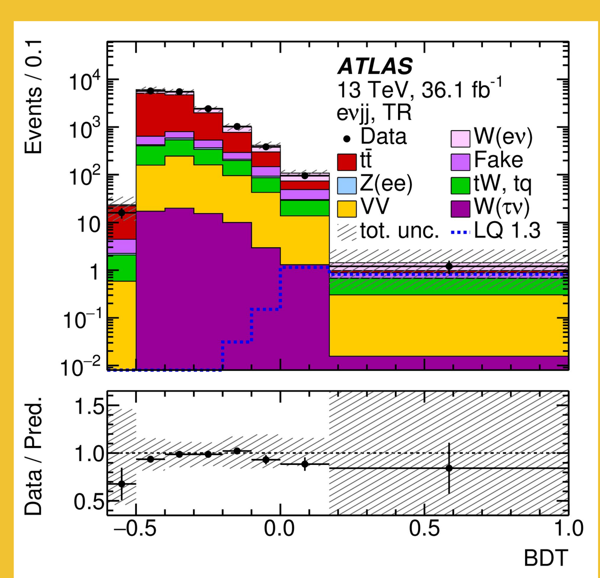


Illustration of reproduced BDT distribution taken from published ATLAS work [3]

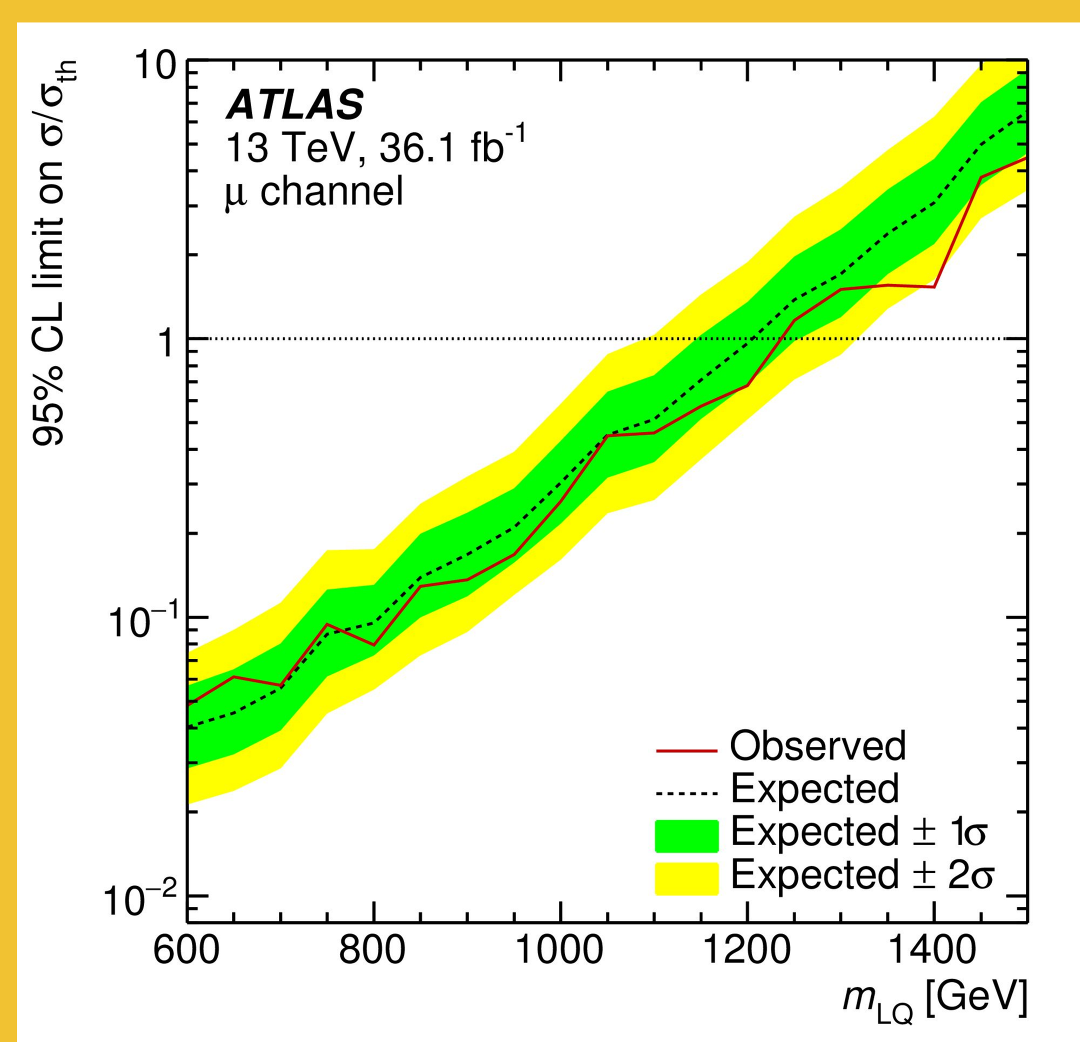


Illustration of reproduced upper limit distribution taken from published ATLAS work [3]

Day 1

- Event generation of LQ signal
- Accessing the trigger and other object information
- Accessing the event content (AOD) and using the grid to generate ntuples

Day 2

- Plotting the Z-peak
- LQ analysis optimization
- Systematic uncertainties

Day 3

- Boosted decision trees (BDT) optimization
- Introduction to statistical tools
- Statistical interpretation of results

References

[1] https://atlassoftwaredocs.web.cern.ch/ASWTutorial/
[2] https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/BPHY-2015-02/
[3] https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/EXOT-2016-08/

Rocky Bala Garg, Ben Carlson, Jason Veatch, Rachel Hyneman, Rachel Smith, Elham Khoda, Miha Muskinja, Alexander Held, Amber Roepe-Gier, Nils Krumnack
(On behalf of the ATLAS Computing Activity)

CHEP, May 2023