

Boosted particle tagging at the LHC based on large R jet substructure using machine learning techniques.

Carlos Fernando Buitrago Cárdenas

March 22, 2023

Abstract

BEGIN ABSTRACT

TABLE OF CONTENTS

1	INTRODUCTION	2
2	THEORETICAL ASPECTS	3
2.1	JETS AND ALGORITHMS	3
2.2	BOOSTED PARTICLES AND FAT JETS	5
2.3	JET TAGGING	6
2.4	FAT JET SUBSTRUCTURE: NSUBJETTINESS	6
2.5	MACHINE LEARNING AND TMVA CLASSIFIERS	6
2.6	THE LHC AND THE ATLAS DETECTOR	6
2.7	EVENT GENERATORS	9
3	EVENT GENERATION	10
3.1	EVENT GENERATION CHAIN	10
3.2	SIGNAL AND BACKGROUND EVENTS	10

Introduction

Theoretical aspects

2.1 JETS AND ALGORITHMS

After being produced in a high-energy event, quarks and gluons fragment and hadronize resulting in a collimated spray of hadrons called a jet. The reason behind the process of hadronization lies in the concept of colour confinement. In quantum chromodynamics (QCD), colour confinement states that only objects with non zero colour charge can propagate as free particles, therefore quarks and gluons are only seen bound together in the form of hadrons. When particles carrying colour charge (namely quarks and gluons) are separated in a high-energy event, new colour carrying particles are spontaneously created from the vacuum in order to form colourless hadrons, thus obeying confinement.

While hadronization is not yet fully understood and a theoretical description of the process is not yet available, there is a number of phenomenological models such as the Lund String Model that do a good job of describing it [1]. The phenomenon can be understood qualitatively through these models by taking into account that the gluon field between colour charges becomes a narrow flux tube as they get separated and eventually it becomes energetically favourable for a new particle to appear rather than extending the tube further, as can be seen in figure 2.1.

The particles resulting from the hadronization of a single quark (parton) tend to travel in

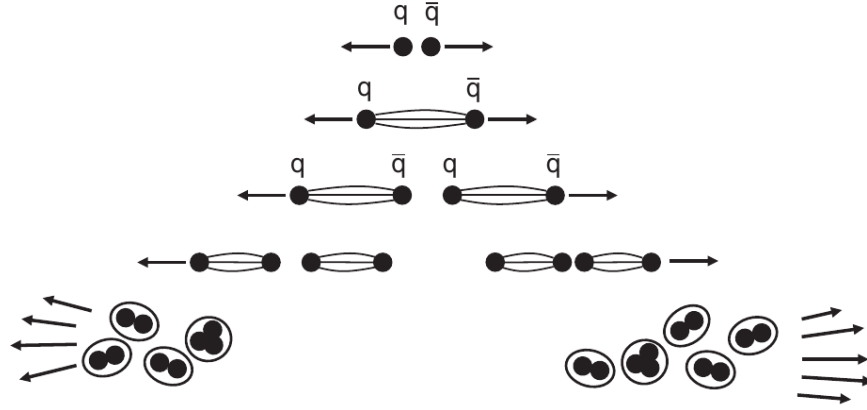


Figure 2.1: Schematic representation of the hadronization process. Picture taken from [2].

the same direction as their parent forming a narrow cone, which is what is known as a jet. In particle detector experiments jets are observed instead of quarks, and their structure is quite visible when looking at the events reconstructed in the detector. Through the measurement of the different properties of a jet it is possible to obtain information about the original parton that originated it. Therefore jets are an essential part of the analyses carried out in collider experiments.

As important as they are, to be used effectively in analyses jets need to be well defined. As stated in the work of Salam [3] a jet definition is constituted by a jet algorithm with its respective parameters and recombination scheme. A jet algorithm is a set of rules that group particles into jets. These algorithms usually involve parameters that govern their behaviour, for example in defining how close two particles need to be in order for them to be considered part of the same jet. Jet algorithms are also related to a certain recombination scheme, which indicates how the momentum is assigned to the object resulting of merging two particles during a clustering process.

Jet algorithms can be usually classified into two broad categories: cone algorithms and sequential recombination algorithms. Cone algorithms originate from the initial idea of Sterman and Weinberg [4]. They are considered "top-down" algorithms, since they group together particles within specific conical angular regions so that the resulting cone is "stable",

meaning that the direction of the cone matches that of the 4-momenta sum of the particles. On the other hand, sequential recombination algorithms are considered "bottom-up", as they iteratively recombine nearby particles in accordance to a certain distance measure.

The anti- k_t algorithm [5] is a sequential recombination algorithm widely used in collider experiments, being also the preferred jet identification algorithm in ATLAS analyses. Since the present study is based on simulated data from this detector (see section 3.1) it is relevant to give a brief overview of this algorithm. The anti- k_t takes as an input a list of N objects and it returns a list of jets, which correspond to clusters of said objects grouped according to specific rules regarding distances between them. The distances used by the algorithm are calculated from the quantities k_{tX} , η_X and ϕ_X which correspond to the transverse momentum, pseudo-rapidity and azimuthal angle of the object X . These distances are d_{ij} (the distance between objects i and j) and d_{iB} (the distance between the object i and the beam), they are defined as follows:

$$d_{ij} = \min(k_{ti}^{-2}, k_{tj}^{-2}) \frac{\Delta_{ij}^2}{R^2} \quad (2.1)$$

$$d_{iB} = k_{ti}^{-2} \quad (2.2)$$

where $\Delta_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$ and R is the radius parameter that sets the size scale of the jets found. The algorithm iteratively forms clusters by identifying the smallest of the two distances for all the objects in the input list. If d_{ij} is the smallest, objects i and j are recombined and replaced in the object list by the recombined object. If on the other hand d_{iB} is the smallest, object i is removed from the object list and marked as a jet.

2.2 BOOSTED PARTICLES AND FAT JETS

Boosted particles are ...

2.3 JET TAGGING

2.4 FAT JET SUBSTRUCTURE: NSUBJETTINESS

2.5 MACHINE LEARNING AND TMVA CLASSIFIERS

2.6 THE LHC AND THE ATLAS DETECTOR

The Large Hadron Collider (LHC) [6] is the world's largest particle accelerator. It was built by the European Organization for Nuclear Research (CERN) in the existing 26.7 km tunnel that was constructed originally for the Large Electron-Positron Collider (LEP), which lies between 50 m and 175 m below the surface beneath the France-Switzerland border. The 3.8 m wide tunnel contains two adjacent beamlines, which allow two high-energy particle beams to travel in opposite directions around the accelerator ring. The beams are guided by about 10.000 superconducting magnets [7], dipole magnets are used to bend the beams, quadrupole magnets are used to focus them and magnets of higher multipole orders are used for smaller corrections in the geometry of the field.

The maximum energy that can be reached by the protons in the beam is limited by the peak dipole field, which has a nominal value of 8.33 T (corresponding to a proton energy of 7 TeV) [8]. However, the actual value of the attainable field depends on external factors that cause beam losses. As such, the highest proton energy achieved as of today is 6.8 TeV, which in turn means an attained center of mass energy of $\sqrt{s}=13$ TeV. In order to achieve said energy, before being injected to the LHC the protons are pre-accelerated in a series of steps in which their energy is successively increased (see figure 2.2). Initially, H^- ions with an energy of 160 MeV are produced in the linear accelerator LINAC4, which feeds the Proton Synchrotron Booster (PSB), where the electrons are stripped from the ions and

the remaining protons are accelerated to 2 GeV and injected into the Proton Synchrotron (PS), which accelerates them to 26 GeV and injects them into the Super Proton Synchrotron (SPS) where they are accelerated to 450 GeV before being finally injected into the main ring.

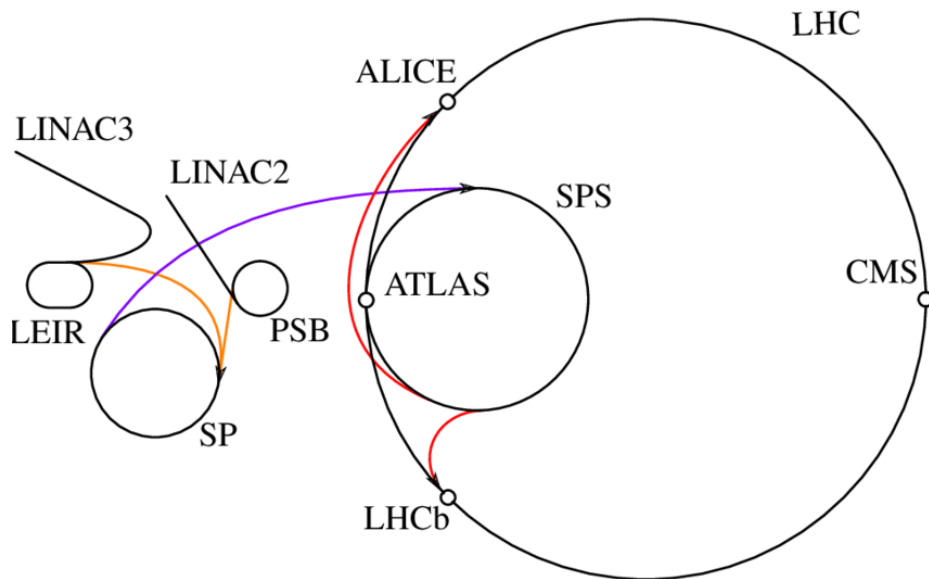


Figure 2.2: Schematic representation of the LHC and its pre-accelerators (in 2020 LINAC2 was replaced by LINAC4). Image taken from [9].

It is also worth noting that the protons do not travel in the form of continuous beams, but rather in bunches of 10^{11} protons each. Under nominal operating conditions each proton beam is composed of 2808 bunches, so that interactions take place at discrete intervals each 25 ns (collision rate of 40 Hz). Taking into account the previous parameters the design (proton-proton) luminosity of the LHC is $10^{34} \text{ cm}^{-2}\text{s}^{-1}$.

The two beams intersect at four different points around the accelerator ring, which is where the collisions occur. Specially strong magnets are used near these points in order to increase the interaction chance. Built around the collision points are seven experiments installed in underground caverns [10]. Of these, the main four are the ATLAS, CMS, LHCb and ALICE detectors. These detectors are used to count, track and characterize all the particles that are produced in the collisions in order to reconstruct the different events as a whole. As was stated previously, this study will be carried out using simulated data from

ATLAS, therefore an overview of the detector will be given below.

The ATLAS (A Toroidal LHC ApparatuS) detector [11] is a multi-purpose detector designed to cover a wide range of physics at the LHC. The detector is forward-backward symmetric with respect to the interaction point, and it has nearly full coverage in solid angle. In order to reconstruct the particles originated from the collisions, ATLAS is composed of multiple layers, each one of them sensitive to different types of particles. The detector layout is shown in figure 2.3.

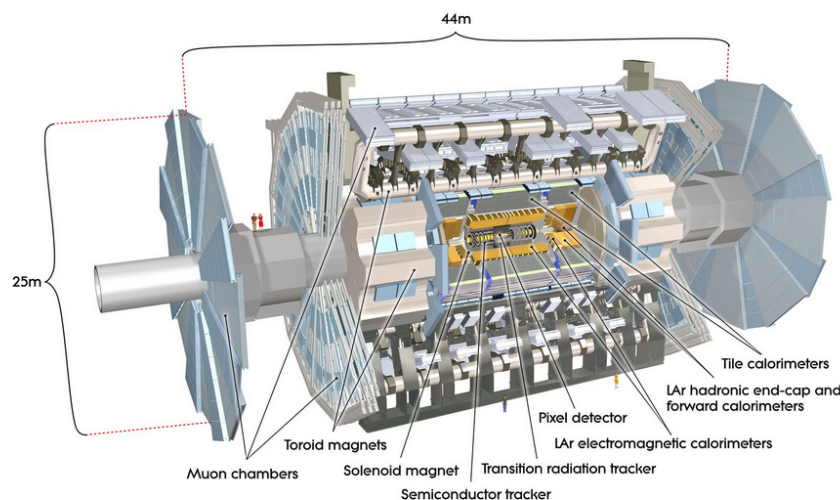


Figure 2.3: Cut-away view of the ATLAS detector. Image taken from [11].

The layers [12] that are closer to the interaction point form what is known as the inner detector, they are immersed in a 2 T magnetic field generated by a solenoid magnet. Charged particles are curved by this field and their trajectories (tracks) are measured by silicon pixel and microstrip detectors, which are surrounded by a transition radiation tracker. Surrounding the inner detector is the calorimeter system, composed by the liquid argon (LAr) electromagnetic calorimeters and the scintillator-tile hadronic calorimeters. LAr technology is also used in the hadronic end-caps (matching the outer limits of the electromagnetic calorimeters) and forward calorimeters, which extend the detection coverage. Both the electromagnetic and hadronic calorimeters use the same principle: when a particle enters the

calorimeter, it showers and deposits energy in the calorimeter cells. The difference between the two kind of calorimeters lies in the materials they are made of and the size of the cells, which determine the particles they target (electromagnetic calorimeters targeting electrons and photons and hadronic calorimeters targeting hadrons). Surrounding the calorimeter system is the muon spectrometer, which is designed to measure the muon tracks (together with neutrinos, muons are the only particles not stopped by the calorimeters). It is composed of high precision tracking chambers immersed in a magnetic field provided by toroid magnets.

2.7 EVENT GENERATORS

Event generation

3.1 EVENT GENERATION CHAIN

3.2 SIGNAL AND BACKGROUND EVENTS

Bibliography

- [1] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand. Parton fragmentation and string dynamics. *Physics Reports*, 97(2-3):31–145, July 1983. ISSN 03701573. doi: 10.1016/0370-1573(83)90080-7.
- [2] Mark Thomson. *Modern Particle Physics*. Cambridge University Press, 2013.
- [3] Gavin P. Salam. Towards jetography. *The European Physical Journal C*, 67(3):637–686, June 2010. ISSN 1434-6052. doi: 10.1140/epjc/s10052-010-1314-6.
- [4] George Sterman and Steven Weinberg. Jets from Quantum Chromodynamics. *Physical Review Letters*, 39(23):1436–1439, December 1977. doi: 10.1103/PhysRevLett.39.1436.
- [5] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti- k_t jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063–063, April 2008. ISSN 1029-8479. doi: 10.1088/1126-6708/2008/04/063.
- [6] Lyndon Evans and Philip Bryant. LHC Machine. *Journal of Instrumentation*, 3(08):S08001, August 2008. ISSN 1748-0221. doi: 10.1088/1748-0221/3/08/S08001.
- [7] Stephen Myers. The Large Hadron Collider 2008–2013. *International Journal of Modern Physics A*, October 2013. doi: 10.1142/S0217751X13300354.
- [8] Oliver S. Bruning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock. LHC Design Report Vol.1: The LHC Main Ring. August 2004. doi: 10.5170/CERN-2004-003-V-1.
- [9] Tino Michael. Determination of muon reconstruction efficiencies in the ATLAS detector using a tag & probe approach in Z to $\mu\mu$ events. Master’s thesis, Dresden, Technische Universität Dresden, 2011.
- [10] Ana Lopes and Melissa Loyse Perrey. FAQ-LHC The guide. Technical report, 2022.
- [11] G. Aad et al. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008. doi: 10.1088/1748-0221/3/08/S08003.
- [12] A. Airapetian et al. ATLAS: Detector and physics performance technical design report. Volume 1. May 1999.