Jets, Substructure, and Searching for Dark Matter at the Large Hadron Collider

by

Siddharth Madhavan Narayanan

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

August 2018

Author .....
Department of Physics
August 32, 2018

Certified by .....
Christoph M. E. Paus
Professor of Physics
Thesis Supervisor

Accepted by .....
Somebody
Chairman, Department Committee on Graduate Theses

Jets, Substructure, and Searching for Dark Matter at the Large Hadron Collider

by

Siddharth Madhavan Narayanan

Submitted to the Department of Physics
on August 32, 2018, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Astrophysical observations of gravitational interactions provide strong evidence for the existence of dark matter (DM). Many theories propose and experiments test the hypothesis that DM may have a particle physics origin, but this remains unproven. One such experiment is the Compact Muon Solenoid (CMS) at the Large Hadron Collider (LHC). If DM couples to particles present in protons, it is possible that DM is produced in collisions at the LHC. Because DM is effectively invisible to CMS, we must look for collisions in which DM is produced in association with one or more Standard Model (SM) particles. This thesis describes two different scenarios for the SM particle hypothesis: a single top quark or two light quarks. Both cases result in complicated detector signatures due to the hadronization of final-state quarks. Improved jet substructure techniques to identify some of these unique signatures are presented. Since the observed data is consistent with SM backgrounds in all searches, we translate this result into the most stringent constraints to date on the relevant beyond-SM models.

Thesis Supervisor: Christoph M. E. Paus
Title: Professor of Physics

Contents

Chapter 1

The CMS experiment at the LHC

1.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [7][1] is a circular particle accelerator, 27 km in circumference and between 40 and 175 m below the surface of the French-Swiss border. Designed to collide protons at a maximum center-of-mass energy $\sqrt{s} = 14$ TeV, the LHC has delivered collisions at $\sqrt{s} = 7,8$ TeV (Run 1) and $\sqrt{s} = 13$ TeV (Run 2); the target energy $\sqrt{s} = 14$ TeV will be reached in Run 3. In addition to protons, the LHC accelerates and collides heavy nuclei (Pb and Xe) at lower values of $\sqrt{s}$. In this thesis, we focus exclusively on data recorded from proton collisions during Run 2.

Protons are brought to the LHC by the multi-stage process [2] depicted in Figure 1.1. Hydrogen atoms are stripped of electrons and accelerated by LINAC2 (a linear accelerator) to a kinetic energy of 50 MeV. LINAC2 then feeds the protons into the Booster ring (final energy of 1.4 GeV), followed by the Proton Synchrotron (26 GeV). From the PS, the protons are injected into the Super Proton Synchrotron (450 GeV). Protons exit the SPS and enter the LHC at one of two places, corresponding to two different beams traveling in opposite directions. The two beams intersect in eight places along the LHC, four of which are instrumented by a detector experiment: CMS, ATLAS, LHCb, and ALICE.

Each proton beam in the LHC is accelerated by eight superconducting cavities

[1] Unless otherwise specified, all technical specifications of the LHC are derived from Reference [7]

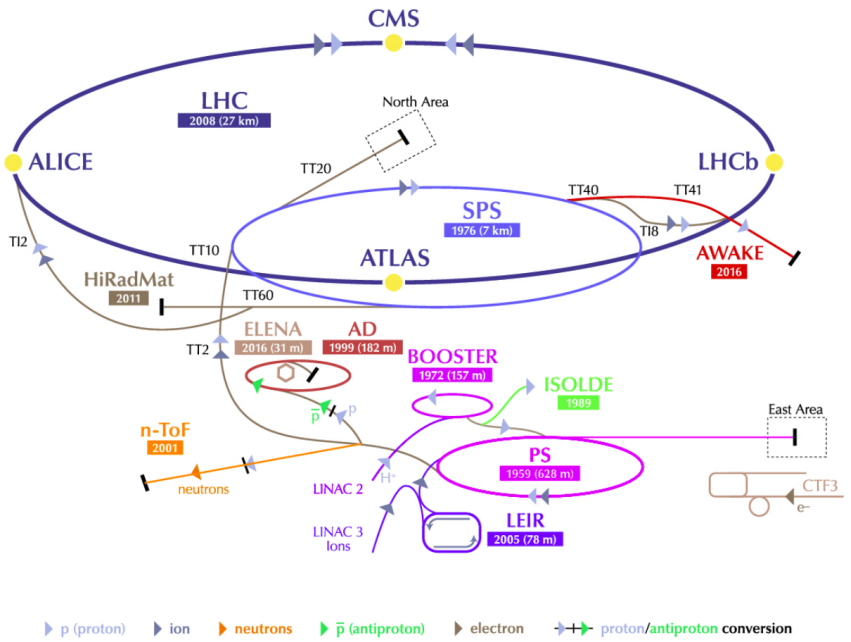

Figure 1.1: Diagram of the CERN accelerator complex, adapted from Reference [6]. The LHC (dark blue) is fed protons (and heavy ions) by a chain of intermediate accelerators, beginning with LINAC2 (dark pink).

exerting radio frequency longitudinal (i.e. parallel to beam direction) electric fields with a frequency of 400 MHz, The maximum RF voltage seen by each beam is 16 MV per revolution. The physical and temporal design of the RF system creates bunches of protons (corresponding to nodes of the oscillating field) approximately 7.5 cm in length and separated by 25 ns. Superconducting NbTi dipole magnets bend the two proton beams in opposite directions as they travel around the ring. Each of the 1232 dipoles is 14 m long and exerts a transverse B field between 0.54 and 8.33 T. To achieve such high B fields, the magnets are cooled to 2 K by superfluid helium. In addition, a number of quadrupole magnets are used to focus and match the beams between the dipoles[2].

In addition to the center-of-mass energy $\sqrt{s}$, the other figure of merit is the number of events producing interesting physics processes, which is defined as:

$$N(pp \to X) = \int dt L \sigma(pp \to X) \tag{1.1}$$

[2]Full details on the various quadrupoles can be found in Table 3.7 of Reference [7].

where σ is the cross section of the relevant process and L is the instantaneous luminosity of the LHC. The cross section is fixed by nature, and so increasing the luminosity is the only handle to increase N. The instantaneous luminosity of two Gaussian beams is given by [7]:

$$L = \frac{N_b^2 n_b f_{\text{rev}} \gamma F}{4\pi\epsilon\beta^*} \tag{1.2}$$

where:

$N_b =$ particles per bunch

$n_b =$ bunches per beam

$f_{\text{rev}} =$ frequency of revolution

$\gamma =$ E/m of beam

$\epsilon =$ emittance of beam

$\beta^* =$ beta function at collision point

$F =$ factor accounting for beam intersection geometry

The instantaneous luminosity evolves as a function of time, primarily due to n_b and N_b being modified by collisions. The total integrated luminosity after time T is:

$$L_{\text{int}} = \int_0^T dt L(t) = L(0)\tau_L \left(1 - e^{-T/\tau_L}\right) \tag{1.3}$$

where $\tau_L \approx 15$ h is the characteristic beam loss timescale and $L(0)$ is the instantaneous luminosity at $T = 0$. The LHC is designed to deliver $L(0) \sim \mathcal{O}(10^{34})$ cm^{-2}s^{-1}. Figure **??** shows the total luminosity delivered by the LHC and recorded by CMS during the 2016 portion of Run 2.

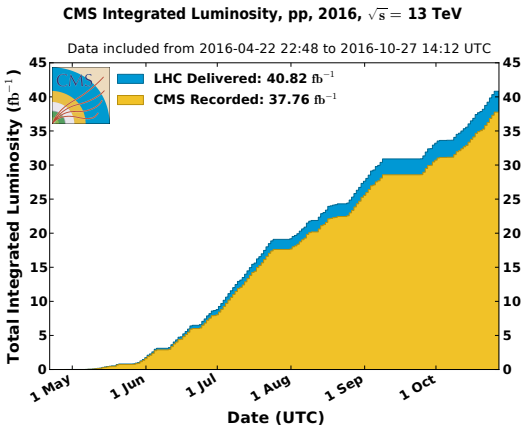

Figure 1.2: Integrated luminosity of the LHC during proton collisions during the 2016 data-taking period [5].

1.2 The Compact Muon Solenoid

The Compact Muon Solenoid (CMS) [3] is one of two general purpose LHC detectors (the other being ATLAS). It is designed to detect and measure stable hadrons, photons, electrons, and muons produced in proton and ion collisions at LHC interaction point 5. From these event descriptions, a number of physics processes can be probed, including SM measurements [???], BSM searches [???], and the discovery of the Higgs boson [???]. In what follows, we will use the (r, ϕ, η) coordinate system with respect to the z axis:

$z =$ distance along beam axis, with $z = 0$ defined to be at the center of the detector

$r =$ distance from the z axis

$\phi =$ azimuthal angle in the plane orthogonal to the z axis

$\eta = -\log \theta/2$ (pseudorapidity), where θ is the angle from the z axis

In this coordinate system, we define x and y to lie in the plane perpendicular to z, with x pointing from the center of the detector to the center of the LHC. As with the pseudorapidity, it is convenient to use quantities invariant under z-boosts, and so we

define the transverse momentum:

$$\vec{p}_{\mathrm{T}} = \begin{pmatrix} p_x \\ p_y \end{pmatrix} \tag{1.4}$$

We will frequently make use of the magnitude of this vector, p_{T}. CMS can detect collision products that are within the fiducial volume of $0 \leq \phi < 2\pi$ and $-5 \leq \eta \leq 5$. Several detector subsystems are used to identify and reconstruct muons, electrons, photons, and charged and neutral hadrons.

Silicon tracker

Starting from the beam pipe, the first of these subsystems is the silicon tracker [4], used to identify charged particles and measure their momenta. The tracker consists of silicon detector geometries: pixels (providing 3D position measurement) and strips (2D). The arrangement of the pixel and strip layers are shown in Figure 1.3. The tracker sits in a near-uniform 3.8 T magnetic field, produced by a superconducting NbTi solenoid. The field lines in the tracker volume are approximately parallel to the beam pipe.

A single silicon pixel has dimensions $285 \times 100 \times 150$ $(\mu\mathrm{m})^3$ (in $r \times r\phi \times z$), leading to a position resolution of $\sim 10 \times 30$ $(\mu\mathrm{m})^2$ (in $r\phi \times z$). The 66 million pixels are arranged into 7 layers: 3 cylindrical "barrels" (at $r = 4.4,\ 7.3,\ 10.2$ cm) and 2×2 "endcap" annulli (at $z = \pm 34.5,\ \pm 46.5$ cm). Outside the pixel layers are the strip layers, consisting of 9.3 million silicon strips arranged into barrels and endcaps. The resolution in $r\phi$ varies between 10 and 50 μm, depending on the location and pitch of the given strip. Certain strip layers contain two layers of strips, rotated through a "stereo" angle (100 mrad) with respect to each other. By matching adjacent hits, the stereo measurement can add a third dimension (z for barrel, r for endcap) to the strip's 2D measurement, with resolution 100-500 μm. There are a total of 10 barrel layers ($0.2 < r < 1$ m) and 24 endcap layers ($0.6 < |z| < 2.8$ m).

Pixels with a signal greater than a tuneable readout threshold (typically around $3000Q_e$) are read out. These pixels are then aggregated with adjacent signals to form

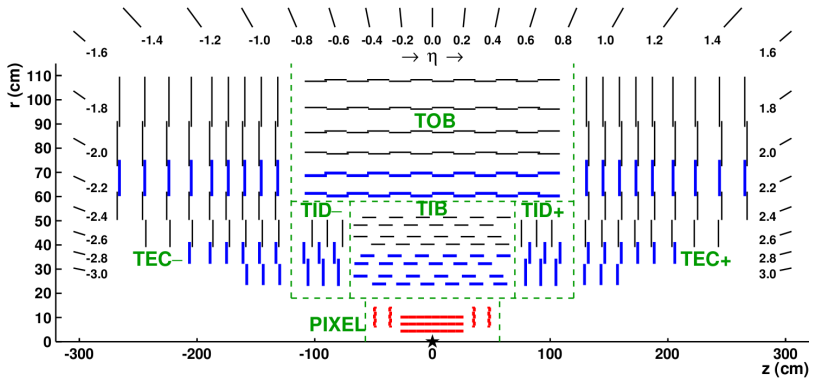

Figure 1.3: Diagram of a slice of the CMS tracking system. The pixel layers are shown in bold red lines. Single-strip (double-strip) layers are indicated by thin black (bold blue) lines. The double-strip modules each consist of two back-to-back strips, rotated with respect to each other, that can provide 3D localization of the hits. Adapted from Reference [4].

pixel clusters, which are further subjected to readout thresholds ($\sim 4000Q_e$). The exact position of the particle in this layer (known as a "hit") is inferred by fitting the charge distribution of the pixels in this cluster to pre-determined templates. A similar method is employed to determine the strip hit positions, with some modifications to account for Lorentz drift of the charges in the silicon detector due to the B-field. The efficiency of reconstructing hits varies with the detector type, location, and particle momentum, but is generally greater than 99% (99.5% if defective modules are not considered).

Tracks are found using an iterative "inside-out" process, where each iteration has five steps:

1. Define seeds using pixel hits, double-strip hits (i.e. hits with 3D information), and an estimate of the beam spot (collision point). At least 3 hits are needed for the seed.

2. Use a Kalman filter [???]to evolve track seeds through the rest of the tracker and find hits, accounting for the B-field and energy loss.

3. Estimate trajectory parameters after finding all hits.

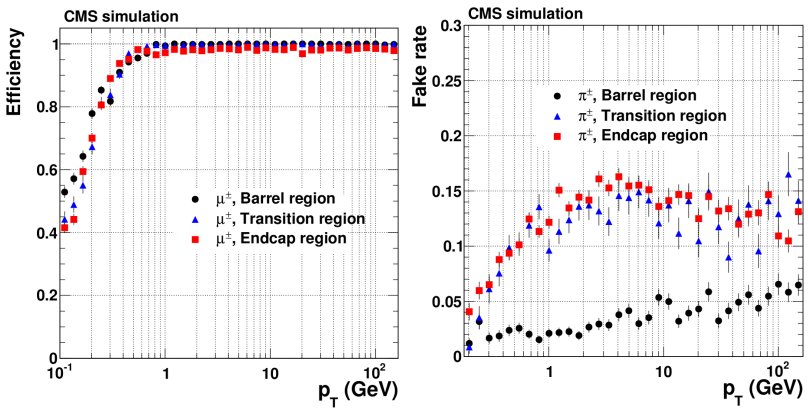

Figure 1.4: Efficiency (fake rate) of the CMS track fit algorithm, evaluated using simulation of muons (charged pions).

4. Decide whether to keep found tracks based on quality requirements (e.g. number of missing hits)

5. Remove hits associated with tracks from hit collection and repeat.

The trajectory parameters referred to in step 3 are the 5 parameters of a helix: ρ (curvature), ϕ_0 (azimuthal angle), λ ($\cot\theta$), d_0 ("impact parameter", minimum r of track), z_0 (minimum $|z|$ of track). The CMS track fit typically has 5-6 iterations, with each successive iteration loosening the seed and track fit requirements to look for more difficult tracks (e.g. missing hits, large d_0). The efficiency and fake rate of this process, as a function of track p_{T}, are shown in Figure 1.4. For muons with $|\eta| < 1.5$ and $p_{\mathrm{T}} > 1$ GeV, the tracking efficiency is over 98%, with a combinatorial fake rate of 2-6%.

Electromagnetic calorimetry

The CMS electromagnetic calorimeter [1] (ECAL) is a homogenous detector with good energy and angular resolution, composed of 76,000 $PbWO_4$ crystals. The crystals are arranged in two sections: a cylindrical barrel (EB) covering $|\eta| < 1.44$ and two endcap annuli (EE) extending to $|\eta| < 3$. This provides slightly more coverage than the tracking volume. Each crystal in the EB (EE) has dimensions $2.2 \times 2.2 \times 23$ ($2.68 \times 2.68 \times 22$) ($\mathrm{cm}^3$), with the long dimension pointing towards the beam. This

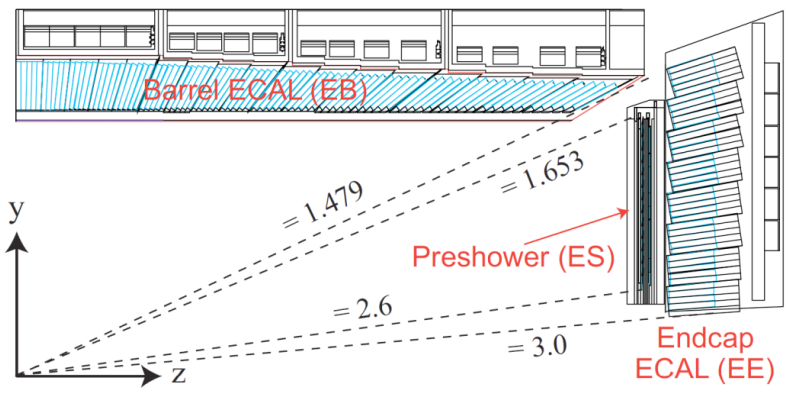

Figure 1.5: One quadrant of the CMS ECAL (symmetric with rotation around z and reflection across $z = 0$). The dashed lines indicate values of η. Adapted from Reference [1].

can be compared to a Moliére radius $r_M = 2.19$ cm and a radiation length of $X_0 = 0.89$ cm. A cross-sectional area comparable to $r_M \times r_M$ facilitates the differentiation of different electromagnetic (EM) showers arising from electrons and photons. The depth of the crystal (in units of X_0) drives the excellent energy resolution, which is determined using a standalone electron beam:

$$\frac{\sigma_E}{E} = \frac{2.8\%}{\sqrt{E/\text{GeV}}} \oplus \frac{12\%}{E/\text{GeV}} \oplus 0.3\% \tag{1.5}$$

Scintillation photons from the $PbWO_4$ crystals are collected by avalanche photodiodes (APDs) in the EB and vacuum phototriodes (VPTs) in the EE, which provide amplification factors of 50 and 10, respectively.

At high momenta, the two photons from a π^0 decay may merge into a single ECAL crystal. This primarily occurs at high $|\eta|$ due to the z-boost of the intitial state. To differentiate one- and two-photon deposits, a "preshower" detector sits in front of the EE ($1.6 < |\eta| < 2.5$). The preshower detector consists of a lead absorber and silicon strips. A photon (or photon pair) initiates a shower in the lead. The shower can be resolved in the silicon strips, which have resolution $\mathcal{O}(1-10)$ mm.

The physical placement of all three ECAL components is shown in Figure 1.5.

Hadronic calorimetry

Muon chambers

Online trigger system

1.3 Simulation of collisions

1.3.1 Physics simulation

1.3.2 Detector simulation

1.4 Particle reconstruction algorithms

Tracks and vertices

Electrons and photons

Jets

Muons

Particle flow algorithm

Missing momentum

Bibliography

[1] A Benaglia et al. The CMS ECAL performance with examples. *Journal of Instrumentation*, 9(02):C02008, 2014.

[2] M. Benedikt, P. Collier, V. Mertens, J. Poole, and K. Schindl. LHC Design Report. 3. The LHC injector chain. 2004.

[3] S. Chatrchyan et al. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

[4] Serguei Chatrchyan et al. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *JINST*, 9(10):P10009, 2014.

[5] CMS Collaboration. `https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideGlobalHLT#Trigger_development_for_Run_2`.

[6] Cinzia De Melis. The CERN accelerator complex. Complexe des accÃľlÃŀrateurs du CERN. Jan 2016. General Photo.

[7] Lyndon Evans and Philip Bryant. LHC Machine. *JINST*, 3:S08001, 2008.