

**Shining Light on Dark Matter,
One Photon at a Time**

by

Brandon Leigh Allen

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

Doctorate of Science in Physics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2019

© Massachusetts Institute of Technology 2019. All rights reserved.

Author
Department of Physics
May 18, 2019

Certified by
Christoph E.M. Paus
Professor
Thesis Supervisor

Accepted by
Nergis Mavalvala
Associate Department Head for Education

**Shining Light on Dark Matter,
One Photon at a Time**

by

Brandon Leigh Allen

Submitted to the Department of Physics
on May 18, 2019, in partial fulfillment of the
requirements for the degree of
Doctorate of Science in Physics

Abstract

A search is conducted for new physics in final states containing a photon and missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV. The data collected by the CMS experiment at the CERN LHC correspond to an integrated luminosity of 35.9 inverse femtobarns. No deviations from the predictions of the standard model are observed. The results are interpreted in the context of dark matter production and limits on new physics parameters are calculated at 95% confidence level. For the two simplified dark matter production models considered, the observed (expected) lower limits on the mediator masses are both 950 (1150) GeV for 1 GeV dark matter mass.

Thesis Supervisor: Christoph E.M. Paus

Title: Professor

Acknowledgments

This is the acknowledgements section. You should replace this with your own acknowledgements.

Contents

1	The CMS Detector	9
1.1	Inner Trackers	11
1.2	Electromagnetic Calorimeter	13
1.2.1	Preshower Detector	15
1.3	Hadronic Calorimeter	16
1.4	Muon Chambers	18
1.5	Online Trigger System	20
1.6	Detector Simulation	22
2	Global Event Reconstruction	23
2.1	Particle Flow Elements	25
2.1.1	Tracks	25
2.1.2	Primary Vertexing	26
2.1.3	Secondary Vertexing	27
2.1.4	ECAL Superclusters	28
2.1.5	HCAL Clusters	29
2.1.6	Muon Segments	30
2.1.7	Isolation	30
2.2	Particle Identification	31
2.2.1	Muons	31
2.2.2	Electrons	32
2.2.3	Isolated Photons	33
2.2.4	Hadrons	36

2.2.5	Jets	37
2.2.6	Missing Tranverse Energy	38
2.2.7	ECAL gain-switch effect	39

Chapter 1

The CMS Detector

The Compact Muon Solenoid (CMS) detector [1] is one of two hermetic, general purpose detectors at the Large Hadron Collider. The primary motivation for the experiment was the discovery of the Higgs boson by observing its decays to photons, electrons, and muons. Towards this end, the detector was built to fulfill the following goals:

- Unambiguous charge identification of muons with momenta up to 1 TeV
- 1 GeV mass resolution on 100 GeV pairs of muons, electrons, and photons
- Efficient triggering and tagging of τ lepton and b quark decays
- Good resolution on the hadronic energy and missing transverse energy
- Sufficient time resolution to deal with 40 MHz of collisions

The CMS detector consists of four main subdetectors: the inner trackers, the electromagnetic calorimeter (ECAL), the hadronic calorimeter (HCAL), and the muon chambers. The first three are within the field volume of the eponymous 3.8 T superconducting NbTi solenoid magnet while the muon chambers are embedded in the return yoke of the magnet. Additionally, there is an online triggering system to reduce readout by over four orders of magnitude from 40 MHz to $\mathcal{O}(1)$ kHz for prompt reconstruction.

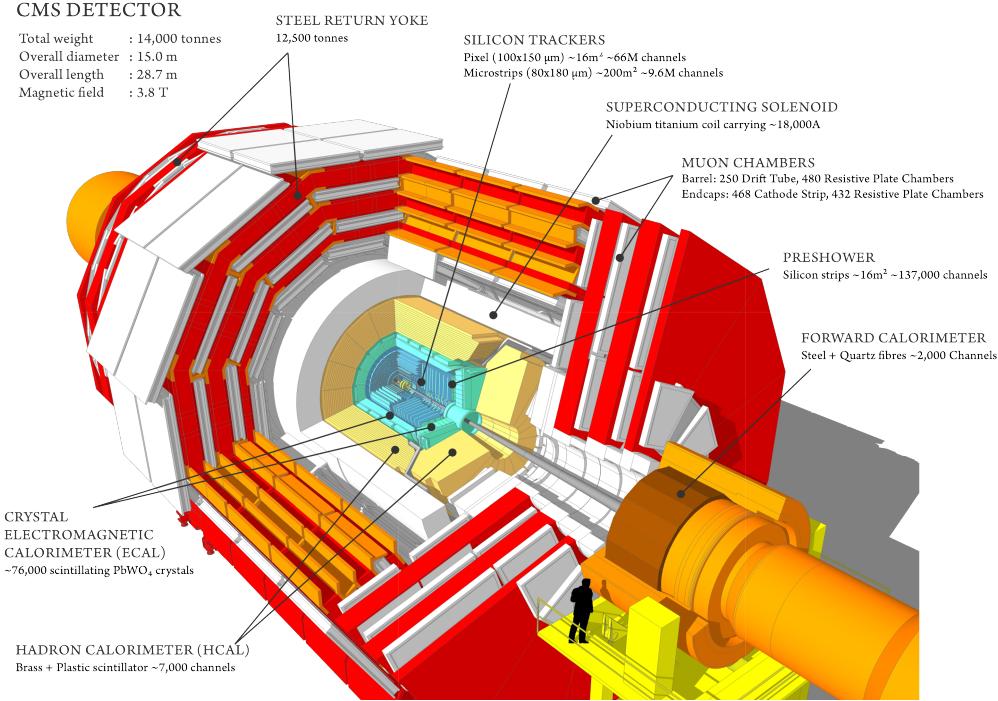


Figure 1-1: A cutaway view of the CMS detector. The labels identify the solenoid as well as the different subdetectors and their components. Reprinted from Reference [2].

The overall layout of the CMS detector is shown in Figure 1-1. The CMS detector has a cylindrical geometry with concentric barrel shaped detectors in the central region and disc shaped detectors in the forward region and a total weight of 12500 tons, a length of 22 m, and a diameter of 15 m. We define a right-handed cartesian coordinate system with the origin at the nominal center of the detector. The z coordinate is along the beam axis, with the positive z -axis pointing counter-clockwise as seen from the sky, and the x and y coordinates are perpendicular to it, with the positive x -axis pointing from the center of the detector to the center of the LHC ring and the positive y -axis pointing upwards. However, we more commonly use the following cylindrical coordinate system when working with the CMS detector:

- distance z along the beam axis
- distance r from the beam axis
- polar angle θ measured with respects to the positive z -axis

- azimuthal angle ϕ in the plane orthogonal to the beam axis with zero at the positive x -axis.

The four-momentum of a particle is $p = (p_x, p_y, p_z, E)$ in the cartesian basis and a particle of mass m produced at rest in the center of the detector has $p = (0, 0, 0, m)$. While the momenta along the beam axis of the two incoming protons are equal, the momenta of the incoming partons involved in the hard scattering often are not as discussed in Section ???. Thus, we define two kinematic quantities that are Lorentz-invariant with respect to a boost along the beam axis: the transverse momentum $\vec{p}_T = p_x \hat{x} + p_y \hat{y}$ with magnitude $p_T = \sqrt{p_x^2 + p_y^2}$ and the pseudorapidity $\eta = -\ln \tan \theta/2$.

In terms of p_T , η , and ϕ , we have the following expressions for our cartesian variables: $p_x = p_T \cos \phi$, $p_y = p_T \sin \phi$, $p_z = p_T \sinh \eta$, and $E = p_T \cosh \eta$, with the last equality assuming the mass of the particle is negligible compared to its momentum. In terms of our Lorentz-invariant coordinates, the four-momentum of a given particle is $p = (p_T, \eta, \phi, E)$. Additionally, the spatial separation of two particles is given by $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$ and the fiducial acceptance of the CMS detector is from $0 \leq \phi < 2\pi$ and $-5 \leq \eta \leq 5$.

1.1 Inner Trackers

Closest to the interaction point, the inner trackers identify charged particles and measures their momenta [1]. Additionally, high resolution tracks are used to identify primary and secondary vertices. The magnetic field in the tracker volume is uniform with strength 3.8 T and field lines parallel to the beam direction. The tracker volume extends to 1.2 m in r and 2.9 m in z , providing coverage for $|\eta| < 2.5$, and is instrumented with silicon pixels and strips. Each silicon sensor is a p - n semiconductor junction with a bias voltage applied. When a charged particle passes through the depletion region of the junction, electron-hole pairs are produced and collected by the readout electronics. A schematic of the inner tracker system is shown in Figure 1-2.

The 66 million individual pixel sensors, each measuring $285 \mu\text{m} \times 100 \mu\text{m} \times 150 \mu\text{m}$ in $r \times r\phi \times z$, are arranged into seven layers: three cylindrical barrels at $r =$

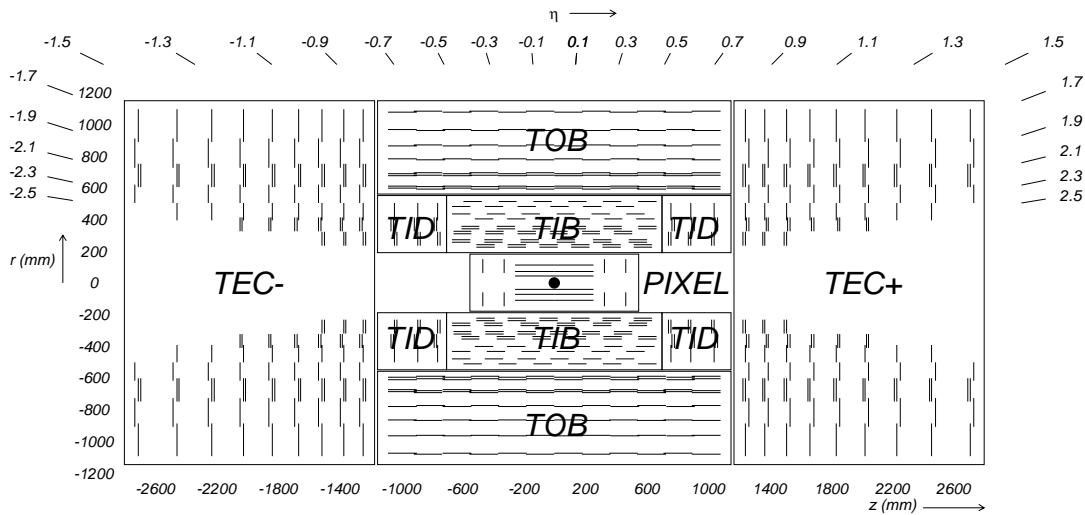


Figure 1-2: A schematic view of the CMS inner tracker system. Silicon pixel and strip detectors are shown. The volumes labeled TIB, TID, TOB, and TEC are all strip trackers. The double lines indicate back-to-back modules that deliver stereo hits. Reprinted from Reference [1].

4.4, 7.3, 10.2 cm and two bi-layer endcap annuli at $z = \pm 34.5, \pm 46.5$ cm. Due to the geometry of the pixel detector, tracks typically cross the sensor at a 20° angle, leading to the charge deposit from a single track to be shared among multiple pixels in the same layer. The exact position of a particle in each layer is determined by interpolating the signals from multiple adjacent pixels with an analog pulse height greater than a tuneable read-out threshold. Thus, each pixel hit is localized to an area of $\sim 15 \mu\text{m} \times 20 \mu\text{m}$ in $r\phi \times z$, providing a much higher spacial resolution than the raw pixel spacing.

The pixels are surrounded 9.3 million silicon strips measuring $10 \text{ cm} \times 80 \mu\text{m}$ arranged in ten cylindrical layers in the barrel and twelve disks in each endcap. The Tracker Inner Barrel (TIB) consists of the first four layers and extends from 20 cm to 55 cm in the radial direction while out six layers constitute the Tracker Outer Barrel (TOB) with an outer radius of 116 cm and extent in $|z|$ of 118 cm. The remaining area in the barrel is covered by the Tracker Inner Disk (TID), consisting of the three disks located from 80 to 90 cm in $|z|$. The Tracker EndCaps (TEC) have nine disks each and cover the region $124 \text{ cm} < |z| < 282 \text{ cm}$.

The majority of the strips are oriented perpendicular to the ϕ direction: parallel to the beam pipe in the barrel region and radially aligned in the endcap region. The strip pitch varies from 80 to 184 μm with the smallest pitch in the innermost layer. This detector geometry provides good resolution in the r - ϕ plane for barrel and the z - ϕ plane for the endcap but little information on the orthogonal directions. To compensate for this, one third of the strips are double-layered with a stereo angle of 100 mrad between the layers. Matching hits between adjacent layers enables a measurement of the z and r coordinates in the barrel and endcap, respectively. The final spacial resolution is 10-50 μm in the direction perpendicular to the strips and 100-530 μm in the parallel direction on the stereo modules.

1.2 Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL) is a homogeneous, hermetic calorimeter composed of 76,000 lead tungstate (PbWO_4) crystals [1]. High density (8.3 g/ cm³) lead tungstate was chosen due to its radiation hardness, fast scintillation decay time constant of 25 ns, small Moliere radius $r_M = 21.9$ mm, and short radiation length $X_0 = 8.9$ mm. All of these factors combine to enable the construction of a compact calorimeter with high granularity and excellent energy resolution.

Figure 1-3 shows the layout of the ECAL. The central barrel region (EB) has 61200 crystals arranged in a 170×360 η - ϕ grid (0.0174 \times 0.0174 granularity) with a coverage up to $|\eta| = 1.44$ while the two endcap annuli (EE) each have 7324 crystals organized in a x - y grid with coverage in the range $1.479 < |\eta| < 3.0$. Each crystal in the EB has a truncated pyramidal shape with a length of 230 mm, a 22 mm \times 22 mm front-face cross-section, and a 26 mm \times 26 mm rear-face cross-section while a crystal in the EE has a cuboid-like shape with a length of 220 mm, a 28.6 mm \times 28.6 mm front-face cross-section, and a 30 mm \times 30 mm rear-face cross-section. The cross-sectional area of approximately one Moliere radius and length of approximately 25 radiation lengths allows just a few crystals to contain the entire transverse and longitudinal development of the shower. To reduce the likelihood of the primary photon or electron

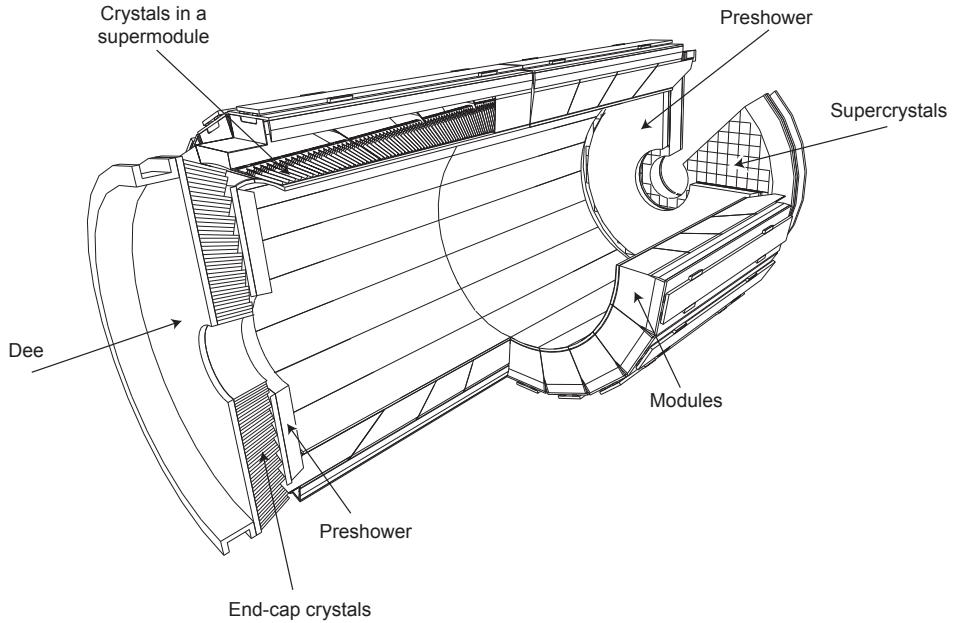


Figure 1-3: The layout of the CMS electromagnetic calorimeter. The barrel and endcap calorimeters are shown. The pre-shower detector sits in front of the endcaps. Reprinted from Reference [1].

emerging from the hard scattering depositing most of its energy in passive material, the crystals do not point directly to the interaction region.

The interactions between the 3.8 T magnetic field and the geometries of the ECAL lead to the selection of different photosensors in the barrel and endcaps. In the endcaps, the magnetic field is parallel to the path of the photoelectrons and has a negligible effect on the gain, while in the barrel, the magnetic field is perpendicular and reduces the gain by a factor proportional to the distance traveled by the photoelectrons. Thus, in the barrel, solid-state reverse-structure avalanche photodiodes (APDs) with a depletion layer of $6.0 \pm 0.5 \mu\text{m}$ are used, while photomultiplier tubes with a single gain stage and a very fine copper mesh anode called vacuum phototriodes (VPTs) are used in the endcaps. Two APDs with an active area of 25 mm^2 are glued to the rear of each crystal in the barrel while only one VPT with an active area of 280 mm^2 is needed per crystal in the endcap. The APDs and VPTs amplify the initial signal of approximately 4.5 photoelectrons per MeV of energy deposit per

crystal by a factor of 50 and 10, respectively.

The small signals from the photodetectors are shaped and amplified in the Multi-Gain Preamplifier (MGPA) and a 12-bit analog-to-digital converter (ADC) samples the pulse every 25 ns. Each output voltage pulse has a length of approximately 300 ns, with the maximum at approximately 75 ns and a slow decay afterwards. The MGPA has multiple gain modes of 12, 6, and 1, and the gain chosen for the output decreases once the signal has saturated the previous gain setting. After the pulse falls below the saturation threshold, the lower gain setting is maintained for the next five samples. This mechanism provides a dynamic signal range from a few MeV to a maximum of 1.5 TeV in the barrel and 3.1 TeV in the endcaps.

The energy resolution of the ECAL was measured using an electron beam to be

$$\frac{\sigma_E}{E} = \frac{2.8\%}{\sqrt{E/\text{GeV}}} \oplus \frac{12\%}{E/\text{GeV}} \oplus 0.3\%, \quad (1.1)$$

where E is the energy of the incident particle and the three terms on the RHS are the stochastic, noise, and constant terms, respectively. The stochastic term is dominated by event-to-event fluctuations in the lateral shower containment and a photo-statistics contribution of 2.1%. Electronic and digitization noise drive the noise term while the constant term comes from a non-uniform longitudinal light collection and intercalibration errors.

1.2.1 Preshower Detector

At high momenta and high $|\eta|$, the two photons from a π^0 decay can merge into a single ECAL crystal due to the large boost in the z -direction of the initial state. By forcing the initiation of an electromagnetic shower in a region with high spacial resolution in front of the ECAL endcaps, the preshower detector can differentiate between one- and two-photon deposits in the region $1.6 < |\eta| < 2.5$. The preshower detector consists of two alternating layers of passive lead absorbers and active silicon strip sensors. The first (second) lead layer is two (one) radiation lengths thick and the subsequent sensor plane has vertical (horizontal) strips of 6 cm length and 1.9 mm

pitch. The silicon strips resolve the shower with a resolution of $\mathcal{O}(1\text{--}10)$ mm, enabling the disambiguation of two nearly collinear photons and the identification of π^0 decays.

1.3 Hadronic Calorimeter

The hadronic calorimeter (HCAL) is a set of four heterogenous calorimeters that provide hermetic coverage when combined [1]: the barrel calorimeter (HB) covering $|\eta| < 1.3$, the endcap calorimeter (HE) covering $1.3 < |\eta| < 3$, the outer calorimeter (HO) covering $|\eta| < 1.3$, and the forward calorimeter (HF) covering $3 < |\eta| < 5$. The region covering $|\eta| < 3$ shall be referred to as the central HCAL. The granularity of the HCAL in $\eta\text{-}\phi$ is 0.087×0.087 for $|\eta| < 1.6$, 0.17×0.17 for $1.6 < |\eta| < 3.0$, and 0.175×0.175 for $3 < |\eta| < 5$. The layout of the HCAL is shown in Figure 1-4.

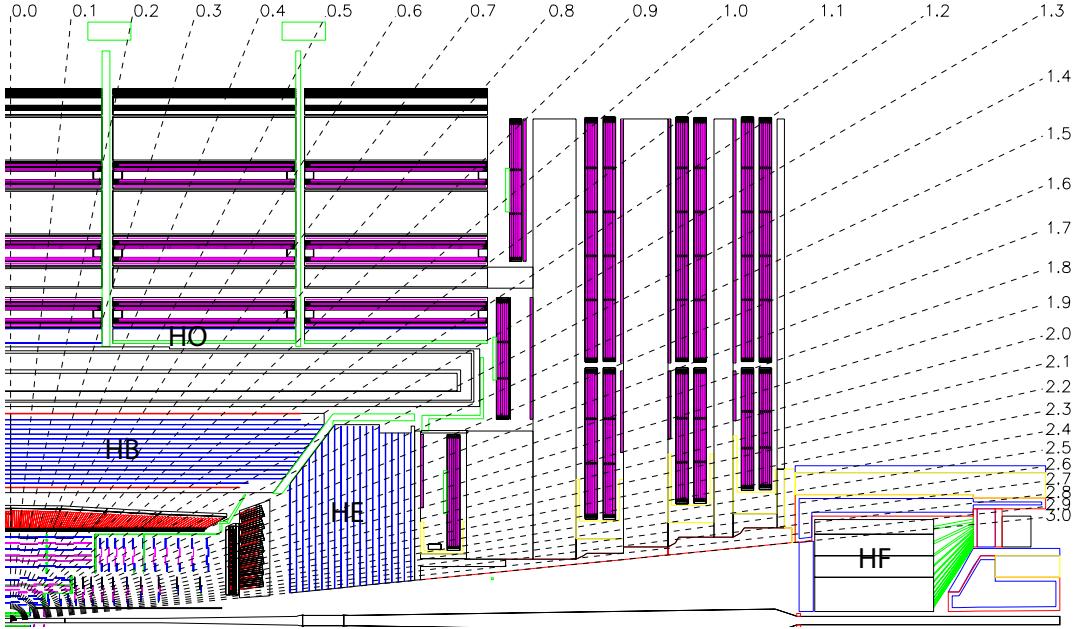


Figure 1-4: The layout of the CMS hadronic calorimeter. The barrel, endcap, outer, and forward calorimeters are shown and labeled. The muon chambers are shown but not labeled. The dashed lines denote different values of pseudorapidity. Reprinted from Reference [1].

The HB and HE are sampling calorimeters with 16 and 17 thin plastic scintillator

layers, respectively, interleaved with thick absorber layers made of a non-magnetic brass alloy with an interaction length $\lambda_I = 1.5$ cm. The layers range in thickness from 40 to 75 mm providing a total absorber depth ranging from a minimum of 5.82 λ_I at $|\eta| = 0$ to a maximum of 10.6 λ_I at $|\eta| = 1.3$ in the barrel and approximately nine interaction lengths throughout the endcaps, with the ECAL contributing another interaction length worth of material. The dimensions of the HB and HE are determined by the requirement that they reside between the ECAL and the solenoid. To circumvent this constraint, an additional layer of scintillator located in the return yoke of the magnet, the HO, utilizes the solenoid material as an additional interaction length of absorber. Hybrid photodiodes (HPDs) are used to read out the scintillator light in the HB and HE while silicon photomultipliers (SiPMs) are used in the HO.

Located 11 meters from the interaction point, the HF is a sampling calorimeter covering the region of phase-space with high pseudorapidity. Since the particle flux here is much higher than in the central region, the HF uses radiation-hard steel absorber instrumented with two sets of scintillating quartz fibers. Charged particles produced by showers in the steel traverse the quartz fibers and emit Cherenkov radiation that is recorded by photomultiplier tubes. To distinguish hadrons from photons and electrons, the second set of fibers starts at a depth of 22 cm.

Since hadrons interact with the ECAL as well as the HCAL, the energy resolution of the detectors must be considered in tandem in the central region. Using a charged particle test beam, the combined ECAL+HCAL energy resolution was measured to be

$$\frac{\sigma_E}{E} = \frac{0.847}{\sqrt{E/\text{GeV}}} \oplus 0.074 \quad (1.2)$$

and the standalone HF energy energy resolution is

$$\frac{\sigma_E}{E} = \frac{1.98}{\sqrt{E/\text{GeV}}} \oplus 0.09. \quad (1.3)$$

1.4 Muon Chambers

The outer most components of CMS are the muon triggering, identification, and detection chambers [1]. These muon detectors are interleaved with the steel return yoke of the magnet resulting in a characteristic *S*-shape for the muon trajectories due to the reversal of magnetic field direction across the solenoid. Signal purity in the muon detectors is high as hadrons, electrons, and photons are stopped by the calorimeters while muons are minimum ionizing particles (MIPs) that lose little energy while traversing the detector. Taking advantage of the large detector volumes required by the outer solenoid radius of 3.5 m, three types of gas ionization chambers are used: drift tubes (DTs) in the barrel covering $|\eta| < 1.2$, cathode strip chambers (CSCs) in the endcaps covering $0.9 < |\eta| < 2.4$, and resistive plate chambers (RPCs) in both covering $|\eta| < 2.1$. The layout of the muon detectors is shown in Figure 1-5.

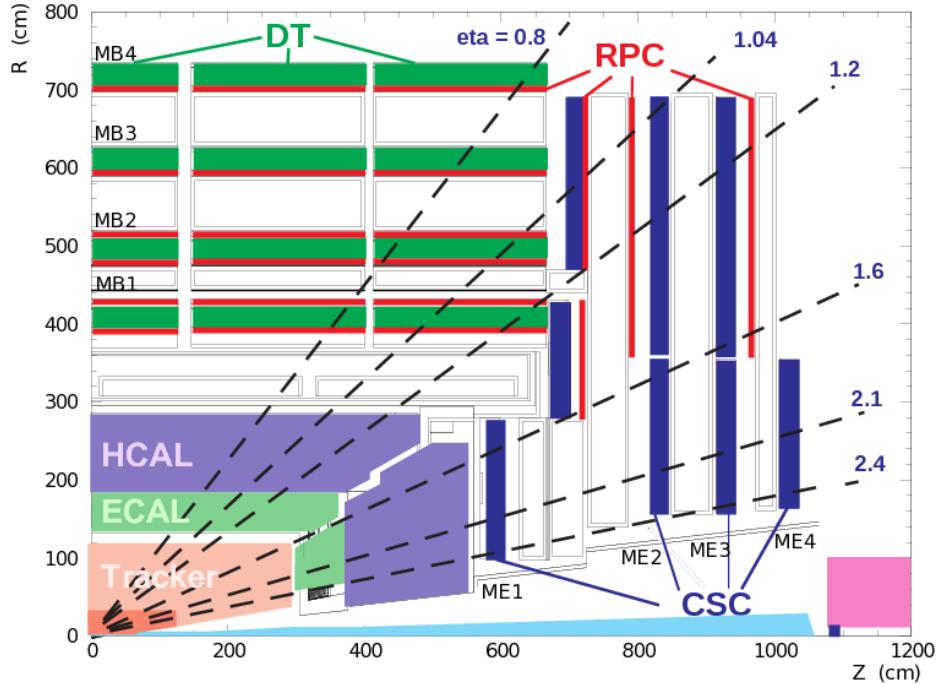


Figure 1-5: The layout of the CMS muon chambers. The four DT stations are labeled MB1-MB4, the four CSC stations are labeled ME1-ME4, and the RPC stations are shown in red. The dashed lines denote different values of pseudorapidity. Reprinted from Reference [1].

The DT chambers consist of rectangular drift cells with transverse dimensions of $42\text{ mm} \times 13\text{ mm}$ filled with a 85:15 Ar/CO₂ mix and a gold/steel anode wire held at a voltage of 3.6 kV. The maximum drift time per cell of 400 ns provides a spatial resolution of approximately 250 μm . A single chamber is made out of superlayers which are in turn made out of four individual cells for a combined resolution of 100 μm . The chambers are arranged into four 2.4 m thick dodecagonal rings called the muon barrel stations. All four stations have two superlayers per chamber that measure position in the $r\text{-}\phi$ plane while the chambers in the inner three stations have an additional superlayer that measures position in the $r\text{-}z$ plane. Together, the four muon stations and iron yokes form a wheel with one between the solenoid and the first iron layer, two in between the yokes, and one outside. The outer barrel is composed of five wheels in total.

Due to their faster response time and better spatial resolution, CSCs are used to handle the higher muon and background fluxes in the endcap region. Each CSC is filled with a 50:40:10 CO₂/Ar/CF₄ mix and instrumented with 80 cathode strips held at voltages of 2.9–3.6 kV relative to the gold-plated tungsten anode wires. The strips run radially outward to measure position in the $r\text{-}\phi$ plane while the wires run perpendicular measure the η and beam-crossing time of the muons. The four muon stations in each endcap are made of CSCs arranged into annuli oriented perpendicular to the beam axis and interspersed between the flux return plates. The three inner stations have multiple annuli with smaller ones fitting inside the larger ones while the fourth station only has one close to the beamline.

An RPC is a parallel-plate double-gap chamber with a time resolution of one nanosecond and a poor spatial resolution. The RPCs are interspersed throughout the DT and CSCs to provide an independent muon trigger system that can identify the correct bunch crossing time.

1.5 Online Trigger System

To achieve an instantaneous luminosity of $\mathcal{O}(10^{34}) \text{ cm}^{-2} \text{ s}^{-1}$, the LHC has bunch crossings every 25 ns for a total data rate of 40 MHz, much higher than the 100 kHz readout rate for the detector and the $\mathcal{O}(1)$ kHz data reconstruction and tape writing rates. Fortunately, uninteresting elastic scattering and QCD inelastic scattering events dominate the approximately $100 \mu\text{b}$ total proton-proton cross-section at the LHC. Meanwhile, most new physics processes have a predicted on the order of picobarns or femtobarns and even the highest rate SM EWK cross-sections are $\mathcal{O}(10)$ nb, so not every event needs to be readout, reconstructed, and written to tape.

A two-stage trigger system selects the events to keep for permanent storage and analysis [1]. The Level 1 hardware-based trigger (L1) selects interesting events based on incomplete detector information to reduce readout and computation times. Events selected by the L1 trigger are fully readout and passed to the high level software-based trigger (HLT) where the full event is reconstructed using the data acquisition system (DAQ), a computing farm with over 20k CPU cores. Events selected by the HLT are sent to the offline computing resources for full reconstruction followed by storage on disk and tape.

The L1 trigger uses field programmable gate arrays (FPGAs) and application specific integrated circuits (ASICs) to make decisions within $4 \mu\text{s}$ of each collision. Each subdetector module has a hardware trigger that reconstructs objects called trigger primitives (TPs). In the two calorimeters, clustered energy deposits from each tower are sent to a regional calorimeter trigger (RCT) which correlates the information from adjacent towers and between the ECAL and HCAL into electron, photon, and jet candidates. The outputs from the RCT are passed to the global calorimeter trigger (GCT) which computes global event variables such as the total transverse energy, the hadronic transverse energy, and the momentum imbalance. Muon track candidates are produced from a simple segment-finding and tracking algorithm in each of the three types of muon chambers. The global muon trigger (GMT) receives the candidates and combines them with information from the GCT

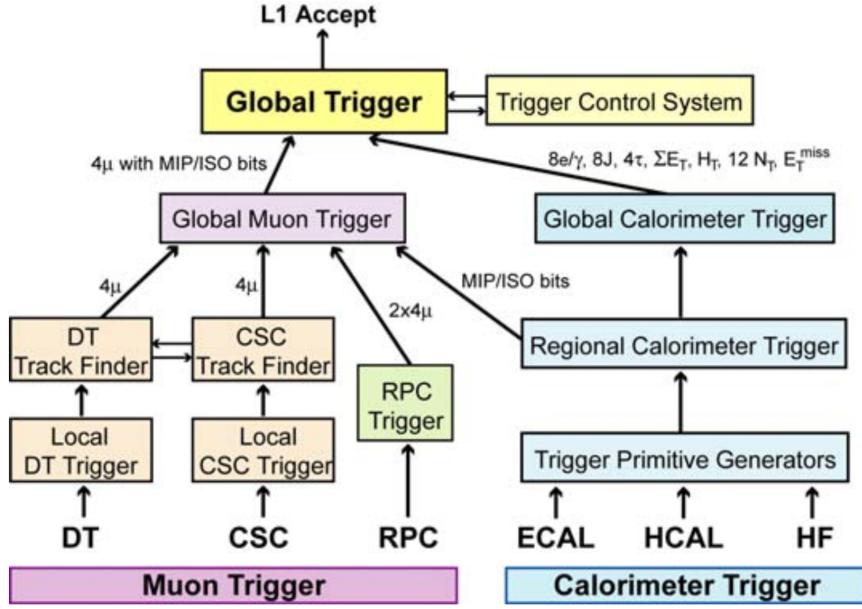


Figure 1-6: The Level-1 trigger architecture with components labeled and information flow indicated. Reprinted from Reference [1].

to produce the final set of muon candidates. The global trigger (GT) decides if an event should be sent to the HLT based on the information it receives from the GCT, the GMT, and the Timing, Trigger, and Control (TTC) system that monitors the readiness of the sub-detectors and the DAQ. The inner tracker is not included because the detector readout and reconstruction process take longer than the time allotted. Figure 1-6 shows a schematic description of the full L1 trigger process.

The HLT uses a version of the offline reconstruction software optimized to process a single event within 200 ms at the cost of some precision. The reconstruction is split into a series of filters that make decisions within regions of interest defined by the L1 trigger information. The HLT implements the desired trigger logic by constructing trigger paths out of these filters. For example, three filters relevant for dark matter searches are (1) a single photon, (2) large momentum imbalance, and (3) large hadronic energy. Combining the first two yields a trigger path targeting the monophoton channel while combining the latter two yields one for the monojet channel. To reduce CPU usage, simple decisions such as those using only the calorimeters or muon information are computed before complex decisions such as tracking.

An event must pass all the filters in a given path to be recorded in an output primary dataset (PD). Trigger paths are organized into PDs such that each PD contains events with similar topologies. Some examples are the SinglePhoton, DoubleMuon, and Jet/HT datasets. A single event that exhibits multiple different physics signatures can pass multiple trigger paths and end up in multiple PDs. These overlaps must be considered in the offline analysis.

1.6 Detector Simulation

The Geant4 program [3, 4] is used to simulate the detector response to the particles produced in collisions. Starting with the output of the particle-level MC described in Section ??, final state particles are propagated through the solenoid’s magnetic field into the passive and active elements of the detector where energy deposition, decay, and showering are simulated. Additional inelastic proton-proton collision are overlaid into an event to simulate the effects of pileup. As the particles interact with the detector, the response of the readout electronics is simulated, including the effects of noise. In order to minimize differences between simulated events and those from collisions, the reconstruction software and output format are the same for both up to the retention of additional truth information from the generators.

Chapter 2

Global Event Reconstruction

In the previous chapter, we discussed the interactions of particles with the individual subdetectors and how these generate electrical signals. Now, we shall discuss the reverse process, namely reconstructing the individual particles or physics objects from the electrical signals recorded by the subdetectors.

Traditionally, each class of physics object is reconstructed using information from a single subdetector: muons from the muon chambers, isolated photons and electrons from the ECAL, jets and missing transverse energy from the HCAL, and secondary vertices from τ lepton and b hadron decays from the tracker. However, as depicted in Figure 2-1, each type of particle interacts with multiple different subdetectors and this information is lost unless the information from all the subdetectors is combined into a single global event description.

The particle flow (PF) algorithm [5] leverages the fine angular granularity of the calorimeters and the excellent momentum resolution of the inner tracker and muon chambers to greatly improve the reconstruction of physics objects and include soft particles that would otherwise be ignored. This is especially advantageous for jet energy measurements as roughly 62% of the jet energy is carried by charged hadrons, approximately 27% by photons, around 10% by neutral hadrons, and about 1.5% by neutrinos.

The distinguishing feature of the PF algorithm is to combine multiple detector signals together into a single PF candidate. The input detector signals are the tracks,

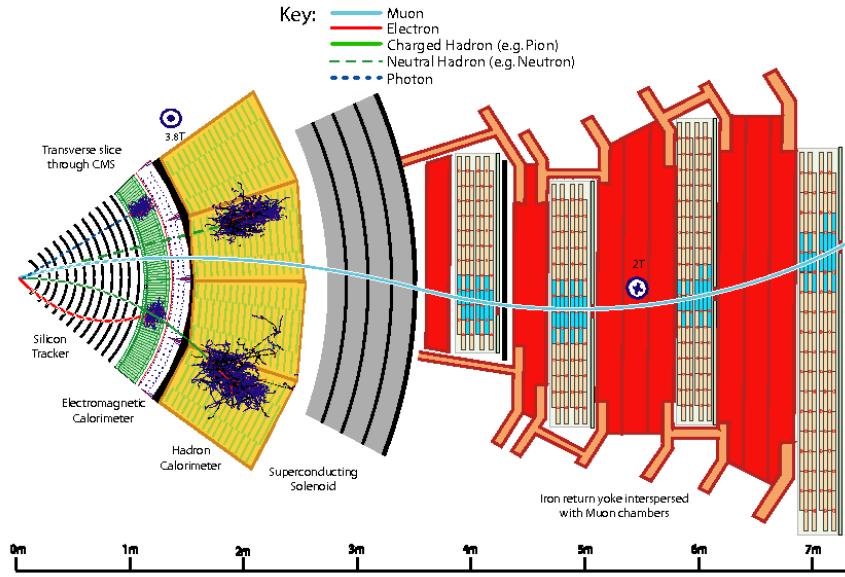


Figure 2-1: A sketch of a transverse slice of the CMS detector showing particle interactions from the interaction point to the muon detector. Reprinted from Reference [5].

vertices, calorimeter clusters, and muon segments described in Section 2.1. Based on their proximity in the η - ϕ , these PF elements are combined into muons, electrons, photons, and hadrons. Muon segments are combined with inner tracks to produce muons, inner tracks are combined with calorimeter clusters to produce electrons and charged hadrons, and calorimeter clusters are correlated to produce photons and neutral hadrons.

The PF algorithm reconstructs particles in regions of the detectors called blocks following the steps described in Section 2.2. After each step any PF elements associated to a PF candidate are removed from the block. For example, clusters associated with photons will not be used when reconstructing neutral hadrons. After all PF candidates are identified, they can be combined into event-wide variables such as jets and the missing transverse energy as described in Sections 2.2.5 and 2.2.6, respectively.

2.1 Particle Flow Elements

2.1.1 Tracks

The Combinatorial Track Finder software [6] is used to reconstruct tracks in an iterative inside-out process. Initial iterations search for tracks that are easy to find, e.g. those with high p_T , and hits associated with these tracks are removed for later iterations, reducing the combinatorial complexity and simplifying the search for more difficult tracks, e.g. greatly displaced ones.

The first step is to form seeds based on pixel hits, double strip hits containing 3D information, and an estimate of the beam spot. Earlier iterations require three pixel hits while later iterations gradually loosen the requirements. The final iterations specifically target increased muon tracking efficiency by including information from the muon chambers.

Next, a Kalman filter is used to find additional hits consistant with the evolution of the track seeds through the rest of the tracker, accounting for the magnetic field, energy loss due to ionization, and multiple scattering. The five parameters used for the helical trajectory evolution are the curvature ρ , the azimuthal angle ϕ_0 , the transverse impact parameter d_0 , the longitudinal impact parameter z_0 , and $\lambda = \cot \theta$, where θ is the polar angle.

After propagating the track through all layers of the detector and finding all associated hits, a Kalman fitter and smoother is used to refit the overall trajectory while a fourth-order Runge-Kutta method is used to extrapolate the trajectory between successive hits. To reduce the fraction of fake tracks, various quality requirements concerning the number of missing hits, the reduced χ^2 of the fit, and compatibility with a primary vertex are applied before proceeding to the next iteration.

Track reconstruction for electrons is more complicated as the Kalman filter is not a good description because of the high rate of non-Gaussian energy loss due to brehmsstrahlung these tracks experience within the tracker [5]. To improve the electron reconstruction efficiency, the electron seed collection is filled both by looking outside-in for ECAL superclusters consistant with track seeds and inside-out track

seeds consistent with superclusters. A Gaussian Sum Filter (GSF) defined to approximate the Bethe-Heitler energy-loss distribution is used to fit the trajectory of electron tracks.

2.1.2 Primary Vertexing

A deterministic annealing (DA) algorithm is used to associate tracks to primary vertices [6]. Tracks must pass additional requirements on the transverse impact parameter d_0 , the number of strip and pixel hits, and the reduced χ of the trajectory fit to be considered when finding primary vertices. The most probable vertex positions at an artificial temperature T are determined by the minimization of the “free energy”

$$F = -T \sum_i^{N_T} \ln \sum_j^{N_V} p_{ij} \rho_j \exp \left[-\frac{1}{T} \left(\frac{z_i^T - z_j^V}{\sigma_i^z} \right)^2 \right] \quad (2.1)$$

where the z_j^V are the vertex positions with weights ρ_j , the z_i^T and σ_i^z are the longitudinal impact parameters and the corresponding uncertainties of the tracks, and the p_{ij} are the probabilities of assigning the track i of N_T to the vertex j of N_V .

The DA algorithm starts with a single vertex at a very high temperature that is gradually decreased. The free energy F is minimized with respects to z_j^K at each new temperature and a vertex is split in two whenever T falls below its critical temperature

$$T_C^j = 2 \sum_i \frac{p_i p_{ij}}{(\sigma_i^z)^2} \left(\frac{z_i^T - z_j^V}{\sigma_i^z} \right)^2 \Bigg/ \sum_i \frac{p_i p_{ij}}{(\sigma_i^z)^2}. \quad (2.2)$$

The annealing procedure with vertex splitting continues down to $T = 4$ and the final assignment of tracks to vertices is performed at $T = 1$ without any further splitting. The vertex designated as *the* primary vertex (PV) of the hard scattering is the one which maximizes

$$S_T = \sum_i (p_T^i)^2 + (p_T^{\text{miss}})^2, \quad (2.3)$$

where p_T^i is the transverse momentum of a track assigned to the vertex and p_T^{miss} is the magnitude of the momentum imbalance in the transverse plane for the vertex.

2.1.3 Secondary Vertexing

Long-lived particles such as b hadrons and τ leptons often produce charged particles in their decays. These charged particles are traced to a secondary vertex at the location of the decay, which is identified by the inclusive vertex fitter (IVF) algorithm [7].

The IVF procedure begins by selecting seed tracks with a 2D impact parameter significance $\sigma_{d_0} \geq 1.2$ and a 3D impact parameter $\sqrt{d_0^2 + z_0^2} \geq 50 \mu\text{m}$. Tracks are assigned to a secondary vertex based on their opening angle with the seed track and distance at closest approach, with the additional stipulation that this distance be smaller for the secondary vertex than for the primary vertex.

To determine the precise position of the secondary vertices, the associated tracks are fitted with the adaptive vertex fitter and any vertices with a flight distance significance less than a certain threshold are discarded. At this point, a track is unassociated from a secondary vertex if the angular distance between the track and the secondary vertex flight direction is greater than 0.4 and if the track's distance at closest approach is larger than the magnitude of its impact parameter.

The secondary vertex position is refitted after track cleaning if there are still at least two tracks associated with the vertex. The last stage of cleaning removes a secondary vertex if it shares at least 20% of its tracks with another and the flight distance significance between the two is less than ten.

2.1.4 ECAL Superclusters

Due to the large amount of material in the tracker, electrons often emit bremsstrahlung photons, photons often convert to electron-positron pairs, and the brehmsstrahlung photons and converted electrons often undergo further conversion and brehmsstrahlung before reaching the ECAL. Because of the bending of electron trajectories in the magnetic field, the resulting electromagnetic (EM) shower is significantly spread in the ϕ -direction and collimated in the η -direction. The ECAL reconstruction algorithm [8, 9] combines the basic cluster from each showered particle into a supercluster representing the initial electron or photon from the hard scattering.

The first step is the identification of a seed crystal with greater transverse energy than its immediate neighbors and above a predefined minimum threshold. The energy of each crystal is determined from calibration constants combined with the amplitude and peak time obtained by fitting the pulse shape of the ten time samples surrounding the triggering bunch crossing.

In the barrel, a supercluster starts with a 5×1 array of crystals in the $\eta\text{-}\phi$ plane centred on the seed crystal. The array is extended around the seed crystal in the ϕ -direction up to $|\Delta\phi| \leq 0.3$ if the energy of the additional crystals exceeds a certain threshold. The contiguous array is grouped into distinct basic clusters each containing a seed array with energy greater than another threshold. The supercluster is the collection of basic threshold found in the $\eta\text{-}\phi$ region centered on the initial seed crystal. Since the crystals in the endcaps are arranged in an $x\text{-}y$ grid, clustering here uses fixed 5×5 matrices of crystals. After a seed cluster is identified, additional, partially overlapping 5×5 matrices are added if their centroid lies within $|\Delta\eta| \leq 0.07$ and $|\Delta\phi| \leq 0.3$. For unconverted photons, both methods produce superclusters that are simple 5×5 matrices.

2.1.5 HCAL Clusters

The purpose of clustering in the HCAL is to measure the energy and direction of neutral hadrons, disentangle neutral hadrons from charged hadron energy deposits, and improve the energy measurement for charged hadrons with poorly reconstructed tracks [5]. Similar to the supercluster algorithm, a cluster in the HCAL is first identified by a seed cell with greater transverse energy than its immediate neighbors and above a predefined minimum threshold. This seed is then grown into a topological cluster by adding cells with at least a corner in common with a cell already in the cluster and energy above twice the noise threshold.

An iterative Gaussian mixture model is used to break each topological cluster of M individual cells is broken into N energy deposits corresponding to individual particles, where N is the number of seeds. Each energy deposit is modeled as a Gaussian distribution \mathcal{N} with amplitude A_i , mean $\vec{\mu}_i$ in the $\eta\text{-}\phi$ plane, and width

σ fixed by the calorimeter resolution. The expected fraction f_{ji} of the energy E_j measured in the cell at position \vec{c}_j from the i th energy deposit is

$$f_{ji} = \frac{\mathcal{N}(\vec{c}_j|A_i, \vec{\mu}_i, \sigma)}{\sum_k^N \mathcal{N}(\vec{c}_j|A_k, \vec{\mu}_k, \sigma)}. \quad (2.4)$$

The amplitude and position of each energy deposit are determined by an analytical maximum-likelihood fit to be

$$A_i = \sum_j^M f_{ji} E_j \quad \left| \quad \vec{\mu}_i = \sum_j^M f_{ji} E_j \vec{c}_j \right. \quad (2.5)$$

where the initial values are the energy and position of the seeds. The process of calculating energy fractions f_{ji} and fitting for the amplitudes A_i and positions $\vec{\mu}_i$ is repeated until convergence, at which point they are taken as the cluster parameters.

2.1.6 Muon Segments

Muon segments are reconstructed from the hits in the muon chambers using a Kalman filter in a similar manner to that described for the inner tracker in Section 2.1.1. A full track constructed in this way is referred to as a standalone muon [5].

2.1.7 Isolation

While not a physics object persay, isolation is a key concept of the PF algorithm that distinguishes prompt leptons and photons originating in the hard scattering from those originating in the decays of hardrons during the parton shower. The latter are surrounded by a large amount of additional hadrons while the former have little hadronic activity in their vicinity, originating mainly from the pileup vertices.

The isolation of a prompt object is the total amount of energy due to additional particles within an annulus of radius $0.01 < \Delta R < 0.4$ around the prompt object, where the lower bound avoids including the prompt object and its radiation in the sum. The isolation is calculated using either the raw energy deposits in the subdetectors or the four-momenta of the PF candidates surrounding the prompt object

depending the stage of the PF algorithm. Prompt objects are required to have an isolation value below a certain threshold, rejecting hadrons misidentified as leptons and photons as well as non-prompt leptons and photons.

The isolation calculation is usually split into three different components based on the types of particles that contribute energy. The photon isolation I_γ is the E_T sum of the PF photons defined in Section 2.2.3 while the charged hadron and neutral hadron isolations I_{CH} and I_{NH} are the p_T sums of the PF charged and neutral hadrons defined in Section 2.2.4, with the additional stipulation that charged hadrons be associated with the primary vertex.

In events with very few tracks, such as one with a single high p_T photon and a large momentum imbalance, it is possible that the identified primary vertex does not correspond to the pp interaction from which the photon object originates because the photon does not figure into the primary vertex calculation from Section 2.1.2. In such cases, the photon object can be surrounded by charged hadrons and still appear isolated under the standard charged hadron isolation. A conservative measure to address such misidentification is to replace I_{CH} with the maximum of the PF charged hadron isolations computed over all reconstructed vertices, e.g. the maximum charged hadron isolation $I_{\text{CH}}^{\max} = \max_{\text{vertices}} I_{\text{CH}}$.

To reduce the pileup dependence of these variables, the median energy density ρ of the pileup interactions in the isolation cone is calculated using the effective areas given in Table 2.1 and subtracted from each isolation sum. Additionally, since the rate of the charged particles originating from pileup interactions is about twice as large as the corresponding rate of the neutral particles, the pileup isolation I_{PU} is defined as the half the p_T sum of the PF charged hadrons *not* associated with the primary vertex. Often selections are placed on the individual isolation components when selecting prompt photons, while the relative combined PF isolation

$$I_{\text{PF}}^{\text{rel.}} = \left(I_{\text{CH}} + \max \left\{ 0, I_{\text{NH}} + I_\gamma - I_{\text{PU}} \right\} \right) / p_T^\ell \quad (2.6)$$

is used when selecting prompt leptons.

Isolation	$ \eta < 1.0$	$1.0 < \eta < 1.479$
I_{CH}	0.0360	0.0377
I_{NH}	0.0597	0.0807
I_{γ}	0.1210	0.1107
I_{CH}^{\max}	0.01064	0.1026

Table 2.1: Effective areas for isolations.

2.2 Particle Identification

2.2.1 Muons

The first step of the PF algorithm reconstructs three types of muon candidates: standalone muons, outside-in global muons, and inside-out tracker muons. To construct a global muon, the algorithm identifies an inner track consistent with the trajectory of a standalone muon evolved inwards using a Kalman filter similar to those discussed in Section 2.1.1. After finding a match, a global muon candidate is created by combining the inner track with the standalone track with a second Kalman filter. Conversely, to construct a tracker muon, the algorithm identifies a muon segment consistent with the trajectory of a inner track with $p_T > 0.5 \text{ GeV}$. Global and tracker muons sharing the same inner track are merged into a single candidate. For muons with $p_T < 200 \text{ GeV}$, the muon momentum is that of the inner track, while the momentum is determined from a global fit of the muon chambers and inner tracker for muons with momentum above this threshold.

Hadrons misidentified muons are rejected through two separate mechanisms. First, the isolation with respects to inner tracks and calorimeter deposits within $\Delta R < 0.3$ is required to be less than 10% of the muon p_T . Non-isolated muons are kept only if certain selections on the reduced χ^2 of the track fit and the two impact parameters d_0 and d_z are satisfied. Finally, misidentified or misreconstructed muons can lead to a spurious imbalance in the transverse momentum. The procedure used to identify and remove these muon candidates is described in Section 2.2.6. The total efficiency of muon reconstruction is 99%.

The work described in this thesis only considers global muons with $p_T > 10 \text{ GeV}$

and $|\eta| < 2.5$. This minimum requirement is only used to reject events containing a muon and is referred to as the veto muon ID. The loose muon ID adds the requirement that the relative combined PF Isolation $I_{\text{PF}}^{\text{rel}}$ must be less than 0.25. must be less than 0.25. In order for a muon to pass the tight ID, it must have $p_T > 30 \text{ GeV}$ and $I_{\text{PF}}^{\text{rel}} < 0.15$ as well as satisfying the additional requirements in Table 2.2.

Variable	Selection	Description
$\chi^2_{\text{track fit}}/N_{\text{dof}}$	< 10	quality of global-muon track fit
$N_{\text{hit}}^{\mu\text{on}}$	> 1	at least one muon-chamber hit
$N_{\text{station}}^{\mu\text{on}}$	> 2	segments in at least two muon stations
d_0	$< 2 \text{ mm}$	reject cosmic ray muons
d_z	$< 5 \text{ mm}$	reject muons from pileup
$N_{\text{hit}}^{\text{pixel}}$	> 1	at least one pixel hit
$N_{\text{hit}}^{\text{tracker}}$	> 5	more than five tracker layers with hits

Table 2.2: Selections for the tight muon ID.

2.2.2 Electrons

Electron candidates are seeded from the GSF tracks described in Section 2.1.1 as long as the corresponding ECAL clusters are not linked to three or more additional tracks. In each block, all ECAL clusters linked to either the supercluster (SC) or one of the GSF track tangents are associated with the candidate to ensure optimal energy containment. Additional tracks linked to these clusters are associated if the track momenta and energies of any linked HCAL clusters are compatible with the electron hypothesis. Any tracks and clusters belonging to identified photon conversions linked to the GSF track tangents are associated as well.

To recover any energy lost during the association process, the total energy of the collected clusters is corrected with analytical functions of E and η . For ECAL-based candidates, the sum of the energies measured in the HCAL cells within $\Delta R < 0.15$ of the supercluster must be less than 10% of the supercluster energy. The final energy of an electron candidate is a weighted average of the corrected ECAL energy and the momentum of the GSF track and the electron direction is that of its GSF track.

The work described in this thesis only considers electrons with $p_T > 10 \text{ GeV}$ and $|\eta| < 2.5$. Additionally, electrons must pass further cuts on the observables listed in Table 2.3. The exact values of the cuts are tuned based on whether the electron is in the barrel or the endcap and to give desired signal efficiencies and background acceptance. The loose ID is tuned to 90% signal efficiency and 0.5% background acceptance, while the tight ID is tuned to 70% signal efficiency and 0.1% background acceptance.

Variable	Description
$\sigma_{inj\eta}$	energy-weighted cell width in the η -direction of the SC
$\Delta\eta$ and $\Delta\phi$	angular separation between the SC seed and the GSF track
H/E	energy ratio of the corresponding ECAL and HCAL towers
$I_{\text{PF}}^{\text{rel.}}$	relative combined PF Isolation
$ 1/E - 1/p $	difference between calorimeter energy and tracker momentum
$N_{\text{hit}}^{\text{miss}}$	number of missing hits in the inner tracker
Conversion veto	presence of tracks originating from a converted photon

Table 2.3: Variables used in selecting electrons.

2.2.3 Isolated Photons

Photon candidates are seeded from the ECAL superclusters (SCs) described in Section 2.1.4 as long as they have no links to GSF tracks and $E_T > 10 \text{ GeV}$. The same cluster and track association process described for electrons in Section 2.2.2 is used for photons, with the photon energy and direction being that of the final supercluster. This is motivated by the observation that the additional energy corrections used to improve the photon energy resolution cause photon candidates with large cluster width to exhibit unphysical energies. Figure 2-2 is a profile of the magnitude of the energy correction in bins of $\sigma_{inj\eta}$, the energy-weighted cell width in the η -direction.

As an illustration, an unphysically large correction is causing the transverse momentum of the photon object in the event shown in Fig. 2-3 to be nearly twice as large as the transverse momentum imbalance, which is supposed to balance the visible, i.e., photon momentum. Photon candidates with wide showers are used to estimate the hadron-to-photon misidentification background, while the photon energy resolution

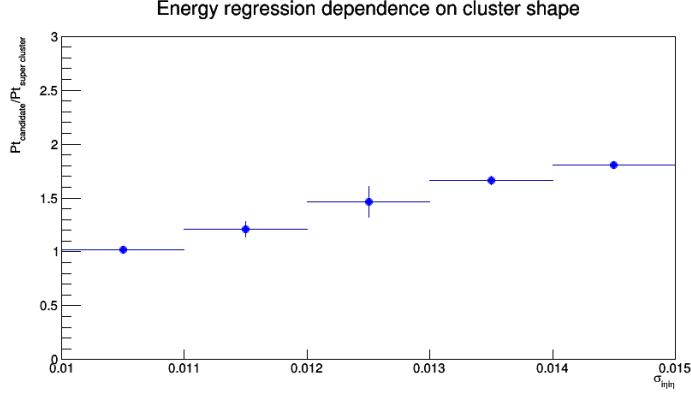


Figure 2-2: Magnitude of the energy correction on the photon object in bins of σ_{inj} .

has an insignificant effect. Therefore, the unbiased supercluster energy was chosen over the corrected photon energy.

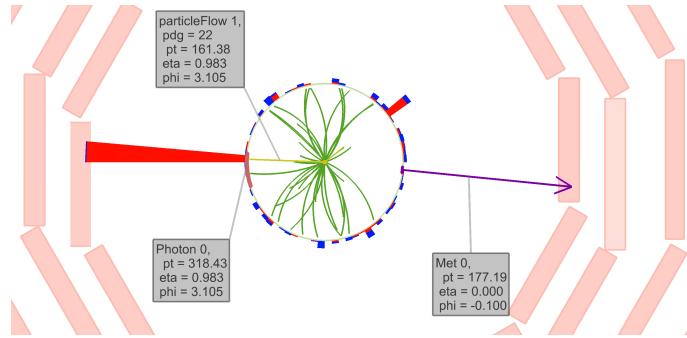


Figure 2-3: An example event where a photon with a wide shower receives a large energy correction.

For the work shown in this thesis, we are only concerned with high- E_T photons from the ECAL Barrel that have a supercluster with $E_T > 175 \text{ GeV}$ and $|\eta| < 1.4442$. To reduce hadron-to-photon misidentification rate, we apply the collection of isolation and shower shape selections in Table 2.4, which will hereby be referred to as the e/γ ID. To reject electrons from the candidate sample, no electron track seeds in the pixel detector can be associated to the supercluster. This is known as the pixel seed veto. To clean the candidate sample from photon objects originating from non-collision sources, we apply the collection of cuts shown in Table 2.5, which combined with the pixel seed veto constitutes the γ -specific ID. The beam halo tagger E_{MIP} is the total

energy deposited in ECAL by a hypothetical beam halo muon that passes through the photon cluster. See Section ?? for more detail on beam halo processes. The lower bounds on $\sigma_{in\eta\eta}$ and $\sigma_{i\phi i\phi}$ as well as the requirement on the cluster seed time $|t_{seed}|$ are employed to reject spurious photon objects arising from the “ECAL spikes” discussed in Section ??.

Variable	Maximum Value
H/E	0.0260
$\sigma_{in\eta\eta}$	0.01040
ρ -corrected I_{CH}^{\max}	1.146
ρ -corrected I_{NH}	$2.792 + 0.0112 \times E_T^\gamma + 0.000028 \times (E_T^\gamma)^2$
ρ -corrected I_γ	$2.176 + 0.0043 \times E_T^\gamma$

Table 2.4: Selections for the e/γ portion of the photon ID. Isolation values and E_T^γ are all in units of GeV.

Variable	Selection	Description
E_{MIP}	< 4.9 GeV	ECAL energy from a hypothetical beam halo muon
$\sigma_{in\eta\eta}$	> 0.001	non-trivial shower width in the η -direction
$\sigma_{i\phi i\phi}$	> 0.001	non-trivial shower width in the ϕ -direction
$ t_{seed} $	< 3 ns	timing of the cluster seed relative to bunch crossing

Table 2.5: Additional selections beyond the pixel seed veto for the γ -specific portion of the photon ID.

2.2.4 Hadrons

The last candidates reconstructed in a given block are the charged and neutral hadrons from fragmentation and hadronization, as well as the non-isolated muons and photons produced from their respective decays.

Inside the tracker acceptance of $|\eta| < 2.5$, all trackless HCAL clusters are reconstructed as neutral hadrons while all trackless ECAL clusters are reconstructed as photons. The preference towards photons is motivated because photons carry 25% of the jet energy and neutral hadrons do not interact strongly with the ECAL. Conversely, outside of the tracker acceptance, it is no longer possible to distinguish

charged and neutral hadrons, so any ECAL clusters linked to HCAL clusters are assumed to arise from unidentified charged hadrons. Thus, only unlinked ECAL clusters are reconstructed as photons and linked ECAL and HCAL clusters are reconstructed as neutral hadrons.

Afterwards, the only remaining PF elements are HCAL clusters linked to one or more tracks and ECAL clusters linked to one of these tracks. A single charged hadron is constructed for each remaining HCAL cluster, with energy equal to the sum of the ECAL and HCAL clusters and momentum equal to the sum of the individual track momenta.

If the energy of the charged hadron exceeds its momentum by an amount larger than the calometric energy resolution, neutral hadrons and photons are added. For excesses greater than 500 MeV, a photon with energy equal to the excess is created. If this photon cannot explain the entire excess, e.g. the excess is larger than the ECAL energy by at least 1 GeV, the remainder is identified as a neutral hadron. After photons and neutral hadrons consume the excess calometric energy, charged hadrons are constructed from the linked tracks with their energy and momentum determined by the track momenta under the charged-pion hypothesis.

If energy and momentum of the charged hadron are compatible, no neutral particles are identified. A charged hadron candidate is created for each track linked to the HCAL cluster, with momenta determined by a χ^2 fit of the tracker and calorimeter measurements. This combination ensures a smooth transition between the tracker-dominated low-energy regime and the calorimeter-dominated high-energy regime while always improving the final energy resolution.

If the momentum of the charged hadron exceeds its energy by three standard deviations, new PF muons are made from any non-isolated global muons failing the cleaning described in Section 2.2.1 with momentum resolution better than 25%. If, after masking the tracks from these muons, the track momentum sum still greatly exceeds the calorimeter energy, all remaining tracks with a p_T uncertainty greater than 1 GeV are identified, sorted in decreasing order of this uncertainty, and sequentially masked until no such tracks remain or the momentum excess disappears, whichever

comes first. At this point, the HCAL cluster is reconstructed according to one of the procedures defined in the preceding paragraphs.

When three or more charged particle candidates are linked to a secondary vertex identified as described in Section 2.1.3, a single primary charged hadron with energy equal to the sum of their energies replaces them in the reconstructed particle list. If an incoming track is associated with the vertex, it determines the direction of the primary charged hadron, which is otherwise determined by the vectorial sum of momenta of the secondary particles. If the momentum of the incoming track is well measured, the energy of undetected secondary particles is estimated and added to the energy of the primary charged particle.

2.2.5 Jets

As discussed in Section ??, jets are produced during the fragmentation and hadronization of colored particles produced in the hard scattering. After all PF candidates have been identified, a sequential recombination algorithm is used in an attempt to cluster these jets. Given an object i in the event E , we define the distance to the beam and the distance to another object j to be

$$d_{iB} = (p_T^i)^{2q} \quad \left| \quad d_{ij} = \min \left\{ (p_T^i)^{2q}, (p_T^j)^{2q} \right\} \frac{(\Delta R_{ij})^2}{R^2}, \right. \quad (2.7)$$

respectively, where q and R are tunable parameters and ΔR_{ij} is the angular distance between the two particles. The distance parameter R is an approximate measure of the cone size ΔR of the jet, while the power of the energy scale q defines the relationship between the relationship between the momentum and angular factors. Jets clustered with $q = -1$ are referred to as anti- k_T jets, those with $q = 0$ as Cambridge-Aachen jets, and those with $q = 1$ as k_T jets. Negative values of q force the clustering of circular jets around hard seeds ensuring that the resulting jet boundaries are resilient with respect to soft radiation. Within CMS, anti- k_T jets with $R = 0.4$ are used to cluster the parton shower from single partons.

The implementation in the FastJet library reduces the computational complex-

ity of clustering from $\mathcal{O}(N^2)$ to $\mathcal{O}(N \log N)$ for jets with hundreds or thousands of constituent particles. First, the two objects i and j with the smallest distance d_{ij} between them are found. If d_{ij} is less than both d_{iB} and d_{jB} , they are removed from E and a single object k with four-momentum $p_\mu^k = p_\mu^i + p_\mu^j$ which is added in their place. Otherwise if $d_{iB} < d_{jB}$, object i is removed from E and added to the set of jet candidates J while object j is kept, and vice versa if $d_{jB} < d_{iB}$. This procedure continues until all objects are removed from E and J contains all possible jet candidates.

2.2.6 Missing Tranverse Energy

The production of neutrinos and dark matter candidates produces a momentum imbalance in the transverse plane. The missing transverse momentum \vec{p}_T^{miss} is defined as the negative vectorial sum of all the PF candidates in the event E such that

$$\vec{p}_T^{\text{miss}} = - \sum_{i \in E} \left(\hat{x} \cdot p_T^i \cos \phi + \hat{y} \cdot p_T^i \sin \phi \right), \quad (2.8)$$

and its magnitude is the missing transverse energy $E_T^{\text{miss}} = |\vec{p}_T^{\text{miss}}|$. In a perfectly reconstructed event, non-zero E_T^{miss} implies the presence of neutrinos or DM candidates; however, the failure to properly reconstruct energy deposits or the reconstruction of PF candidates with incorrect energy results in events with large amount of fake E_T^{miss} .

One last cleaning of the PF candidates is conducted in an attempt to fix these events. To remove muons from cosmic rays, muon candidates with trajectories more than 1 cm away from the beam axis are removed if the measured E_T^{miss} is consequently reduced by half. For muons with $p_T > 20 \text{ GeV}$, the choice of subdetector used to estimate momemtum is reviewed and the smallest available estimate used if it reduces the measured E_T^{miss} by half. Additionally, the assignment of charged hadrons and neutral hadrons is reconsidered to ensure a charged hadron is not reconstructed as a muon and neutral hadron and vice versa.

Fake E_T^{miss} can persist in an event even after the final cleaning of PF candidates. At this point, events are checked against a known set of filters identifying possible sources of fake E_T^{miss} not captured by the PF algorithm. One set of filters is the HCAL

and ECAL filters that identify events with calorimeter clusters caused by noise from the shape and timing of the energy distribution. Another such filter is the beam halo filter that identifies energy deposits from muons produced from interactions between the beam and the machine that travel parallel to the beam. These muons are identified by their localization in ϕ and a longitudinal track left in the ECAL endcaps and the CSCs. Applying these filters removes essentially all remaining events with fake E_T^{miss} while rejecting less than 1% of events with real E_T^{miss} .

2.2.7 ECAL gain-switch effect

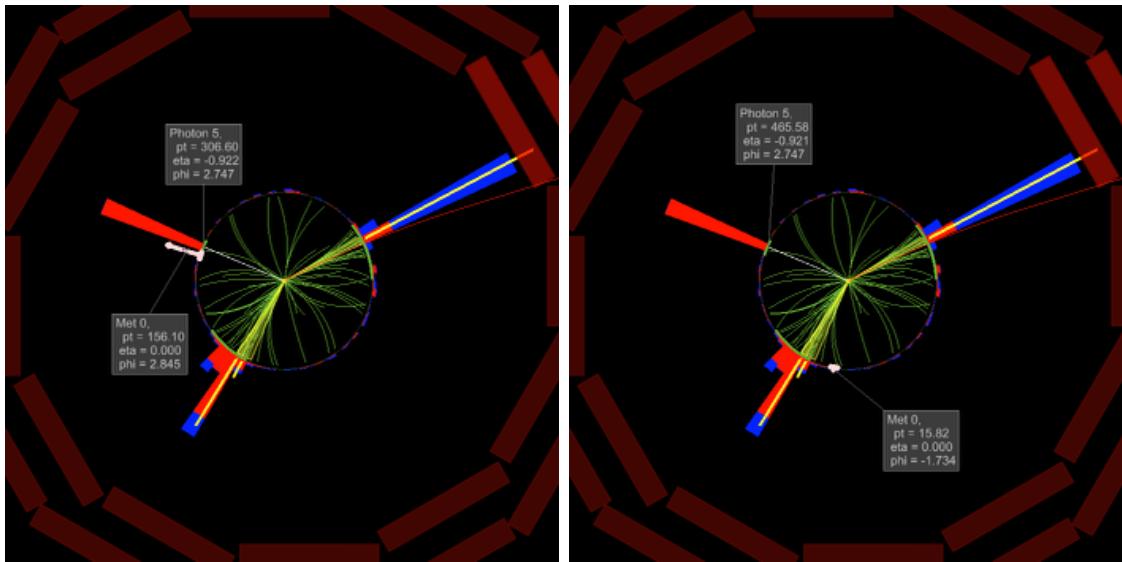


Figure 2-4: Two event displays comparing the same event, reconstructed without (left) and with (right) the fix for ECAL gain-switch effect.

The “multi-fit” algorithm for ECAL hit reconstruction was found to have an unexpected behavior when there is a large energy deposit onto a single ECAL crystal, such that the electronic signal converted at the frontend electronics is sourced partially from channels of the preamplifier with lower gains (6 or 1) than the default (12) channel. In the most dramatic cases, pulse misreconstruction would result in underestimation by hundreds of GeV of photon p_T . This effect is mitigated in the reprocessed data set used for this analysis by identifying ECAL clusters whose seed crystal hit had a switch of gains, and performing an alternative pulse reconstruction

when possible.

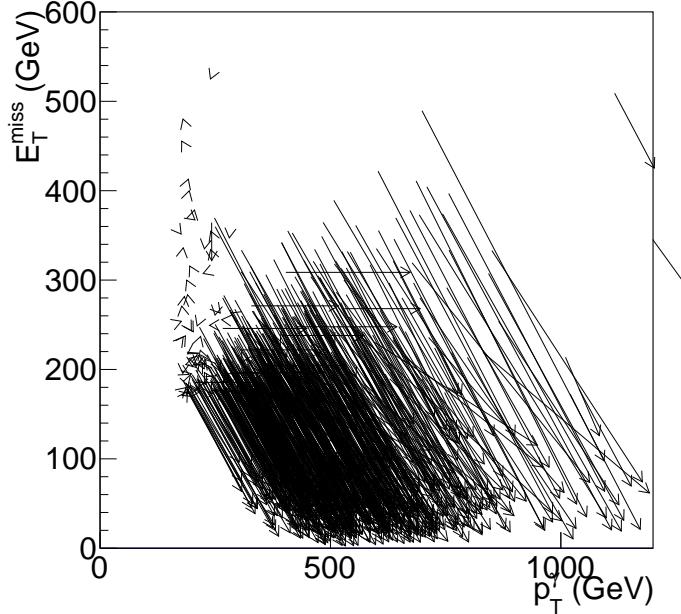


Figure 2-5: The change in reconstructed photon p_T and E_T^{miss} for events in the bin $\Delta\phi(\gamma, p_T^{\text{miss}}) < 0.05$ of the distribution in Figure ???. Each arrow represents a single event, the tail (head) of the arrow corresponding to $(E_T^\gamma, E_T^{\text{miss}})$ coordinates in the datasets without (with) the fix for the gain-switch problem.

The gain-switch problem affected the analyses documented in this thesis, since large underestimation of the energy of a photon in an otherwise typical $\gamma+\text{jets}$ event would introduce large missing transverse momentum to the event, typical collinear to the affected photon. Figures 2-4 and 2-5 are the visualization of how the new dataset changes the reconstructed photon energy and E_T^{miss} .

Bibliography

- ¹S. Chatrchyan et al., “The CMS experiment at the CERN LHC”, Journal of Instrumentation **3**, S08004–S08004 (2008).
- ²*CMS detector design*, (2019) <http://cms.web.cern.ch/news/cms-detector-design> (visited on 11/23/2011).
- ³S. Agostinelli et al., “Geant4ffffdffffda simulation toolkit”, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **506**, 250–303 (2003).
- ⁴J. Allison et al., “Geant4 developments and applications”, IEEE Transactions on Nuclear Science **53**, 270–278 (2006).
- ⁵A. Sirunyan et al., “Particle-flow reconstruction and global event description with the CMS detector”, Journal of Instrumentation **12**, P10003–P10003 (2017).
- ⁶T. C. Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker”, Journal of Instrumentation **9**, P10009–P10009 (2014).
- ⁷A. Sirunyan et al., “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV”, Journal of Instrumentation **13**, P05011–P05011 (2018).
- ⁸T. C. Collaboration, “Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV”, Journal of Instrumentation **10**, P06005–P06005 (2015).

⁹T. C. Collaboration, “Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV”, Journal of Instrumentation **10**, P08010–P08010 (2015).