

**Searching for Dark Matter with the CMS Detector  
in proton-proton collisions containing  
a single high- $p_T$  photon and large  $E_T^{\text{miss}}$**

by

Brandon Leigh Allen

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



**Searching for Dark Matter with the CMS Detector  
in proton-proton collisions containing  
a single high- $p_T$  photon and large  $E_T^{\text{miss}}$**

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 Philosophy 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

<b>1</b>	<b>Dark Matter</b>	<b>9</b>
1.1	Astrophysical Evidence . . . . .	10
1.1.1	Galactic Clusters . . . . .	10
1.1.2	Galactic Rotation Curves . . . . .	11
1.1.3	Gravitational Lensing . . . . .	12
1.1.4	Cluster Collisions . . . . .	13
1.2	Relic Density . . . . .	16
1.3	Dark Matter Candidates . . . . .	18
1.3.1	Weakly-Interacting Massive Particles . . . . .	19
1.3.2	Axions . . . . .	19
1.3.3	Sterile Neutrinos . . . . .	20
1.4	Simplified Models for LHC . . . . .	22
1.5	Non-collider Searches . . . . .	24
1.5.1	Direct Detection . . . . .	24
1.5.2	Indirect Detection . . . . .	27
1.6	Summary . . . . .	29
<b>2</b>	<b>The Large Hadron Collider</b>	<b>31</b>
2.1	Experimental Apparatus . . . . .	31
2.2	Collider Phenomenology . . . . .	33
2.2.1	Parton Distribution Functions . . . . .	35
2.2.2	Hard Scattering . . . . .	36
2.2.3	Parton Shower . . . . .	37

2.2.4	Hadronization . . . . .	38
<b>3</b>	<b>The CMS Detector</b>	<b>39</b>
3.1	Inner Trackers . . . . .	41
3.2	Electromagnetic Calorimeter . . . . .	43
3.2.1	Preshower Detector . . . . .	45
3.3	Hadronic Calorimeter . . . . .	46
3.4	Muon Chambers . . . . .	48
3.5	Online Trigger System . . . . .	50
3.6	Detector Simulation . . . . .	53
<b>4</b>	<b>Global Event Reconstruction</b>	<b>55</b>
4.1	Particle Flow Elements . . . . .	57
4.1.1	Tracks . . . . .	57
4.1.2	Primary Vertexing . . . . .	58
4.1.3	Secondary Vertexing . . . . .	59
4.1.4	ECAL Superclusters . . . . .	59
4.1.5	HCAL Clusters . . . . .	60
4.1.6	Muon Segments . . . . .	61
4.1.7	Isolation . . . . .	62
4.2	Particle Identification . . . . .	63
4.2.1	Muons . . . . .	63
4.2.2	Electrons . . . . .	65
4.2.3	Isolated Photons . . . . .	66
4.2.4	PF Hadrons . . . . .	68
4.2.5	Jets . . . . .	70
4.2.6	Missing Transverse Energy . . . . .	71
4.2.7	ECAL gain-switch effect . . . . .	72
<b>Bibliography</b>		<b>75</b>

# Chapter 1

## Dark Matter

As a theory of the fundamental particles and forces of nature, the Standard Model should also help explain physics at the largest scales. The  $\Lambda$ CDM model [**Condon2018**] best explains all current cosmological observations including the structure of the cosmic microwave background; the abundances of hydrogen, helium, and lithium; the large-scale structure in the distribution of galaxies, and the accelerating expansion of the universe. However, results from many cosmological observations show that baryonic matter (matter consisting of various combinations of protons, neutrons, and electrons) only contributes  $\sim 5\%$  of the total energy of the universe, with radiation (photons and relativistic neutrinos) contributing less than a hundredth of a percent.

The remaining 95% of energy comes from just two sources:  $\sim 27\%$  from non-relativistic non-baryonic matter referred to as dark matter and  $\sim 68\%$  from an unknown form of energy that permeates all of space referred to as dark energy. Current observations show that dark energy is uniform in space and time producing a similar effect to that of the cosmological constant in the Einstein field equations of general relativity. Not much else is known about dark energy, although there are many experiments attempting to discover additional properties. The work in this thesis shall focus on trying to explain dark matter.

Dark matter cannot be explained by the 17 particles of the Standard Model, yet its gravitational effects have been observed in many circumstances. The rest of this chapter will cover the astrophysical evidence for dark matter (Sections 1.1 and 1.2),

various dark matter candidates (Section 1.3) and the models investigated at the LHC (Section 1.4), and non-collider searches for dark matter (Section 1.5).

## 1.1 Astrophysical Evidence

All existing evidence for dark matter comes from astrophysical observations of its gravitational effects on the universe at various length scales. We shall focus on four different sources of evidence: the average velocity of galaxies in clusters, the rotation curves of spiral galaxies, strong gravitational lensing, and merging galactic clusters. The evidence presented here is not exhaustive, see Reference [Roos2010] for more detail.

### 1.1.1 Galactic Clusters

Galactic clusters are the largest gravitational bound systems, with the orbital velocities of the individual clusters determined by the total gravitational mass of the cluster. Applying the Virial Theorem gives the explicit relation

$$v^2 = \frac{GM}{2r}, \quad (1.1)$$

where  $v$  is the average orbital velocity of a galaxy in the cluster,  $r$  is the average separation between galaxies in the cluster,  $M$  is the total gravitational mass of the cluster, and  $G$  is the Newtonian constant of gravitation. In 1933, Fritz Zwicky measured the average orbital velocity of the Coma cluster and using Equation 1.1 calculated that its mass was a factor of ten larger than the observed visible mass, leading to the conclusion that the majority of the cluster consisted of non-luminous matter [Zwicky1933]. Today studies show that stars only contribute 1% of the total cluster mass, with a hot, baryonic intracluster medium and dark matter contributing the remaining 14% and 85% of the total cluster mass, respectively [Roos2010].

### 1.1.2 Galactic Rotation Curves

Spiral galaxies are stable gravitational bound systems with stars and interstellar gas rotating around the galactic center in nearly circular orbits in a single plane. For these galaxies, the orbit of an individual star is stable when the gravitational force acting on the star balances the centripetal acceleration of the star. With this condition, the expected stellar velocity  $v$  is a function of distance  $r$  from the galactic center given by

$$v = \sqrt{\frac{GM(r)}{r}}, \quad (1.2)$$

where  $M(r)$  is the total gravitational mass inside radius  $r$ . Thus, past a certain critical radius  $r_c$ , the stellar velocity should fall with as  $r^{1/2}$  as the mass of the galaxy is no longer increasing significantly [**Damico2009**].

In 1980, Vera Rubin and Kent Ford observed that instead of decreasing at distances outside the visible galaxy, the stellar velocity stayed constant out to a very great distance, necessitating an additional non-luminous source of mass [**Rubin1980**]. The most common explanation for this missing mass is the existence of an isotropic dark matter halo surrounding the galaxy. With the inclusion of interstellar gas, the total mass inside radius  $r$  is given by

$$M(r) = 4\pi \int_0^r dr' (r')^2 [\rho_S(r') + \rho_g(r') + \rho_{DM}(r')], \quad (1.3)$$

where  $\rho_S$ ,  $\rho_g$ , and  $\rho_{DM}$  are the density profiles of the stars, interstellar gas, and dark matter in the galaxy, respectively. Once these densities have been specified, it is possible to plot the fraction of the total stellar velocity due to each mass source as a function of distance from the galactic center. Figure 1-1 shows the results of doing this using the observed stellar and interstellar mass density profiles and the expected density from an isotropic dark matter halo for two different spiral galaxies. In both cases, this reproduces the observed flat galactic rotation curve incredibly well, lending strong support for the existence of galactic dark matter halos.

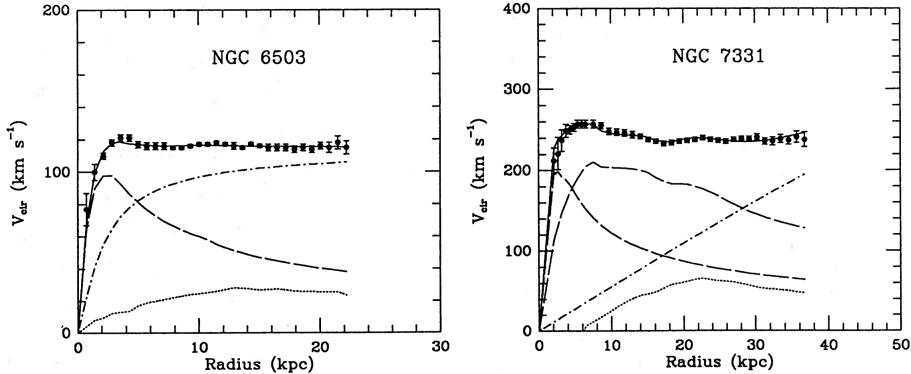


Figure 1-1: The observed (points) and fitted (solid line) rotation curves for two sample galaxies. The fit consists of three components: the stellar component (dashed), the interstellar gas (dotted), and the dark matter halo (dash-dotted). Reprinted from Reference [Begeman1991].

### 1.1.3 Gravitational Lensing

As a consequence of Einstein's equivalence principle, a massive body will deflect light, a phenomenon known as gravitational lensing. In the language of general relativity, this means that the photons take the path given by the geodesic lines following the curvature of space-time due to the massive body. For most observations of gravitational lensing due to astrophysical bodies, the physical size of the lensing object is much smaller than the distance between observer, lens, and source allowing us to use the thin lens approximation. Approximating the lens as a planar distribution of matter, the angular deflection is given by

$$\vec{\alpha}(\vec{x}) = \frac{4G}{c^2} \int d^2x' \frac{\vec{x} - \vec{x}'}{|\vec{x} - \vec{x}'|^2} \int dz \rho(\vec{x}', z) \quad (1.4)$$

where  $\vec{x}$  is a two-dimensional vector in the plane of the lens,  $z$  is the perpendicular distance from the plane of the lens, and  $\rho$  is the three dimensional density [Massey2010]. If the source is treated as a point mass, this reduces to

$$\alpha = \frac{4G}{c^2} \cdot \frac{M}{b} \quad (1.5)$$

where  $b = |\vec{x} - \vec{x}'|$  is the impact parameter and  $M$  is the total mass of the object. Thus, measuring the angle of deflection due to gravitational lensing around an astrophysical object provides an independent measurement of the total mass of the body which can be compared to the mass of the luminous objects in the body.

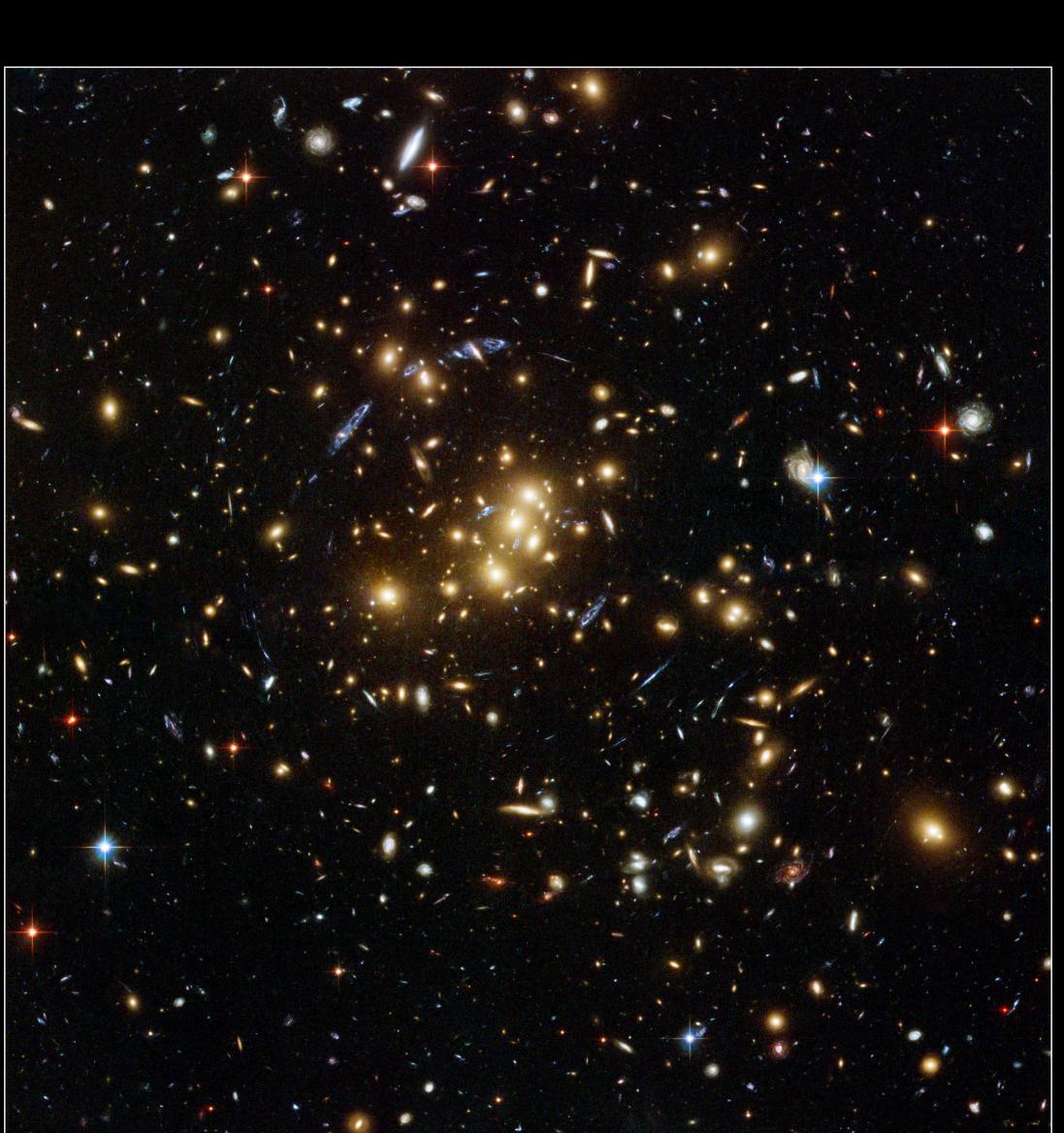
Depending on the mass of the deflecting body and impact parameter, the size of deflection can fall into three different regimes. The first of these is called strong lensing where the curving of space-time is so strong that light can travel multiple paths around the lens and still reach the observer. If the source is directly behind a circular lens, light travels around all sides of the lens and appears as an Einstein ring, while if the source is offset or the lens is non-circular, the source will instead appear in multiple locations as if viewed from slightly different angles. An example of strong lensing is shown in Figure 1-2.

The next regime is known as weak lensing, where the deflection is enough to distort the image of the source but not enough to result in multiple images. The shear of this distortion can be converted into a map of the projected mass distribution. True weak lensing results in circular “*E-mode*” patterns while sources of systematic uncertainty produce both “*E-mode*” and curl-like “*B-mode*” patterns. Thus, requiring a zero “*B-mode*” contribution assures that the measured mass distribution is accurate. Figure 1-3 shows the observed shear of half a million galaxies measured in the Hubble Space Telescope COSMOS survey.

The final regime is the microlensing that occurs when a lens moves relative to a luminous source. As the lens passes in front of the source, it will temporarily increase the apparent luminosity of the source, enabling a mass measurement of the lens. Microlensing results show that rocky exoplanets orbit other stars and that these planets cannot form the bulk of dark matter in the Milky Way.

#### 1.1.4 Cluster Collisions

Gravitational lensing measurements of galactic cluster collisions provide support for dark matter and help constrain its properties. Figure 1-4 shows the merging cluster 1E0657-558. By comparing the weak lensing reconstruction of the gravitational po-



**Galaxy Cluster Cl 0024+17 (ZwCl 0024+1652)**  
*Hubble Space Telescope • ACS/WFC*

NASA, ESA, and M.J. Jee (Johns Hopkins University)

STScI-PRC07-17b

Figure 1-2: Strong gravitational lensing around galaxy cluster CL0024+17, consisting of the gravitationally bound yellow, elliptical galaxies. The elongated blue objects are from much more distant galaxies behind the cluster which are distorted into arcs due to gravitational lensing from the dark matter halo surrounding the cluster. Figure credit: NASA, ESA, M.J. Jee and H. Ford (Johns Hopkins University)

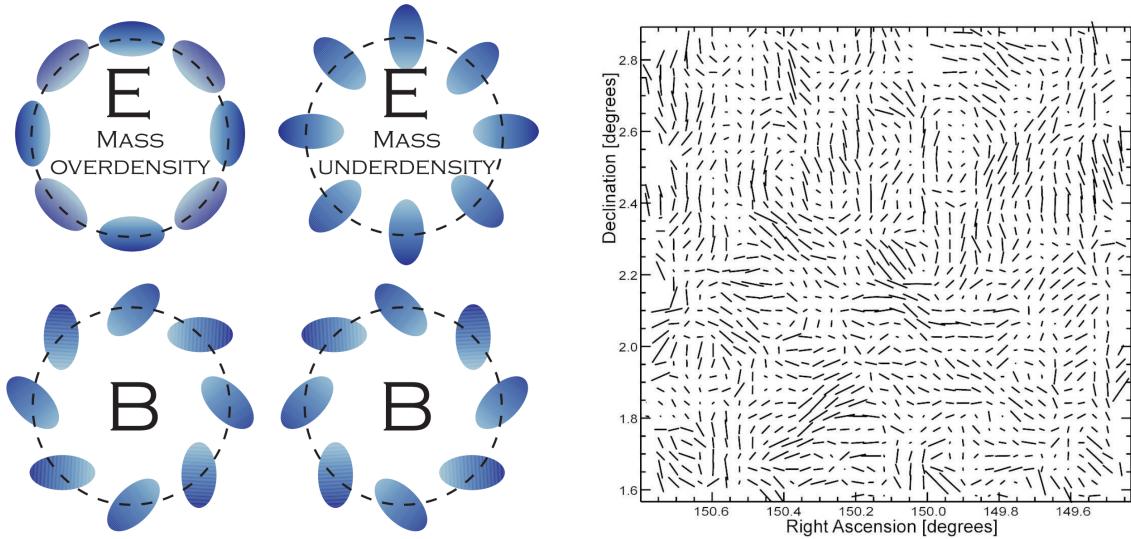


Figure 1-3: Left: Examples of circular “E-mode” and curl-like “B-mode” patterns. Right: The observed ellipticities of half a million distant galaxies within the 2 square degree Hubble Space Telescope COSMOS survey. Reprinted fom Reference [Massey2010].

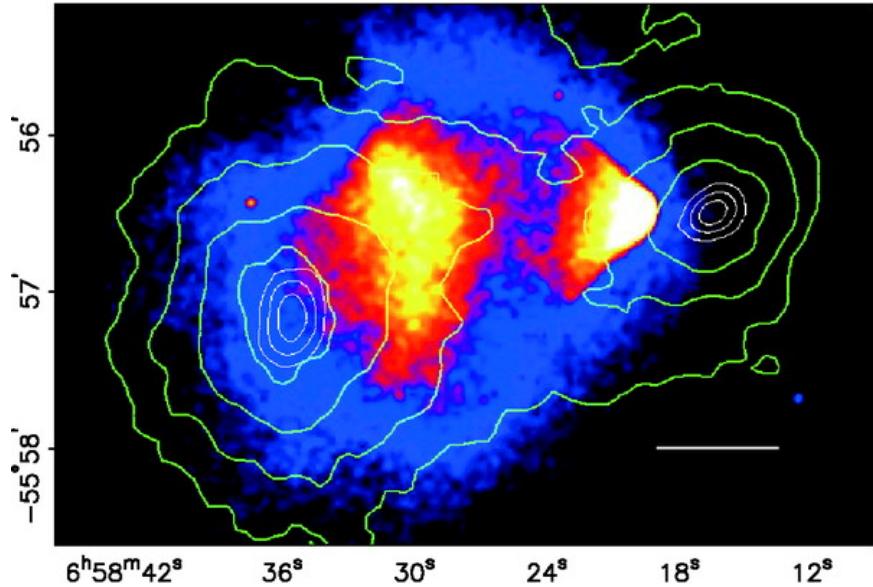


Figure 1-4: The merging cluster 1E0657-558. The green contours show the weak lensing reconstruction of the gravitational potential of the cluster. The colors indicate the X-ray temperature of the plasma, changing from blue to white as the plasma goes from coolest to hottest. The smaller “bullet” cluster on the right which traversed through the larger cluster on the left. Reprinted from Reference [Clowe2006].

tential of the cluster shown in green contours against temperature color gradient of the X-ray emitting interstellar plasma, it was learned that the gravitational potential of the cluster does not track the dominant baryonic mass contribution coming from the plasma. Instead, the gravitational potential tracks the smaller stellar baryonic mass component. Dark matter must be the dominant gravitational source in the cluster since the center of total mass is offset from the center of baryonic mass. Furthermore, the observation of two gravitational mass centers places strong constraints on the self-interaction of dark matter requiring that the observed mass must have a self-interaction collisional cross section  $\sigma/m < 1.25 \text{ cm}^2\text{g}^{-1}$  at a 68% confidence level [**Clowe2006**].

## 1.2 Relic Density

During the early universe, dark matter (DM) was in thermal equilibrium with the rest of the SM particles with a number density  $n_\chi$  given by

$$n_\chi^{\text{eq}} = \frac{g}{(2\pi)^3} \int f(\vec{p}) d^3\vec{p}, \quad (1.6)$$

where  $g$  is the number of internal degrees of freedom of the DM particle  $\chi$  and  $f(\vec{p})$  is either the Fermi-Dirac or Bose-Einstein distribution, depending on the quantum statistics of the DM particle [**Jungman1996**]. At very high temperatures relative to the mass  $m_\chi$  of the DM particle, dark matter and SM particles rapidly convert back and forth with a DM annihilation rate  $\Gamma = \langle \sigma_A v \rangle \cdot n_\chi$ , where  $\langle \sigma_A v \rangle$  is the thermally averaged product of the total cross section for annihilation  $\sigma_A$  and the relative velocity  $v$  of the dark matter particles. After the temperature drops below  $m_\chi$ , the DM annihilation rate  $\Gamma$  drops below the Hubble expansion rate  $H$  of the universe and the DM particles stop annihilating and freeze-out of equilibrium with the SM particles, leaving the DM relic density that we observe today. During the freeze-out process, the time dependence of the number density  $n_\chi$  is described by the

Boltzmann equation [**Jungman1996**] as follows

$$\frac{dn_\chi}{dt} + 3Hn_\chi = -\langle\sigma_A v\rangle [(n_\chi)^2 - (n_\chi^{\text{eq}})^2]. \quad (1.7)$$

The term on the left-hand side accounts for the reduction in density due to the expansion of the universe. The two terms on the right-hand side account for the change in density due to annihilation and product of DM particles to and from SM particles, respectively.

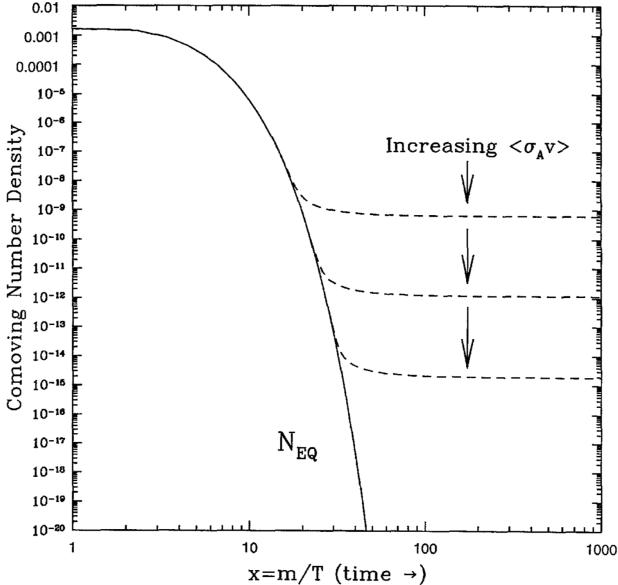


Figure 1-5: Number density of dark matter in the early universe as a function of time. The solid curves are the equilibrium abundance while the dashed curves are the actual abundance after freeze-out. Reprinted from Reference [**Jungman1996**].

Figure 1-5 shows the calculated DM number density  $n_\chi$  as a function of time in the early universe. As the annihilation cross section increases, the relic density decreases as the dark matter particles stay in equilibrium longer. Assuming that  $\langle\sigma_A v\rangle$  is independent of energy, an order-of-magnitude estimate of the relic density is given by

$$\Omega_\chi \cdot h^2 = \frac{m_\chi n_\chi}{\rho_c} \simeq \frac{10^{-27} \text{ cm}^3 \text{ s}^{-1}}{\langle\sigma_A v\rangle}, \quad (1.8)$$

where  $h = H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1})$  is the reduced Hubble constant and  $\rho_c = 3H^2/(8\pi G)$

is the critical density of the universe [**Jungman1996**]. Thus, to first order the DM relic density depends only on two things: the total DM annihilation cross section  $\sigma_A$  and the mass  $m_\chi$  of the DM candidate, as after freeze-out  $T \ll m_\chi$  and the velocity  $v$  is strictly proportional to  $m_\chi$ . The latest Planck results measure that  $\Omega_\chi \cdot h^2 = 0.1200 \pm 0.0012$  [**Planck2018**], providing strong constraints on the possible values of  $\sigma_A$  and  $m_\chi$ .

### 1.3 Dark Matter Candidates

Dark matter has not been directly observed so far. In the previous sections, we collected all known facts about dark matter to restrict the pool of potential candidates. Any serious dark matter candidate must satisfy the following criteria:

- No or extremely weak interactions with photons, e.g. be ***dark***
- Weak baryonic interactions to preserve the DM halos
- Weak self-interactions
- The observed relic density

These four criteria place stringent requirements on dark matter candidates. The light neutrinos, the only SM particles satisfying the first three conditions, are excluded as the total neutrino relic density has a bound of  $\Omega_\nu \cdot h^2 \leq 0.00067$  at 95% confidence level from analysis of the CMB anisotropies. Big Bang Nucleosynthesis and gravitational microlensing have mostly exclude non-luminous baryonic matter from forming the bulk of dark matter. Thus, most theories of dark matter propose a new fundamental particle as a dark matter candidate. The following sections discuss the most common dark matter candidates, namely weakly-interacting massive particles (WIMPs), axions, and sterile neutrinos.

### 1.3.1 Weakly-Interacting Massive Particles

The annihilation cross-section of new particle  $\chi$  with electroweak scale interactions is approximately

$$\langle \sigma_A v \rangle \approx \left( \frac{\alpha \cdot g_\chi^2}{m_\chi} \right)^2 = \left( \frac{e^2}{4\pi} \cdot \frac{(0.8)^2}{100 \text{ GeV}} \right)^2 \simeq 10^{-26} \text{ cm}^3 \text{ s}^{-1}, \quad (1.9)$$

where  $\alpha = e^2/4\pi$  is the fine-structure constant,  $m_\chi = 100 \text{ GeV}$  is the mass of the particle, and  $g_\chi \approx 0.8$  is the effective coupling for the four-point interaction  $\chi\bar{\chi} \rightarrow f\bar{f}$ . Plugging this into Equation 1.8, we obtain  $\Omega_\chi \cdot h^2 \sim 0.1$ , which is very close to the observed value. This numerical coincidence, known as the “WIMP miracle”, motivates a new weakly-interacting massive particle (WIMP) as a good DM candidate [Jungman1996].

The only criteria to be a WIMP beyond the generic definition of dark matter is that the mass and interaction strength of the new particle must be approximately that of the EWK scale. Thus, a great many new physics models have WIMPs natively, such as the neutralino in supersymmetry and the lightest Kaluza-Klein particle in theories of universal extra dimensions. In this thesis, we shall focus on a set of simplified models that describe WIMP interactions in a relatively model-independent manner so that our results can be reinterpreted in as many theories as possible.

### 1.3.2 Axions

The hypothetical axion particle introduced in Section ?? is a dark matter candidate if the axion decay constant  $f_a$  is large enough as all axion-SM couplings are inversely proportional to it. Constraints from the observed duration of the neutrino burst from supernova SN 1987A require that  $f_a \gtrsim 10^9 \text{ GeV}$  [Raffelt2008], sufficiently high that the axion lifetime exceeds the age of the universe by many orders of magnitude. Thus, the axion is a viable DM candidate due to its long lifetime and weak couplings to SM particles.

After accounting for kinematic mixing with the  $\pi^0$  and  $\eta$  mesons, the axion mass

is given by

$$m_a = \left( \frac{\sqrt{m_u m_d}}{m_u + m_d} \right) \left( \frac{f_\pi}{f_a} \right) m_\pi \simeq \left( \frac{10^7 \text{ GeV}}{f_a} \right) \text{ eV}, \quad (1.10)$$

where  $f_\pi$  is the pion decay constant and  $m_u$ ,  $m_d$ , and  $m_\pi$  are the masses of the up quark, the down quark, and the neutral pion, respectively. From this, we see that the axion mass is inversely proportional to  $f_a$ , leading to an upper limit of  $m_a \lesssim 10 \text{ meV}$ . The Planck measurements of the cosmic microwave background also provide a lower (upper) bound on the axion mass  $m_a \gtrsim 10 \mu\text{eV}$  (axion decay constant  $f_a \lesssim 10^{12} \text{ GeV}$ ), otherwise the axion abundance is too high.

The axion obtains a two-photon vertex through loops involving virtual quarks and gluons of the form

$$\mathcal{L}_{a\gamma\gamma} = \frac{1}{4} g_{a\gamma\gamma} \epsilon_{\mu\nu\rho\sigma} F^{\mu\nu} F^{\rho\sigma} a = -g_{a\gamma\gamma} (\vec{E} \cdot \vec{B}) a, \quad (1.11)$$

where  $F^{\mu\nu}$  is the electromagnetic field-strength tensor,  $\vec{E}$  and  $\vec{B}$  are the electric and magnetic fields, respectively, and the coupling constant is

$$g_{a\gamma\gamma} = -\frac{\alpha}{3\pi f_a} \left( \frac{m_u + 4m_d}{m_u + m_d} \right). \quad (1.12)$$

From this, we can see that the axion's coupling to the photon is incredibly small unless in the presence of a strong electromagnetic field. Thus, astrophysical axions are dark unless they enter a region with such a field. Searches for axions such as CAST[CAST2017] and ADMX[ADMX2018a, ADMX2018b] exploit this to try to observe axion-to-photon conversions.

### 1.3.3 Sterile Neutrinos

In Section ??, we introduced mass for the SM fermions through the Higgs mechanism in order to preserve  $SU(2)_L$  gauge invariance. However, since the right-handed neutrinos are singlets under the full  $SU(3)_C \times SU(2)_L \times U(1)_Y$  gauge of the SM, it is possible to add explicit Majorana mass terms in addition to the Dirac mass terms in

the SM Lagrangian [**Drewes2013**] as follows

$$\begin{aligned}
\mathcal{L}_{\nu_R} &= i\bar{\nu}_R \not{\partial} \nu_R - (\bar{\ell}_L Y_\nu \phi_c \nu_R + \bar{\nu}_R Y_\nu^\dagger \phi_c^\dagger \ell_L) - \frac{1}{2} \left( \bar{\nu}_R^c M_M \nu_R + \bar{\nu}_R M_M^\dagger \nu_R^c \right) \\
&= i\bar{\nu}_R \not{\partial} \nu_R - (\bar{\nu}_L M_\nu \nu_R + \bar{\nu}_R M_\nu^\dagger \nu_L) - \frac{1}{2} \left( \bar{\nu}_R^c M_M \nu_R + \bar{\nu}_R M_M^\dagger \nu_R^c \right) \\
&= i\bar{\nu}_R \not{\partial} \nu_R - \frac{1}{2} \left[ \begin{pmatrix} \bar{\nu}_L & \bar{\nu}_R^c \end{pmatrix} \begin{pmatrix} 0 & M_\nu \\ M_\nu^T & M_M \end{pmatrix} \begin{pmatrix} \nu_L^c \\ \nu_R \end{pmatrix} + h.c. \right], \tag{1.13}
\end{aligned}$$

where  $\nu_R^c$  are the charge conjugates of the right-handed neutrinos,  $h.c.$  stands for hermitian conjugate, and  $Y_\nu$ ,  $M_\nu$ , and  $M_M$  are the Yukawa, Dirac mass, and Majorana mass matrices for the neutrinos, respectively.

In the limit  $M_M \gg M_\nu$ , the combined mass matrix has two distinct sets of eigenvalues: the sterile neutrinos  $\nu_s$  with masses  $m_s \simeq M_M$  and the active neutrinos  $\nu_a$  with masses  $m_a \simeq M_\nu^2/M_M$ . The active and sterile neutrinos are related to the left- and right-handed neutrinos by the active-sterile mixing matrix  $\theta = M_\nu M_M^{-1}$  as follows

$$\begin{aligned}
|\nu_a\rangle &= \cos \theta |\nu_L\rangle + \sin \theta |\nu_R\rangle \\
|\nu_s\rangle &= -\sin \theta |\nu_L\rangle + \cos \theta |\nu_R\rangle. \tag{1.14}
\end{aligned}$$

In this formulation, the active neutrinos are the observed neutrinos of the SM while the sterile neutrino states are a promising dark matter candidate as they only interact with the SM through neutrino oscillations. Furthermore, the sterile neutrinos must have  $m_s \simeq$  keV or else they cannot account for the observed masses of DM-dominated objects without violating the Pauli exclusion principle [**Boyarsky2009**]. Fortunately, this constraint also means that the oscillation rate into active neutrinos is low enough that the sterile neutrino lifetime is longer than the age of the universe. The Majorana mass term also allows for additional CP-violating phases beyond the Dirac phases. Astrophysical detection, accelerator production, and neutrinoless double  $\beta$  decay experiments all place constraints on sterile neutrino properties [**Boyarsky2019**].

## 1.4 Simplified Models for LHC

Without loss of generality, we assume that dark matter consists of a single Dirac fermion WIMP species  $\chi$  with mass  $m_{\text{DM}}$  [18]. We are able to do this because collider searches are only minimally sensitive to the details of the dark matter candidate. Instead, collider searches are mainly sensitive to a new mediator particle with mass  $M_{\text{med}}$  that connects the dark sector to the Standard Model through couplings  $g_q$  and  $g_{\text{DM}}$  to the SM quarks and  $\chi$ , respectively. In general, this mediator can have any spin structure; for purposes of simplicity, we limit ourselves to the observed mediators in the SM: a spin-0 scalar  $S$ , a spin-0 pseudoscalar  $P$ , a spin-1 vector  $V_\mu$ , and a spin-1 axial-vector  $A_\mu$  (not to be confused with SM photon) [**Harris2015**]. The Lagrangians for these possible mediators are

$$\mathcal{L}_S = \frac{1}{2} M_{\text{med}}^2 S^2 + g_{\text{DM}} \bar{\chi} S \chi + g_q \sum_q \bar{q} Y_q S q \quad (1.15)$$

$$\mathcal{L}_P = \frac{1}{2} M_{\text{med}}^2 P^2 + g_{\text{DM}} \bar{\chi} \gamma^5 P \chi + i g_q \sum_q \bar{q} Y_q \gamma^5 P q \quad (1.16)$$

$$\mathcal{L}_V = \frac{1}{2} M_{\text{med}}^2 V_\mu V^\mu + g_{\text{DM}} \bar{\chi} \not{V} \chi + g_q \sum_q \bar{q} \not{V} q \quad (1.17)$$

$$\mathcal{L}_A = \frac{1}{2} M_{\text{med}}^2 A_\mu A^\mu + g_{\text{DM}} \bar{\chi} \gamma^5 \not{A} \chi + g_q \sum_q \bar{q} \gamma^5 \not{A} q, \quad (1.18)$$

where  $\{q\}$  and  $\{Y_q\}$  are the SM quarks and their associated Yukawa couplings and we have assumed that the coupling  $g_q$  is universal to all quarks without any loss of generality. The width of the mediators is determined by the coupling constants and mass of the mediator. Thus, these four models are described by only four parameters in addition to the spin structure of their couplings:  $\{g_q, g_{\text{DM}}, m_{\text{DM}}, M_{\text{med}}\}$ . For the results shown in this thesis, we fix  $g_q = 0.25$  and  $g_{\text{DM}} = 1$  to satisfy the narrow width approximation, provide complementarity to dijet and dilepton searches for new mediators, and enable easy comparison between other dark matter searches from ATLAS and CMS. Meanwhile, we scan the mass parameters  $m_{\text{DM}}$  and  $M_{\text{med}}$  between 1-1000 GeV and 10-2000 GeV, respectively [18].

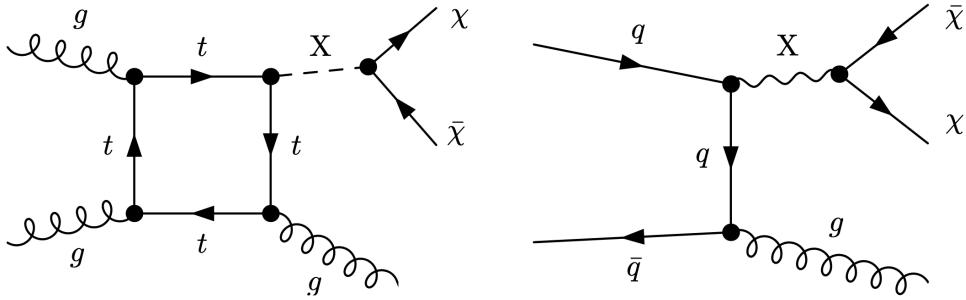


Figure 1-6: Feynmann diagrams for possible dark matter production modes at the LHC, with  $X$  denoting the mediator and  $\chi$  the dark matter particle. The left diagram is most common for scalar and pseudoscalar mediators, while the right diagram is most common for vector and axial mediators. For the right diagram, the gluon can be replaced with any of the EWK gauge boson whereas for the left diagram, other production modes dominate for the EWK gauge bosons. Reprinted from Reference [Harris2015].

As the CMS detector is focused on observing particles through electromagnetic and strong interactions, the production of DM particles alone does not result in an observable signature in the detector. Instead, we look for events where they are produced in association with a visible SM particle with the resulting signature of the SM particle recoiling against an invisible state. The standard mechanism for producing this final state is for one of the incoming partons to radiate the visible SM particle, usually a quark or a gauge boson, as shown in Figure 1-6. The possible final states split into the categories shown in Table 1.1 with unique final states and interaction strengths.

Name	Strength	Final State
Monojet	$\mathcal{O}(\alpha_s)$	$j + \chi\bar{\chi}$
Monophoton	$\mathcal{O}(\alpha)$	$\gamma + \chi\bar{\chi}$
Mono- $Z$	$\mathcal{O}(\alpha)/M_Z^2$	$\ell\ell + \chi\bar{\chi}$

Table 1.1: Final states of DM production with approximate interaction strength.

The monojet final state has the highest production rate but also the largest amount of background from SM processes. The Mono- $Z$  final state is the cleanest signature but also ahhs a substantially suppressed production rate due to the mass of the  $Z$  boson and the small branching ratio to leptons. The monophoton final state lands in the

middle on both metrics: the production rate is  $\sim 1/10$  that of the monojet final state but also has fewer backgrounds from SM processes. Additionally, the monophoton state has the “advantage” of being able to distinguish between the spin-0 and spin-1 mediators by the “virtue” of the  $gg \rightarrow S\gamma$  and  $gg \rightarrow P\gamma$  processes being forbidden by Furry’s theorem [Gabrielli2016].

## 1.5 Non-collider Searches

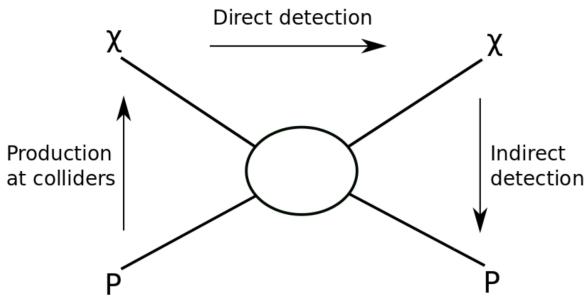


Figure 1-7: Possible dark matter detection channels. Reprinted from Reference [Undagoitia2015].

The goal of a collider-based search for dark matter is to create pairs of DM candidates through quark-antiquark annihilation, i.e.  $q\bar{q} \rightarrow \chi\bar{\chi}$ . However, there are other ways you can look at this interaction when attempting to find dark matter, as shown in Figure 1-7. The longest running searches for DM particles have been direct detection searches, where one attempts to observe the scattering of a DM particle off of a heavy nucleus, i.e.  $q\chi \rightarrow q\chi$ . Another method involves the inverse process utilized at hadron colliders, the observation of SM particles produced in the annihilation of astrophysical DM pairs, i.e.  $\chi\bar{\chi} \rightarrow q\bar{q}$ . In this section, we shall give a brief overview of the leading experiments for these two approaches.

### 1.5.1 Direct Detection

Direct detection experiments aim to observe a collision between a DM particle  $\chi$  and an atomic nucleus  $N$ . Current experiments target nuclear recoils in the  $1 -$

100 keV range and there are various different technologies that are used to identify the recoiling nucleus, as shown in Figure 1-8. The three main signals are light from a scintillating material, charge from an ionization reaction, and phonons from thermal excitations [Undagoitia2015]. Many detectors utilize multiple signals in order to enhance background rejection.

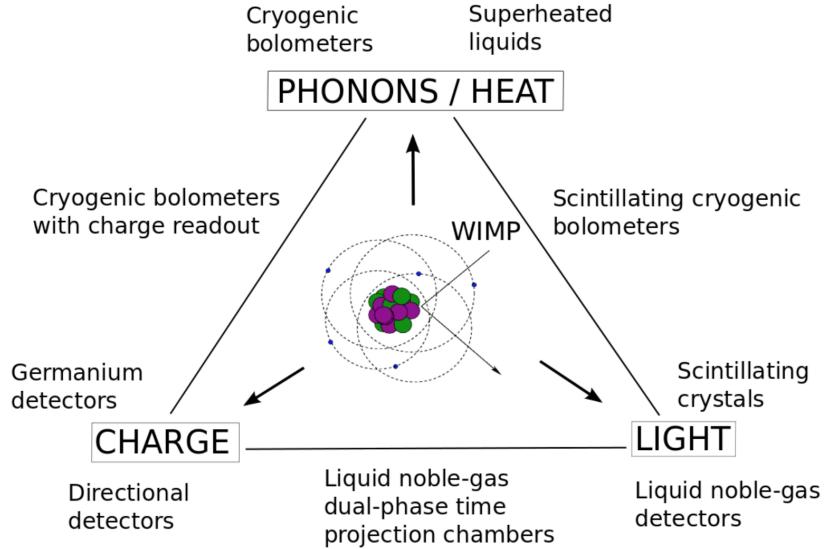


Figure 1-8: Possible signals measured and technologies used in direct detection experiments. Reprinted from Reference [Undagoitia2015].

Another consideration is whether the interaction is spin-independent (SI) or spin-dependent (SD), corresponding to a vector/scalar or axial/pseudoscalar mediator, respectively. The SI DM-nucleus cross-section scales as the number of nucleons squared and thus experiments focusing on SI interactions use materials with a large atomic number. Conversely, the spin-dependent DM-nucleus cross-section depends on the total nuclear spin and experiments focusing on SD interactions use materials with an odd number of nucleons, particularly those with unpaired protons and neutrons. The remainder of this section will discuss three proto-typical direct detection experiments.

The Large Underground Xenon (LUX) experiment [20] is a dual-phase xenon time projection chamber that utilizes both scintillation light and free electrons from ionization to detect nuclear recoils. The active detector volume has 250 kg of ultrapure

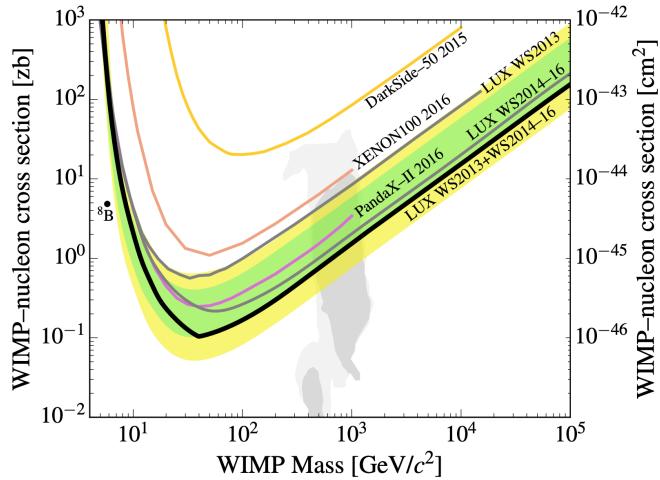


Figure 1-9: Latest SI independent results from the LUX experiment (brazilian flag bands). Reprinted from Reference [20].

liquid xenon, a large enough volume that the detector is self-shielding. The presence of naturally occurring odd isotopes of xenon with unpaired nucleons enable the measurements of spin-dependent interactions are possible with an effective volume of approximately half the active detector volume. The latest LUX results are shown in Figure 1-9.

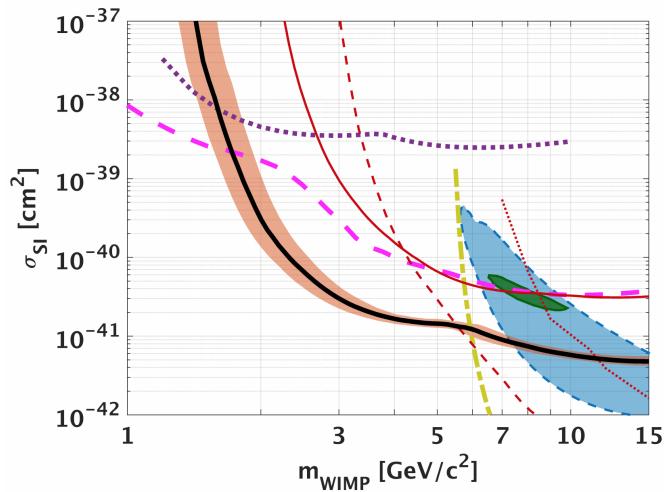


Figure 1-10: Latest SI independent results from the CDMSlite experiment (black line with salmon-colored band). Reprinted from Reference [19].

The CDMSlite experiment [19] uses a cryogenic germanium bolometer that detects

phonons and ionization from nuclear recoils. The presence of both signals allows for particle identification, but the most sensitive results only use the phonon signals. CDMSlite is most sensitive to dark matter with masses between 1-10 GeV as shown in Figure 1-10.

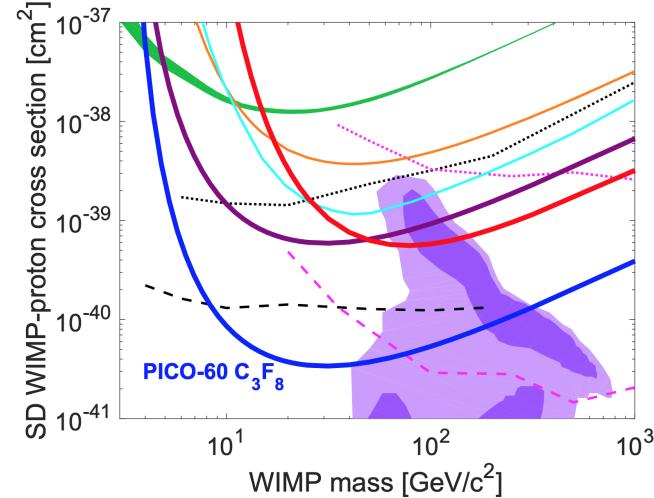


Figure 1-11: Latest SD independent results from the PICO-60 experiment (thick blue line). Reprinted from Reference [24].

The PICO-60 experiment [24] uses a bubble chamber filled with  $\text{C}_3\text{F}_8$  to observe nuclear recoils with a threshold of 13.6 keV. This threshold is high enough to reject backgrounds from minimum ionizing particles while still having sufficient signal efficiency. Flourine has exactly one unpaired proton and no unpaired neutrons which greatly enhances the sensitivity to the proton-coupling SD interactions. The latest results from PICO-60 are shown in Figure 1-11.

### 1.5.2 Indirect Detection

We discussed the process of dark matter annihilation in the early universe and found that the relic density  $\Omega_\chi \cdot h^2$  depended on the average annihilation cross section  $\langle \sigma_A v \rangle$ . We found that a 100 GeV WIMP particle must have  $\langle \sigma_A v \rangle \sim 10^{-26} \text{ cm}^3 \text{ s}^{-1}$  in order to produce the observed relic density. While our calculation assumed that the DM number density did not change appreciably after freezeout, there is still a non-zero annihilation rate of  $n_\chi^2 \langle \sigma_A v \rangle / 2$  for dark matter to SM particles. With this decay rate

and the observed relic density, there should be  $\mathcal{O}(10)$  particles from DM annihilation reaching Earth per year per square meter [**Slatyer2017**]. Thus, it is possible to learn about dark matter by observing the particle flux near Earth.

While dark matter can decay to a pair of any SM particles, the only ones that live long enough to reach Earth are photons, neutrinos, electrons, positrons, protons, and anti-protons, whether produced directly or from decays of short-lived particles. While it is possible to measure the flux of all of these particles, it is only possible to observe the energy spectrum and source locations of photons and neutrinos because galactic magnetic fields obscure these for the charge particles. Thus, experiments such as the Alpha Magnetic Spectrometer (AMS-02) [**AMS2014**] look for a rise in the positron and anti-proton fractions of the cosmic ray flux to find signs of WIMP annihilation. Similarly, gamma ray telescopes such as the Fermi Large Area Telescope (Fermi-LAT) [**Fermi2015**] search for high-energy photons originating from dwarf spheroidal galaxies as they have especially high ratios of dark matter to luminous matter. Both experiments have tentative signals that are yet to be confirmed.

The neutrino channel differs from the other two because it is the only channel sensitive to the annihilation of WIMPs that have been gravitationally captured by the sun because only neutrinos escape the sun with their original energy spectrum intact [**Gaskins2016**]. Since solar WIMPs are in equilibrium, the annihilation flux depends only on the DM-nucleon cross-section instead of the annihilation cross-section, and since the sun consists primarily of hydrogen, observations of neutrinos from solar WIMPs are especially sensitive to spin-dependent processes. Thus, because astrophysical uncertainties are very small for neutrinos originating from the sun versus galactic sources, results from solar neutrino observatories are competitive with those from direct detection experiments.

The IceCube neutrino observatory [25] is a Cherenkov light detector consisting of a cubic kilometer of Antarctic ice instrumented with 79 strings of photomultiplier tubes. Figure 1-12 shows the latest results from IceCube, with each curves assuming annihilation into a different SM particle pair.

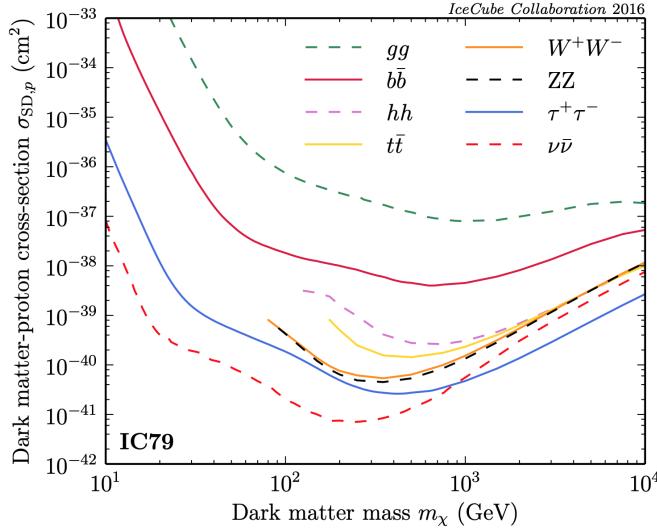


Figure 1-12: Latest SD independent results from the IceCube experiment. Reprinted from Reference [25].

## 1.6 Summary

Copious astrophysical evidence exists for a non-relativistic non-baryonic type of matter referred to as dark matter that constitutes  $\sim 27\%$  of the energy of the universe. An attractive candidate for dark matter is a new beyond the Standard Model particle that has weak interactions with photons, baryons, and itself and a relic density of  $\Omega_\chi \cdot h^2 = 0.1200 \pm 0.0012$ . While other candidates exist, a new weakly-interacting massive particle is a dark matter candidate found in many theories of new physics. Collider, direct detection, and indirect detection searches can all be interpreted in the context of simplified models, which feature a new Dirac fermion DM particle and a massive mediator. The remainder of this thesis focuses on a search for dark matter produced in conjunction with a single photon at the Large Hadron Collider using the Compact Muon Solenoid detector.



# Chapter 2

## The Large Hadron Collider

### 2.1 Experimental Apparatus

The Large Hadron Collider (LHC) is a circular proton-proton collider, 27 km in circumference and between 40 and 175 m below the surface, located at the European Organization for Nuclear Research (CERN) on the French-Swiss border near the city of Geneva [**Evans2008**]. Designed to collide protons at a maximum center-of-mass energy  $\sqrt{s} = 14 \text{ TeV}$ , the LHC has delivered collisions at  $\sqrt{s} = 7, 8 \text{ TeV}$  during Run 1 (2010-2012) and at  $\sqrt{s} = 13 \text{ TeV}$  during Run 2 (2015-2018). While the LHC is primarily a proton-proton collider, lead (Pb) ion beams of energy of up to 2.8 TeV per nucleon are used to produce lead-lead and proton-lead collisions. In this thesis, we focus exclusively on data recorded from proton-proton collisions during Run 2.

The LHC is the final stage of the CERN accelerator complex [**Benedikt2004**] depicted in Figure 2-1. Hydrogen atoms are stripped of their electrons and accelerated to an energy of 50 MeV by the LINAC2 linear acceleration. Following this, they are injected into the Booster ring, the Proton Synchrotron (PS), and the Super Proton Synchrotron (SPS) and accelerated to 1.4, 26, and 450 GeV, respectively. After the SPS, the protons are injected into the two counter-circulating rings of the LHC in up to 2808 discrete bunches with a bunch spacing of 25 ns. The two beams intersect in eight places along the LHC with detector experiments CMS, ATLAS, LHCb, and ALICE each located at an intersection point.

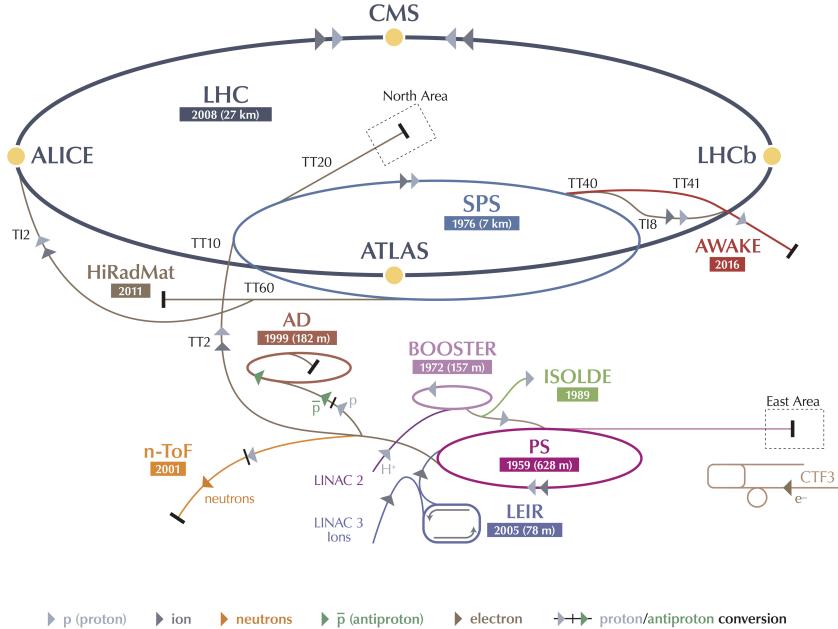


Figure 2-1: A schematic representation of the CERN accelerator complex. The LHC (dark blue) is fed protons by a chain of intermediate accelerators, beginning with LINAC2 (light pink). Reprinted from the CERN Document Server [Mobs2018].

The LHC is a synchrotron containing 1232 superconducting NbTi dipole magnets measuring 15 m in length, each with a peak dipole field of 8.33 T. There are an additional 492 quadrupole magnets measuring 5-7 m in length which focus the beams in between the dipole magnets. Due to space limitations in the tunnels, the beam pipes are magnetically coupled and the magnets share the same superfluid liquid helium cryostatic system needed to reach the 1.9 K temperature required to achieve the desired magnetic field strength and cool sufficient amount of power out of the system.

The number of events produced at the LHC is given by

$$N(pp \rightarrow X) = \int dt L(t) \sigma(pp \rightarrow X), \quad (2.1)$$

where  $\sigma$  is the cross section of the process and  $L$  is the instantaneous luminosity of

the machine given by

$$L = \frac{N_b^2 n_b f_{\text{rev}} \gamma}{4\pi \epsilon \beta^*} \times F, \quad (2.2)$$

where  $N_b$  is the number of particles per bunch ( $\mathcal{O}(10^{11})$ ),  $n_b$  is the number of bunches per beam,  $f_{\text{rev}}$  is frequency of revolution,  $\gamma$  is the Lorentz factor of the beam,  $\epsilon$  is transverse emittance of the beam,  $\beta^*$  is beta function of the beam at the collision point, and  $F$  is the geometric luminosity reduction factor due to the crossing angle at the interaction point. The instantaneous luminosity decreases exponentially as a function of time due to  $N_b$  and  $n_b$  being reduced by collisions. The LHC is designed to deliver an initial instantaneous luminosity of  $\mathcal{O}(10^{34}) \text{ cm}^{-2} \text{ s}^{-1}$ , achieved by having multiple inelastic proton-proton interactions per bunch crossing known as pileup.

As all known cross sections are time-independent, the total number of events is directly proportional to the integrated luminosity given by

$$L_{\text{int}} = \int_0^T dt L(t) = L(0) \tau_L (1 - e^{-T/\tau_L}), \quad (2.3)$$

where  $T$  is the time since starting collisions,  $L(0)$  is the initial instantaneous luminosity, and  $\tau_L \approx 15 \text{ h}$  the characteristic beam loss timescale for the LHC. The total luminosity delivered by the LHC and recorded by CMS during the 2016 is shown in Figure 2-2.

## 2.2 Collider Phenomenology

The proton is a composite particle consisting of valence quarks, sea quarks, and gluons, collectively referred to as partons. When colliding protons at the LHC, we are actually interested in the inelastic scattering of a pair of partons from the incident protons. Each parton  $a, b$  carries a fraction of the momentum of the incoming proton  $x_{a,b}$  following the particle-dependent parton distribution functions (PDFs)  $f_{a,b}$ . The differential cross section for  $2 \rightarrow N$  parton scattering process [Perelstein2010] is

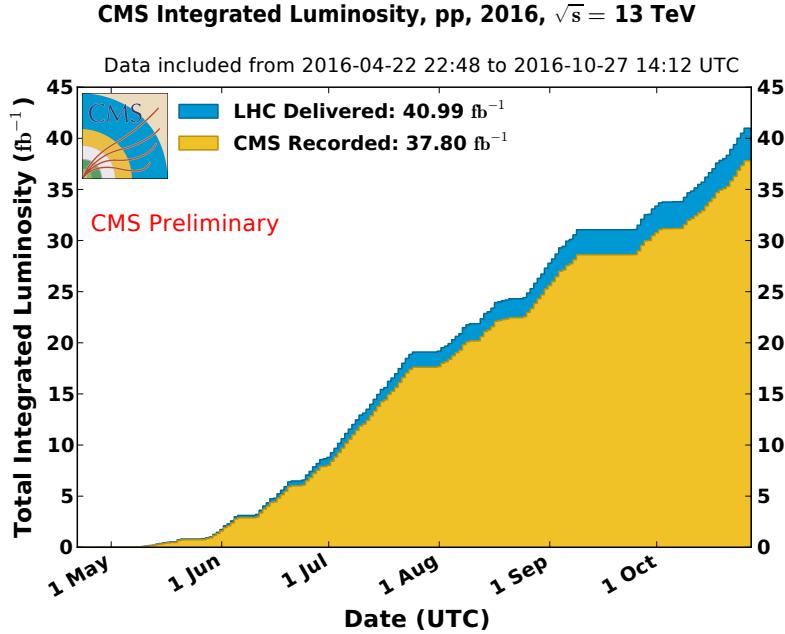


Figure 2-2: The total integrated luminosity of the LHC during proton-proton collisions during 2016 [LumiTwiki]. While a total luminosity of  $41 \text{ fb}^{-1}$  was collected, only a subset during which the detector operated well enough is used in this thesis. This corresponds to  $36 \text{ fb}^{-1}$  of data.

given by

$$d\sigma(ab \rightarrow \{c_i\}) = \frac{(2\pi)^4}{2s} \left( \prod_i \frac{d^3 p_i}{(2\pi)^3} \right) \cdot \delta^4 \left( k_a + k_b - \sum_i p_i \right) \cdot |\mathcal{M}(ab \rightarrow \{c_i\})|^2 \quad (2.4)$$

where  $k_{a,b} = x_{a,b}\sqrt{s}$  are the momenta of the incoming partons,  $\{p_i\}$  are the momenta of the outgoing partons  $\{c_i\}$ , and  $\mathcal{M}$  is the matrix element of the process.

This parton level scattering, called the hard scattering process, is perturbatively calculable through standard QFT methods. However, the hard scattering does not include any effects related to the PDFs of the incoming partons or the decay and hadronization of the outgoing partons into the final state particles (called the parton shower), both of which involve non-perturbative aspects of QCD. Fortunately, the collinear factorization theorem [Collins1989] states that the probability of obtaining the final state  $X(\Theta)$  from a hadron collision can be calculated as the product of the probability that specific partons  $a, b$  are involved in the interaction, the probability for the hard scattering to produce outgoing partons  $\{c_i\}$ , and the formation of final

state hadrons from these outgoing partons. The factorization process is not unique and requires the choice of an arbitrary energy scale  $\mu_F$ , which defines a lower bound for interactions to be considered part of the hard scattering.

Including the effects from PDFs and parton showering (PS), the general cross section for  $pp \rightarrow X(\Theta)$  is

$$\begin{aligned} \frac{d\sigma}{d\Theta}(pp \rightarrow X(\Theta)) &= \sum_{a,b} \int dx_a f_a(x_a, \mu_F) \cdot dx_b f_b(x_b, \mu_F) \\ &\quad \times d\sigma(ab \rightarrow \{c_i\}) \times D(\{c_i\} \rightarrow X(\Theta)), \end{aligned} \quad (2.5)$$

where the sum is over the initial state partons and  $D$  is the fragmentation function that describes parton shower process resulting in the observed final state.

The following sections discuss the simulation of the three main elements of Equation 2.5: the parton distribution functions  $f_a$ , the hard scattering cross section  $d\sigma$ , and the parton shower and hadronization processes that contribute to the fragmentation function  $D$ .

### 2.2.1 Parton Distribution Functions

Due to soft collinear emissions from the partons, the behavior of the parton distribution functions depends on the factorization scale. Denoting the gluon PDF as  $g(x, \mu_F)$  and the PDF for quark flavor  $i$  as  $q_i(x, \mu_F)$ , the analytic behavior of the PDFs is given by the DGLAP [Dokshitzer1977, Gribov1972, Altarelli1977] evolution equations

$$\mu_F \frac{d}{d\mu_F} \begin{pmatrix} q_i(x, \mu_F) \\ g(x, \mu_F) \end{pmatrix} = \frac{\alpha_s}{2\pi} \int_x^1 \frac{dy}{y} \begin{pmatrix} P_{qq}(x/y) & P_{qg}(x/y) \\ P_{gq}(x/y) & P_{gg}(x/y) \end{pmatrix} \begin{pmatrix} q_i(y, \mu_F) \\ g(y, \mu_F) \end{pmatrix} \quad (2.6)$$

where  $y$  is the fraction of momentum carried by initial parton and the  $P$  matrix elements are the splitting kernels defined by

$$\left. \begin{array}{l} P_{qq}(z) = \frac{4}{3} \left( \frac{1+z^2}{1-z} \right) \\ P_{gq}(z) = \frac{1}{2} (z^2 + (1+z)^2) \end{array} \right| \begin{array}{l} P_{qg}(z) = \frac{4}{3} \left( \frac{1+(1-z)^2}{z} \right) \\ P_{gg}(z) = 6 \left( \frac{1-z}{z} + \frac{z}{1-z} + z(1-z) \right). \end{array} \quad (2.7)$$

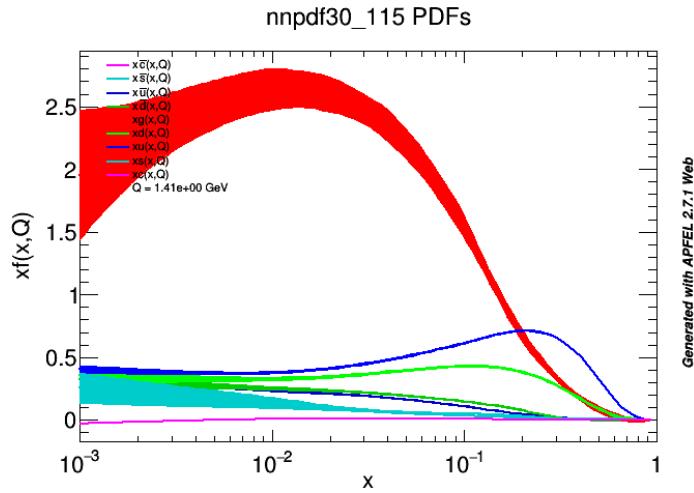


Figure 2-3: The various quark and gluon PDFs for the proton, as a function of momentum fraction  $x$ . The specific PDF set is the NNPDF3.0 118 NLO PDF set.

The DGLAP equations cannot be solved analytically at a fixed scale. Instead, parameterized functional forms are fitted to data from many experiments. The results presented in this thesis use the NNPDF3.0 PDF set provided by the NNPDF collaboration [Ball2015]. Figure 2-3 shows the quark and gluon PDFs for the proton. As  $x \rightarrow 0$ , the gluon fraction dominates while near  $x \approx 0.3$ , the up-quark fraction  $u(x, \mu_F)$  approaches  $\frac{2}{3}$ , the down-quark fraction  $d(x, \mu_F)$  approaches  $\frac{1}{3}$ , and the gluon and sea quark fractions approach zero.

### 2.2.2 Hard Scattering

The hard scattering process is simulated using Monte Carlo generators that sample events with probability proportional to the phase space and matrix element. For

the results contained in this thesis, the primary hard interaction is simulated using the MADGRAPH5 aMC@NLO generator [**Alwall2014**, **Frederix2012**], which can simulate to leading order (LO) in EW vertices and up to next-to-leading order (NLO) in QCD vertices.

### 2.2.3 Parton Shower

The parton shower is a sequence of splittings where one outgoing parton  $c_i$  emits a second soft and/or collinear particle  $j$  [**Sjostrand2015**]. Each splitting has an associated splitting kernel  $P_{c_i \rightarrow c_{ij}}(z)$ , where  $z$  is the momentum fraction carried by the initial parton. The allowed QCD splittings are  $q \rightarrow qg$ ,  $g \rightarrow q\bar{q}$ , and  $g \rightarrow gg$  and the allowed QED splittings are  $f \rightarrow f\gamma$  and  $\gamma \rightarrow f\bar{f}$ . The kernels associated with these splittings are

$$\left| \begin{array}{l} P_{q \rightarrow qg}(z) = \frac{4}{3} \left( \frac{1+z^2}{1-z} \right) \\ P_{g \rightarrow q\bar{q}}(z) = \frac{1}{2} (z^2 + (1-z)^2) \\ P_{g \rightarrow gg}(z) = 3 \left( \frac{(1-z)(1-z)^2}{z(1-z)} \right) \\ P_{f \rightarrow f\gamma}(z) = Q_f^2 \left( \frac{1+z^2}{1-z} \right) \\ P_{\gamma \rightarrow f\bar{f}}(z) = N_C Q_f^2 (z^2 + (1-z)^2) \end{array} \right. \quad (2.8)$$

where  $Q_f$  is the charge of the fermion and  $N_C$  is the number of color states the fermion can occupy (3 for quarks and 1 for leptons). The cross section of a splitting is given by

$$\frac{d\sigma(ab \rightarrow \{c_i\}j)}{d\sigma(ab \rightarrow \{c_i\})} = P_{c_i \rightarrow c_{ij}}(z) \cdot \frac{\alpha_s}{2\pi} \cdot \frac{d\theta}{\theta} \cdot dz \quad (2.9)$$

where  $\theta$  is the opening angle between  $c_i$  and  $j$ . These cross sections diverge as  $\theta \rightarrow 0$  and  $z \rightarrow 1$ , meaning bare quarks producing many soft and collinear gluons. Then, these gluons further split to  $gg$  and  $q\bar{q}$  pairs, which in turn emit even more soft and collinear gluons and photons. This process continues until the energy of the outgoing partons reaches  $\Lambda_{\text{QCD}}$  at which point hadronization occurs. The final state particles from the shower of a single parton are often collimated into a narrow cone that is reconstructed as a single physics object called a jet.

## 2.2.4 Hadronization

The QCD potential between two quarks can be approximated as  $V(\vec{r}) \approx \kappa r$ , where  $\kappa$  has been measured to be approximately 1 GeV/fm. The linear behavior of the potential is due to the attractive interactions between the gluons mediating the quark-quark interaction which confine the color field between the quarks into a tube 1 fm in diameter. As the quarks separate, the energy contained in this gluon tube increases linearly until it exceeds the mass of a  $q\bar{q}$  pair. At this point, a new  $q\bar{q}$  pair pops into existence through a quantum mechanical tunneling process, splitting the tube in two. Due to the difference in quark masses, only up, down, and strange quarks are produced, in a 10:10:3 ratio. This process continues until the energy of all the quarks have low enough energy to combine into stable hadrons.

The above procedure is a qualitative description of the Lund string model [**Anderson1983**]. The Pythia event generator models hadronization using the Lund string model as well as the parton shower effects described in the previous section. All results in this thesis use the Pythia 8.2 program [**Sjostrand2015**] to simulate the parton shower and hadronization processes.

# Chapter 3

## The CMS Detector

The Compact Muon Solenoid (CMS) detector [CMS2008] 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  $\tau$  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.

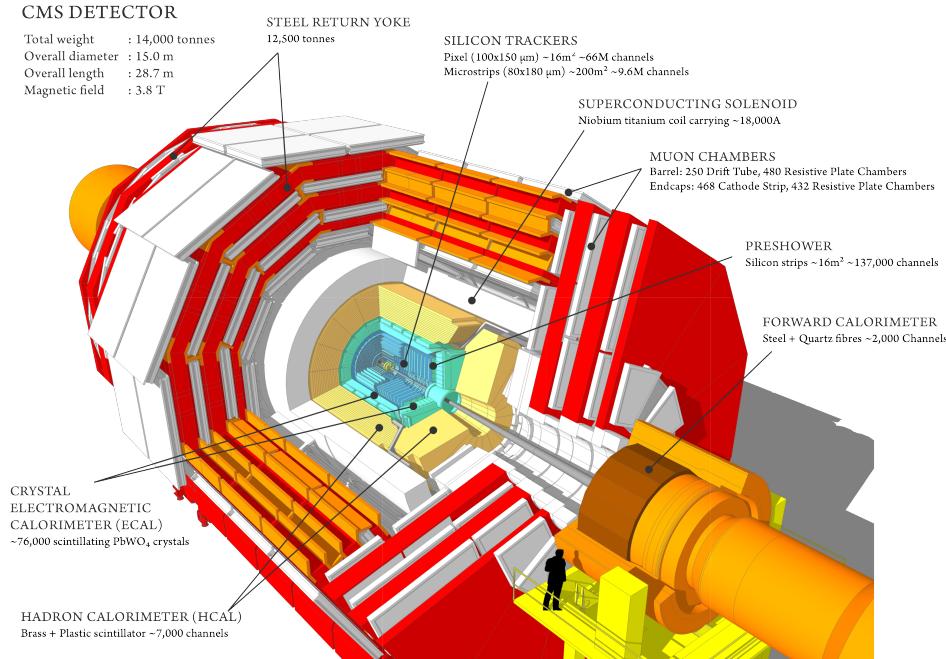


Figure 3-1: A cutaway view of the CMS detector. The labels identify the solenoid as well as the different subdetectors and their components. Reprinted from Reference [CmsWeb].

The overall layout of the CMS detector is shown in Figure 3-1. 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  $\theta$  measured with respects to the positive  $z$ -axis

- azimuthal angle  $\phi$  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 are not as discussed in Section 2.2. 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  $p_T$ ,  $\eta$ , and  $\phi$ , 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  $p = (p_T, \eta, \phi, E)$ . 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  $-5 \leq \eta \leq 5$ .

### 3.1 Inner Trackers

Closest to the interaction point, the inner trackers identify charged particles and measure their four momenta and the best estimate for the point of origin [**CMS2008**]. 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 in the innermost layers and strips in the remaining outer layers. 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 3-2.

The 66 million individual pixel sensors, each measuring  $285 \mu\text{m} \times 100 \mu\text{m} \times 150 \mu\text{m}$

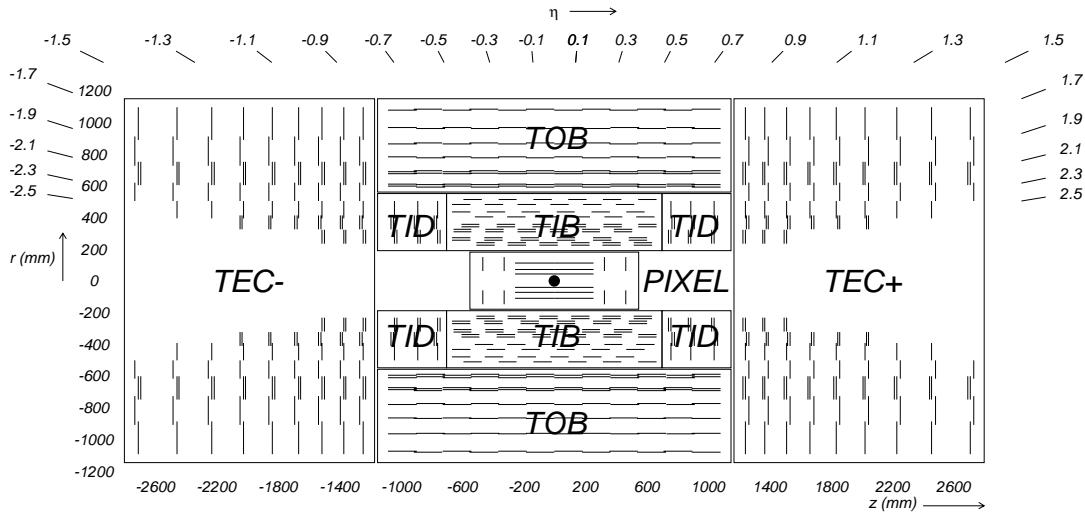


Figure 3-2: A schematic view of the CMS inner tracker system. Silicon pixel and strip detectors are shown. The volumes labeled TIB, TID, TOB, and TEC are all strip trackers. The double lines indicate back-to-back modules that deliver stereo hits. Reprinted from Reference [CMS2008].

in  $r \times r\phi \times z$ , are arranged into seven layers: three cylindrical barrels at  $r = 4.4, 7.3, 10.2$  cm and two bi-layer endcap disks at  $z = \pm 34.5, \pm 46.5$  cm. Due to the geometry of the pixel detector, tracks typically cross the sensor at a  $20^\circ$  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 by 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 the outer six layers constitute the Tracker Outer Barrel (TOB) with an outer radius of 116 cm and an 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 from  $124\text{ cm} < |z| < 282\text{ cm}$ .

The majority of the strips are oriented perpendicular to the  $\phi$  direction: parallel to the beam pipe in the barrel region and radially aligned in the endcap region. The strip pitch varies from  $80$  to  $184\,\mu\text{m}$  with the smallest pitch in the innermost layer. This detector geometry provides good resolution in the  $r$ - $\phi$  plane for barrel and the  $z$ - $\phi$  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  $10$ - $50\,\mu\text{m}$  in the direction perpendicular to the strips and  $100$ - $530\,\mu\text{m}$  in the parallel direction on the stereo modules.

## 3.2 Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL) is a homogeneous, hermetic calorimeter composed of 76,000 lead tungstate ( $\text{PbWO}_4$ ) crystals [CMS2008]. High density ( $8.3\,\text{g}/\text{cm}^3$ ) lead tungstate was chosen due to its radiation hardness, fast scintillation decay time constant of  $25\,\text{ns}$ , small Moliere radius  $r_M = 21.9\,\text{mm}$ , and short radiation length  $X_0 = 8.9\,\text{mm}$ . All of these factors combine to enable the construction of a compact calorimeter with high granularity and excellent energy resolution.

Figure 3-3 shows the layout of the ECAL. The central barrel region (EB) has 61200 crystals arranged in a  $170 \times 360$   $\eta$ - $\phi$  grid ( $0.0174 \times 0.0174$  granularity) with a coverage up to  $|\eta| = 1.44$  while the two endcap disks (EE) each have 7324 crystals organized in a  $x$ - $y$  grid with coverage in the range  $1.479 < |\eta| < 3.0$ . Each crystal has a truncated pyramidal shape with a length of  $230\,\text{mm}$ , a  $22\,\text{mm} \times 22\,\text{mm}$  front-face cross-section, and a  $26\,\text{mm} \times 26\,\text{mm}$  rear-face cross-section in the EB and a length of  $220\,\text{mm}$ , a  $28.6\,\text{mm} \times 28.6\,\text{mm}$  front-face cross-section, and a  $30\,\text{mm} \times 30\,\text{mm}$  rear-face cross-section in the EE. 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

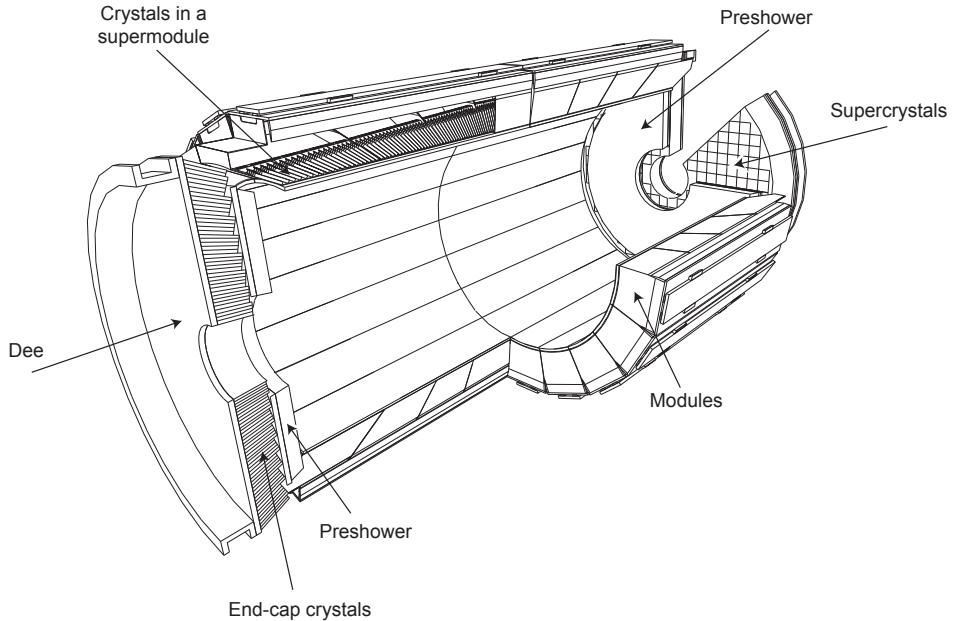


Figure 3-3: The layout of the CMS electromagnetic calorimeter. The barrel and endcap calorimeters are shown. The pre-shower detector sits in front of the endcaps. Reprinted from Reference [CMS2008].

the likelihood of the primary photon or electron emerging from the hard scattering passing through the cracks between two crystals or depositing a large fraction of its energy in passive material, the crystals do not point directly to the interaction region.

The interplay between the 3.8 T magnetic field and the differing 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  $25 \text{ mm}^2$  are glued to the rear of each crystal in the barrel while only one VPT with an active area of  $280 \text{ mm}^2$  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

$$\frac{\sigma_E}{E} = \frac{2.8\%}{\sqrt{E/\text{GeV}}} \oplus \frac{12\%}{E/\text{GeV}} \oplus 0.3\%, \quad (3.1)$$

where  $E$  is the energy of the incident particle and the three terms on the right-hand side 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 photostatistics 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.

### 3.2.1 Preshower Detector

At high momenta and high  $|\eta|$ , the two photons from a  $\pi^0$  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  $\mathcal{O}(1 - 10)$  mm, enabling the disambiguation of two nearly collinear photons and the identification of  $\pi^0$  decays.

### 3.3 Hadronic Calorimeter

The hadronic calorimeter (HCAL) is a set of four heterogenous calorimeters that provide hermetic coverage when combined [CMS2008]: 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  $0.087 \times 0.087$  for  $|\eta| < 1.6$ ,  $0.17 \times 0.17$  for  $1.6 < |\eta| < 3.0$ , and  $0.175 \times 0.175$  for  $3 < |\eta| < 5$ . The layout of the HCAL is shown in Figure 3-4.

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  $5.82 \lambda_I$  at  $|\eta| = 0$  to a maximum of  $10.6 \lambda_I$  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 and close to the beam line, 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.

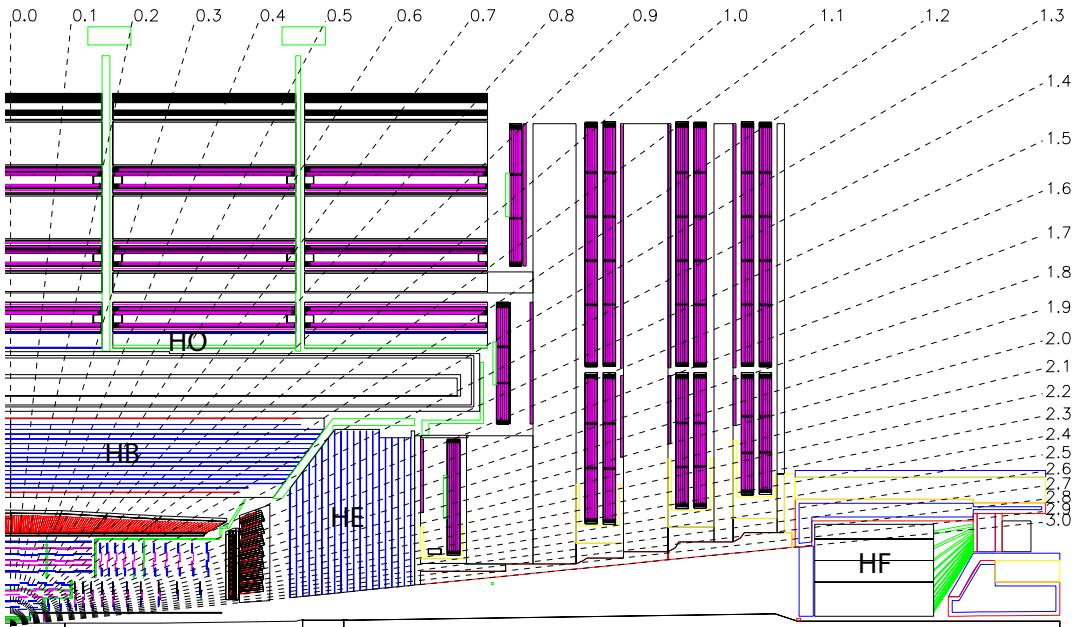


Figure 3-4: The layout of the CMS hadronic calorimeter. The barrel, endcap, outer, and forward calorimeters are shown and labeled. The muon chambers are shown but not labeled. The dashed lines denote different values of pseudorapidity. Reprinted from Reference [CMS2008].

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 [CMS2008], the combined ECAL+HCAL energy resolution was measured to be

$$\frac{\sigma_E}{E} = \frac{0.847}{\sqrt{E/\text{GeV}}} \oplus 0.074 \quad (3.2)$$

and the standalone HF energy resolution is

$$\frac{\sigma_E}{E} = \frac{1.98}{\sqrt{E/\text{GeV}}} \oplus 0.09. \quad (3.3)$$

### 3.4 Muon Chambers

The outer most components of CMS are the muon triggering, identification, and detection chambers [CMS2008]. 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 because 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 coering  $0.9 < |\eta| < 2.4$ , and resistive plate chambers (RPCs) in both covering  $|\eta| < 2.1$ . The layout of the muon detectors is shown in Figure 3-5.

The DT chambers consist of rectangular drift cells with tranverse dimensions of 42 mm  $\times$  13 mm filled with a 85:15 Ar:CO<sub>2</sub> mix and a gold/steel anode wire held at a voltage of 3.6 kV. The maximum drift time per cell of about 400 ns provides a spatial

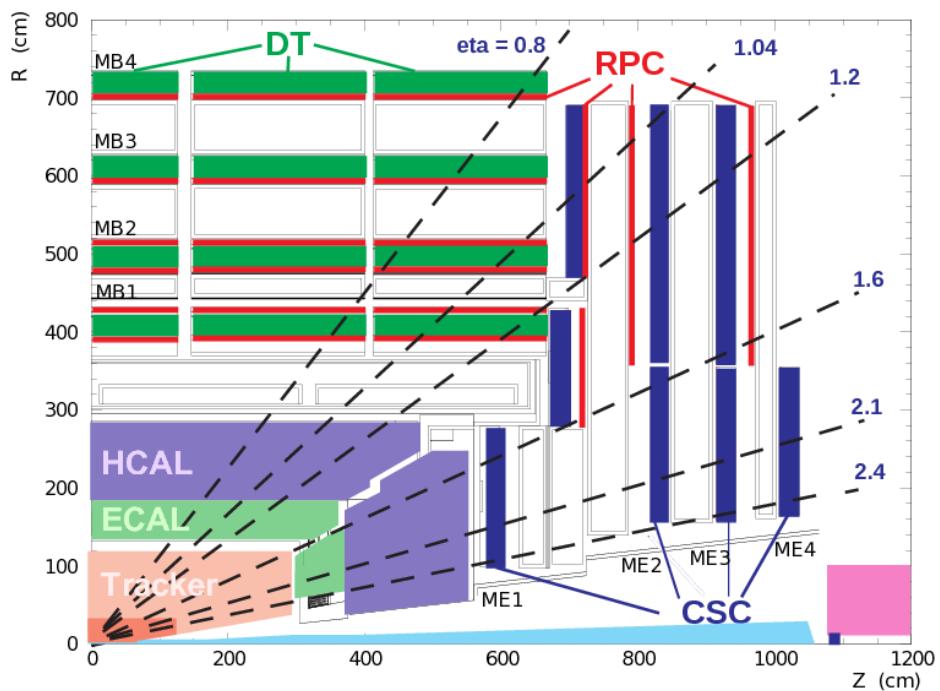


Figure 3-5: The layout of the CMS muon chambers. The four DT stations are labeled MB1-MB4, the four CSC stations are labeled ME1-ME4, and the RPC stations are shown in red. The dashed lines denote different values of pseudorapidity. Reprinted from Reference [CMS2008].

resolution of approximately  $250\,\mu\text{m}$ . A single chamber is made out of superlayers which are in turn made out of four individual cells for a combined resolution of  $100\,\mu\text{m}$ . The chambers are arranged into four  $2.4\,\text{m}$  thick dodecagonal rings called the muon barrel stations. All four stations have two superlayers per chamber that measure position in the  $r$ - $\phi$  plane while the chambers in the inner three stations have an additional superlayer that measures position in the  $r$ - $z$  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  $\text{CO}_2:\text{Ar}:\text{CF}_4$  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  $r$ - $\phi$  plane while the wires that run perpendicular measure the  $\eta$  and beam-crossing time of the muons. The four muon stations in each endcap are made of CSCs arranged into disks oriented perpendicular to the beam axis and interspersed between the flux return plates. The three inner stations have multiple disks 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 an excellent 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, a task the DT and CSCs cannot accomplish because of their 400 ns time resolution.

### 3.5 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. Most new physics processes have a predicted cross-section on the order of picobarns and even the highest rate SM EWK cross-sections are  $\mathcal{O}(10) \text{ nb}$ . Thus, not every event needs to be readout, reconstructed, and written to tape to capture all of the interesting physics at the LHC.

A two-stage trigger system selects the events to keep for permanent storage and analysis [CMS2008]. The Level 1 hardware-based trigger (L1) selects interesting events based on less precise and less granular 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. The readout is highly parallelized because events are stored in a processing pipeline until the L1 decision is made. Events selected by the HLT are sent to the offline computing resources for full reconstruction followed by storage on disk and tape.

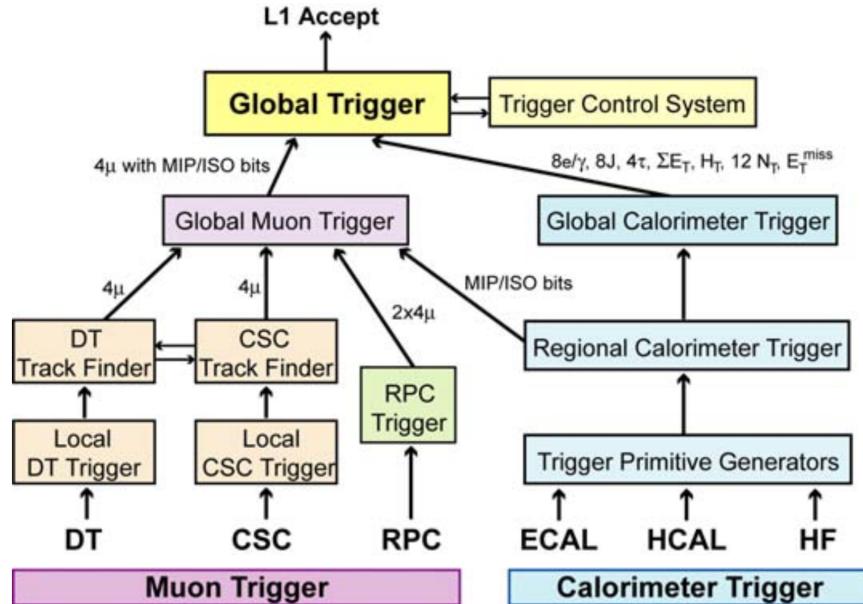


Figure 3-6: The Level-1 trigger architecture with components labeled and information flow indicated. Reprinted from Reference [CMS2008].

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 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 3-6 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 interested 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 targetting 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.

### 3.6 Detector Simulation

The Geant4 program [**Geant2003**, **Geant2006**] 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 2.2, 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 onto 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. The reconstruction software and output format are the same for both up to the retention of additional truth information from the generators.



# Chapter 4

## 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 reconstruction of 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  $\tau$  lepton and  $b$  hadron decays from the tracker. However, as depicted in Figure 4-1, each type of particle interacts with multiple different subdetectors and the best reconstruction is achieved by combining the information from all subdetectors into a single global event description.

The Particle Flow (PF) algorithm [**PF2017**] 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 because 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, to achieve the most accurate measurement

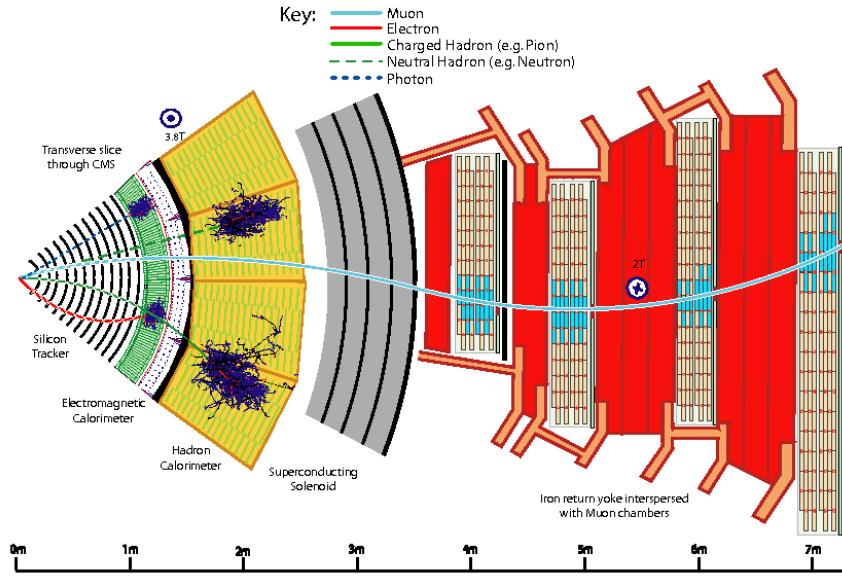


Figure 4-1: A sketch of a transverse slice of the CMS detector showing particle interactions from the interaction point to the muon detector. Reprinted from Reference [PF2017].

of a single particle and avoid double counting. The input detector signals are the tracks, vertices, calorimeter clusters, and muon segments described in Section 4.1. Based on their proximity in the  $\eta\text{-}\phi$ , 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 4.2. After each step, all 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 4.2.5 and 4.2.6, respectively.

## 4.1 Particle Flow Elements

### 4.1.1 Tracks

The combinatorial track finder software [**Tracker2014**] is used to reconstruct tracks in an iterative inside-out process. Initial interactions 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  $\rho$ , the azimuthal angle  $\phi_0$ , the transverse impact parameter  $d_0$ , the longitudinal impact parameter  $z_0$ , and  $\lambda = \cot \theta$ , where  $\theta$  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  $\chi^2$  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 Bremsstrahlung that electrons experience within the tracker [**PF2017**]. 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.

#### 4.1.2 Primary Vertexing

A deterministic annealing (DA) algorithm is used to associate tracks to primary vertices [Tracker2014]. Tracks must pass additional requirements on the transverse impact parameter  $d_0$ , the number of strip and pixel hits, and the reduced  $\chi$  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”

$$F = -T \sum_i^{N_T} \ln \sum_j^{N_V} p_{ij} \rho_j \exp \left[ -\frac{1}{T} \left( \frac{z_i^T - z_j^V}{\sigma_i^z} \right)^2 \right], \quad (4.1)$$

where the  $z_j^V$  are the vertex positions with weights  $\rho_j$ , the  $z_i^T$  and  $\sigma_i^z$  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  $N_T$  total tracks to the vertex  $j$  of  $N_V$  total vertices.

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

$$T_C^j = 2 \sum_i \frac{p_i p_{ij}}{(\sigma_i^z)^2} \left( \frac{z_i^T - z_j^V}{\sigma_i^z} \right)^2 \Bigg/ \sum_i \frac{p_i p_{ij}}{(\sigma_i^z)^2}. \quad (4.2)$$

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

$$S_T = \sum_i (p_T^i)^2 + (p_T^{\text{miss}})^2, \quad (4.3)$$

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

### 4.1.3 Secondary Vertexing

Somewhat longer-lived particles such as  $b$  hadrons and  $\tau$  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 [Vertex2018].

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  $\sqrt{d_0^2 + z_0^2} \geq 50 \mu\text{m}$  to the primary vertex. Tracks are assigned to a secondary vertex based on their opening angle with the seed track and distance of 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 of 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 and any vertices with only one track are removed. 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.

### 4.1.4 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. Electron trajectories bend in the magnetic field which leads to the resulting electromagnetic (EM) shower being significantly spread in the  $\phi$ -direction and collimated in the  $\eta$ -direction. The ECAL reconstruction algorithm [**Electrons2015**, **Photons2015**] combines the basic cluster from each showered particle into a supercluster representing the initial electron or photon from the hard scattering.

At the start of the clustering process, seed crystals are identified as those with a greater transverse energy than their 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 strip of five crystals in the  $\eta$ -direction centered on the seed crystal. This strip is extended into a  $5 \times N$  matrix around the seed crystal in the  $\phi$ -direction up to  $|\Delta\phi| \leq 0.3$  by adding the adjacent strips of five crystals in  $\eta$  if the energy of the crystals in the strip exceeds a certain threshold. This contiguous  $5 \times N$  matrix is grouped into distinct basic clusters each containing a seed strip with energy greater than another threshold. The supercluster is the collection of basic clusters found in the  $5 \times N$  matrix centered on the initial seed crystal. Since the crystals in the endcaps are arranged in an  $x$ - $y$  grid, clustering here uses fixed  $5 \times 5$  matrices of crystals. After a seed cluster is identified, additional, partially overlapping  $5 \times 5$  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  $5 \times 5$  matrices.

#### 4.1.5 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 [**PF2017**]. Similar to the supercluster algorithm, seed cells in the HCAL are identified as those with a greater transverse energy than their 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.

A single topological cluster can result from several particles leaving overlapping energy deposits close to each other. To disentangle the separate contributions, an iterative Gaussian mixture model is used to break each topological cluster of  $M$  individual cells 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  $\eta\phi$  plane, and width  $\sigma$  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

$$f_{ji} = \frac{\mathcal{N}(\vec{c}_j|A_i, \vec{\mu}_i, \sigma)}{\sum_k^N \mathcal{N}(\vec{c}_j|A_k, \vec{\mu}_k, \sigma)}. \quad (4.4)$$

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

$$A_i = \sum_j^M f_{ji} E_j \quad \left| \quad \vec{\mu}_i = \sum_j^M f_{ji} E_j \vec{c}_j, \right. \quad (4.5)$$

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.

#### 4.1.6 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 4.1.1. A full track constructed in this way is referred to as a standalone muon [PF2017].

### 4.1.7 Isolation

While not a physics object in the traditional sense of representing a fundamental particle from the hard scattering, isolation is a powerful variable used in the PF algorithm to distinguish prompt leptons and photons originating in the hard scattering from those originating in the decays of hadrons 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  $I_\gamma$  is the  $E_T$  sum of the PF photons defined in Section 4.2.3 while the charged hadron and neutral hadron isolations  $I_{\text{CH}}$  and  $I_{\text{NH}}$  are the  $p_T$  sums of the PF charged and neutral hadrons defined in Section 4.2.4, 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 4.1.2. 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  $I_{\text{CH}}$  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  $\rho$  of the pileup interactions in the isolation cone is calculated using the effective areas given in Table 4.1 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  $I_{\text{PU}}$  is defined as the half the  $p_{\text{T}}$  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

$$I_{\text{PF}}^{\text{rel}} = \left( I_{\text{CH}} + \max \left\{ 0, I_{\text{NH}} + I_{\gamma} - I_{\text{PU}} \right\} \right) / p_{\text{T}}^{\ell} \quad (4.6)$$

is used when selecting prompt leptons.

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

Table 4.1: Effective areas for isolations.

## 4.2 Particle Identification

### 4.2.1 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 4.1.1. 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  $p_T > 0.5$  GeV. Global and tracker muons sharing the same inner track are merged into a single candidate. For muons with  $p_T < 200$  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 as 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  $\chi^2$  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 4.2.6. The total efficiency of muon reconstruction is 99%.

The work described in this thesis only considers global muons with  $p_T > 10$  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. In order for a muon to pass the tight ID, it must have  $p_T > 30$  GeV and  $I_{\text{PF}}^{\text{rel}} < 0.15$  as well as satisfying the additional requirements in Table 4.2.

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 mm	reject cosmic ray muons
$d_z$	< 5 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 4.2: Selections for the tight muon ID.

### 4.2.2 Electrons

Electron candidates are seeded from the GSF tracks described in Section 4.1.1 as long as the corresponding ECAL clusters are not linked to three or more additional tracks. In each PF 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 ECAL 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  $\eta$ . 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  $|\eta| < 2.5$  that pass cuts on the observables listed in Table 4.3. 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. Additionally, a tight electron must have  $p_T > 30 \text{ GeV}$ .

Variable	Description
$\sigma_{i\eta i\eta}$	energy-weighted cell width in the $\eta$ -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 4.3: Variables used in selecting electrons.

### 4.2.3 Isolated Photons

Photon candidates are seeded from the ECAL superclusters (SCs) described in Section 4.1.4 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 4.2.2 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 4-2 is a profile of the magnitude of the energy correction in bins of  $\sigma_{i\eta i\eta}$ , the energy-weighted cell width in the  $\eta$ -direction.

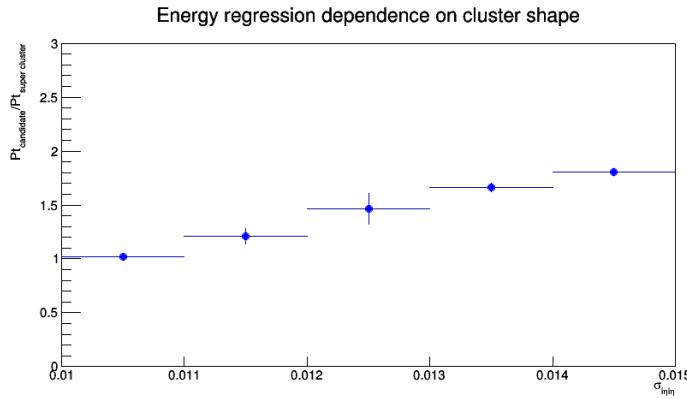


Figure 4-2: Magnitude of the energy correction on the photon object in bins of  $\sigma_{i\eta i\eta}$ .

As an illustration, an unphysically large correction is causing the transverse momentum of the photon object in the event shown in Fig. 4-3 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 has an insignificant effect. Therefore, the unbiased supercluster energy was chosen over the corrected photon energy.

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 4.4, which will hereby be referred to as the  $e/\gamma$

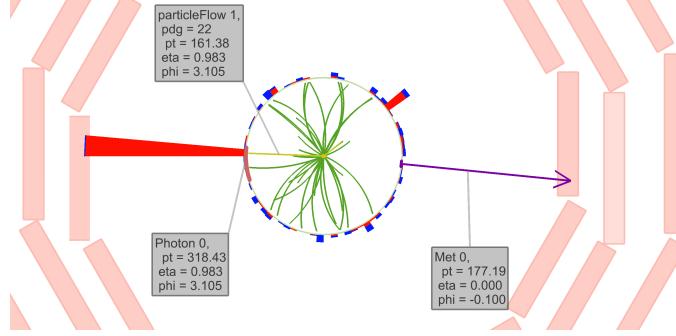


Figure 4-3: An example event where a photon with a wide shower receives a large energy correction.

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 4.5, which combined with the pixel seed veto constitutes the  $\gamma$ -specific ID. The beam halo tagger  $E_{\text{MIP}}$  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  $\sigma_{i\eta i\eta}$  and  $\sigma_{i\phi i\phi}$  as well as the requirement on the cluster seed time  $|t_{\text{seed}}|$  are employed to reject spurious photon objects arising from the “ECAL spikes” discussed in Section ??.

Variable	Maximum Value
$H/E$	0.0260
$\sigma_{i\eta i\eta}$	0.01040
$\rho$ -corrected $I_{\text{CH}}^{\max}$	1.146
$\rho$ -corrected $I_{\text{NH}}$	$2.792 + 0.0112 \times E_{\text{T}}^{\gamma} + 0.000028 \times (E_{\text{T}}^{\gamma})^2$
$\rho$ -corrected $I_{\gamma}$	$2.176 + 0.0043 \times E_{\text{T}}^{\gamma}$

Table 4.4: Selections for the  $e/\gamma$  portion of the photon ID. Isolation values and  $E_{\text{T}}^{\gamma}$  are all in units of GeV.

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

Table 4.5: Additional selections beyond the pixel seed veto for the  $\gamma$ -specific portion of the photon ID.

#### 4.2.4 PF Hadrons

The last candidates reconstructed in a given PF 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 justified because they 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. The energy of this hadron can exceed its momentum due to the presence of neutral particles. Conversely, the momentum can exceed the energy due to the presence of muons from heavy flavor decays or because of mismeasured tracks. The next few paragraphs describe how reconstruction is handled in each of the three cases.

If energy and momentum of the charged hadron are compatible, no neutral par-

ticles or muons are identified. A charged hadron candidate is created for each track linked to the HCAL cluster, with momenta determined by a  $\chi^2$  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 energy of the charged hadron exceeds its momentum by an amount larger than the calorimetric 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 calorimetric 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 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 4.2.1 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 4.1.3, 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.

### 4.2.5 Jets

As discussed in Section 2.2, 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

$$d_{iB} = (p_T^i)^{2q} \quad \left| \quad d_{ij} = \min \left\{ (p_T^i)^{2q}, (p_T^j)^{2q} \right\} \frac{(\Delta R_{ij})^2}{R^2}, \right. \quad (4.7)$$

respectively, where  $q$  and  $R$  are tunable parameters and  $\Delta R_{ij}$  is the angular distance between the two particles. The distance parameter  $R$  is an approximate measure of the cone size  $\Delta R$  of the jet, while the power of the energy scale  $q$  determines the relative importance of 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 [**Cacciari2012**] reduces the computational complexity 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.

#### 4.2.6 Missing Tranverse Energy

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

$$\vec{p}_T^{\text{miss}} = - \sum_{i \in E} \left( \hat{x} \cdot p_T^i \cos \phi + \hat{y} \cdot p_T^i \sin \phi \right), \quad (4.8)$$

and its magnitude is the missing transverse energy  $E_T^{\text{miss}} = |\vec{p}_T^{\text{miss}}|$ . In a perfectly reconstructed event, non-zero  $E_T^{\text{miss}}$  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  $E_T^{\text{miss}}$ .

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  $E_T^{\text{miss}}$  is 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 the measured  $E_T^{\text{miss}}$  is reduced 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  $E_T^{\text{miss}}$  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  $E_T^{\text{miss}}$  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  $\phi$  and a longitudinal track left in the ECAL endcaps and the CSCs. Applying these filters removes essentially all remaining events with fake  $E_T^{\text{miss}}$  while rejecting less than 1% of events with real  $E_T^{\text{miss}}$ .

#### 4.2.7 ECAL gain-switch effect

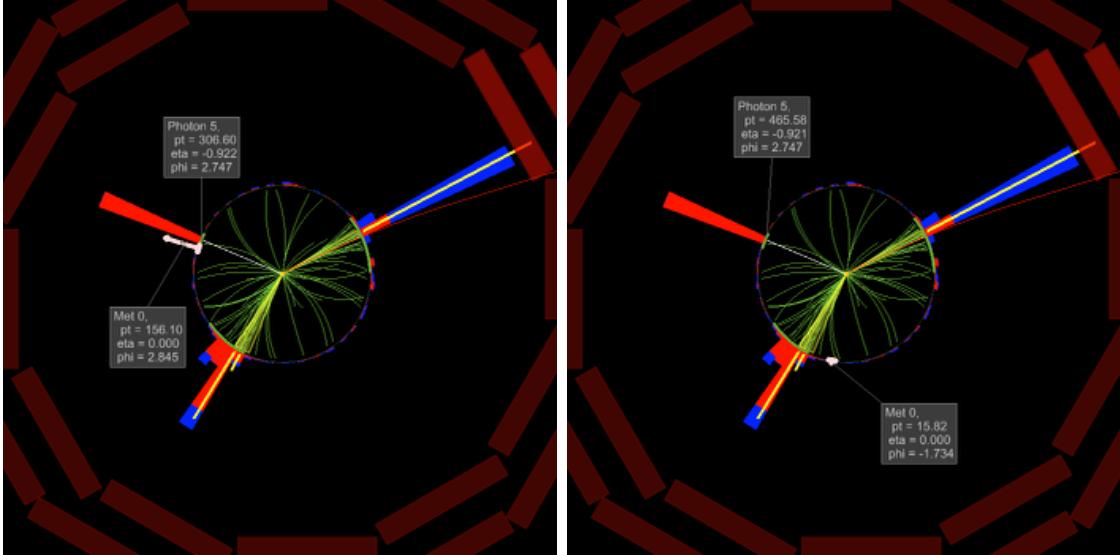


Figure 4-4: Two event displays comparing the same event, reconstructed without (left) and with (right) the fix for ECAL gain-switch effect.

The “multi-fit” algorithm for ECAL hit reconstruction was found to have an unexpected behavior when there is a large energy deposit into a single ECAL crystal. The electronic signal converted at the frontend electronics is partially subject to the default 12 gain amplification and partially subject to the lower gain amplification of 6 or 1. The “multi-fit” algorithm does not account for the switch in amplification and misreconstructs the pulse as a result. 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.

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, typically collinear to the affected photon. Figures 4-4 and 4-5 are the visualization of how the new dataset changes the reconstructed photon energy and  $E_{\text{T}}^{\text{miss}}$ .

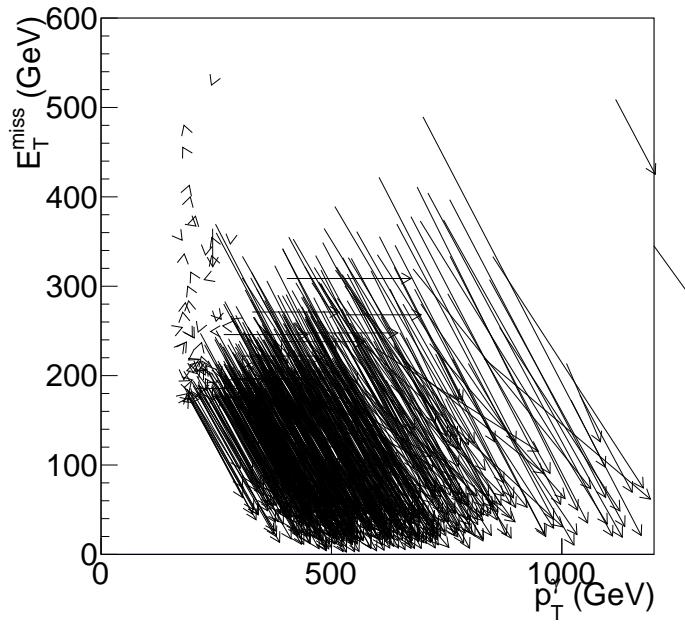


Figure 4-5: The change in reconstructed photon  $p_T$  and  $E_T^{\text{miss}}$  for events in the bin  $\Delta\phi(\gamma, p_T^{\text{miss}}) < 0.05$  of the distribution on the right of Figure ???. Each arrow represents a single event, the tail (head) of the arrow corresponding to  $(E_T^\gamma, E_T^{\text{miss}})$  coordinates in the datasets without (with) the fix for the gain-switch problem.



# Bibliography

- <sup>1</sup>*CMS luminosity measurements for the 2016 data taking period*, CMS Physics Analysis Summary CMS-PAS-LUM-17-001 (CERN, 2017).
- <sup>2</sup>A. M. Sirunyan et al., “Search for new physics in final states with an energetic jet or a hadronically decaying  $w$  or  $z$  boson and transverse momentum imbalance at  $\sqrt{s} = 13$  TeV”, Phys. Rev. D **97**, 092005 (2018).
- <sup>3</sup>G. Bozzi, S. Catani, G. Ferrera, D. de Florian, and M. Grazzini, “Production of Drell-Yan lepton pairs in hadron collisions: Transverse-momentum resummation at next-to-next-to-leading logarithmic accuracy”, Phys. Lett. **B696**, 207–213 (2011).
- <sup>4</sup>A. Denner, S. Dittmaier, M. Hecht, and C. Pasold, “NLO QCD and electroweak corrections to  $W + \gamma$  production with leptonic  $W$ -boson decays”, JHEP **04**, 018 (2015).
- <sup>5</sup>A. Denner, S. Dittmaier, M. Hecht, and C. Pasold, “NLO QCD and electroweak corrections to  $Z + \gamma$  production with leptonic  $Z$ -boson decays”, JHEP **02**, 057 (2016).
- <sup>6</sup>J. M. Lindert et al., “Precise predictions for  $V +$  jets dark matter backgrounds”, Eur. Phys. J. **C77**, 829 (2017).
- <sup>7</sup>M. J. Ambrose et al., “A novel beam halo monitor for the CMS experiment at the LHC”, JINST **10**, P11011 (2015).
- <sup>8</sup>D. Petyt et al., *Characterization and treatment of anomalous signals in the cms electromagnetic calorimeter*, CMS Note 2010/357 (2010).
- <sup>9</sup>D. A. Petyt et al., “Mitigation of anomalous APD signals in the CMS electromagnetic calorimeter”, J. Phys.: Conf. Series **404**, 012043 (2012).

- <sup>10</sup>K. Belasco et al., “The RooStats Project”, PoS **ACAT2010**, Comments: 11 pages, 3 figures, ACAT2010 Conference Proceedings, 057 (2010).
- <sup>11</sup>T. Junk, “Confidence level computation for combining searches with small statistics”, Nucl. Instrum. Meth. A **434**, 435 (1999).
- <sup>12</sup>A. L. Read, “Presentation of search results: the  $\text{cl}_s$  technique”, J. Phys. G **28**, 2693 (2002).
- <sup>13</sup>G. Cowan, K. Cranmer, E. Gross, and O. Vitells, “Asymptotic formulae for likelihood-based tests of new physics”, Eur. Phys. J. C **71**, [Erratum: 10.1140/epjc/s10052-013-2501-z], 1554 (2011).
- <sup>14</sup>P. A. R. Ade et al., “Planck 2015 results. XIII. Cosmological parameters”, Astron. Astrophys. **594**, A13 (2016).
- <sup>15</sup>M. Backović, K. Kong, and M. McCaskey, “MadDM v.1.0: computation of dark matter relic abundance using MadGraph5”, Phys. Dark Univ. **5-6**, 18 (2014).
- <sup>16</sup>M. Aaboud et al., “Search for dark matter at  $\sqrt{s} = 13$  tev in final states containing an energetic photon and large missing transverse momentum with the atlas detector”, The European Physical Journal C **77**, 393 (2017).
- <sup>17</sup>G. Busoni et al., “Recommendations on presenting LHC searches for missing transverse energy signals using simplified  $s$ -channel models of dark matter”, 2016.
- <sup>18</sup>D. Abercrombie et al., “Dark Matter Benchmark Models for Early LHC Run-2 Searches: Report of the ATLAS/CMS Dark Matter Forum”, edited by A. Boveia, C. Doglioni, S. Lowette, S. Malik, and S. Mrenna (2015).
- <sup>19</sup>R. Agnese et al., “New results from the search for low-mass weakly interacting massive particles with the cdms low ionization threshold experiment”, Phys. Rev. Lett. **116**, 071301 (2016).
- <sup>20</sup>D. S. Akerib et al., “Results from a search for dark matter in the complete lux exposure”, Phys. Rev. Lett. **118**, 021303 (2017).
- <sup>21</sup>X. Cui et al., “Dark matter results from 54-ton-day exposure of PandaX-II experiment”, Phys. Rev. Lett. **119**, 181302 (2017).

<sup>22</sup>E. Aprile et al., “Dark matter search results from a one Tonne  $\times$  Year exposure of XENON1T”, 2018.

<sup>23</sup>G. Angloher et al., “Results on light dark matter particles with a low-threshold CRESST-II detector”, Eur. Phys. J. C **76**, 25 (2016).

<sup>24</sup>C. Amole et al., “Dark matter search results from the PICO–60 C3f8 bubble chamber”, Phys. Rev. Lett. **118**, 251301 (2017).

<sup>25</sup>M. Aartsen et al., “Improved limits on dark matter annihilation in the sun with the 79-string IceCube detector and implications for supersymmetry”, Journal of Cosmology and Astroparticle Physics **2016**, 022–022 (2016).

<sup>26</sup>E. Behnke et al., “Final results of the PICASSO dark matter search experiment”, Astropart. Phys. **90**, 85 (2017).

<sup>27</sup>K. Choi et al., “Search for neutrinos from annihilation of captured low-mass dark matter particles in the Sun by Super-Kamiokande”, Phys. Rev. Lett. **114**, 141301 (2015).

<sup>28</sup>The CMS Collaboration, *Simplified likelihood for the re-interpretation of public CMS results*, CMS Note CMS-NOTE-2017-001 (2017).