
**Searching for diboson resonances in the
all-hadronic final state
and
a Lorentz invariance based deep neural network for
W-tagging**

Dissertation
zur
Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der
Mathematisch-naturwissenschaftlichen Fakultät
der
Universität Zürich

von
Thea Klaeboe Arrestad
aus
Norwegen

Promotionskommission:
Prof. Dr. Benjamin Kilminster
Prof. Dr. Florencia Canelli
Prof. Dr. Jesse Thaler
Dr. Andreas Hinzmann

Zürich 2018

Abstract

Zusammenfassung

Die vorliegende Doktorarbeit stellt blabla.....

Contents

1 The Standard Model and Beyond	2
1.1 The Standard Model Lagrangian	2
1.1.1 Particles and fields	2
1.1.2 Electroweak theory	2
1.1.3 The Higgs Mechanism	2
1.1.4 Quantum Chromodynamics	2
1.2 Beyond Standard Model Physics	2
1.2.1 The hierarchy problem and the gravitational force	2
1.2.2 Theories of New Physics	2
1.2.3 Heavy Vector Triplet formalism	2
2 Experimental setup	3
2.1 The Large Hadron Collider	3
2.2 The CMS detector	5
2.2.1 Coordinate system	6
2.2.2 Tracking detectors	6
2.2.3 Electromagnetic calorimeter	8
2.2.4 Hadronic calorimeter	10
2.2.5 Muon chambers	10
2.3 Trigger system: From collision to disk	10
3 Object reconstruction	11
3.1 The Particle Flow Algorithm	11
3.2 Jet reconstruction	11
3.3 Monte Carlo Simulation	11
4 Jet substructure	12
4.1 Jet substructure techniques	12
4.1.1 Jet grooming	12
4.1.2 Jet substructure	12
4.2 W-tagging algorithms	12
4.2.1 Performance	12
4.2.2 Validation	12
5 Diboson resonance searches in CMS	13
5.1 Searches for heavy resonances decaying to dibosons	13
5.1.1 Diboson resonance searches in Run 1: A small bump	13
5.1.2 Searches in the all-hadronic final state	13
5.2 Search I: First search for diboson resonances at 13 TeV	13
5.3 Search II: Developing a new vector boson tagger	13

5.4	Search III: A novel multi-dimensional search	13
6	LoLa: A novel Machine Learning W-tagger for future analyses	14
6.1	Machine Learning: The future of high-performance, low-latency taggers . . .	14
6.2	LoLa: A Lorentz Invariance Based Deep Neural Network for W-tagging . . .	14
6.2.1	Inputs	14
6.2.2	The Combination Layer	14
6.2.3	The Lorentz Layer	14
6.3	Decorrelating from mass and p_T	14
6.4	Performance	14
7	Summary	15
	Bibliography	16

Introduction

A short explanation of NP decaying to dibosons (Vprime, Gravition), jets in the boosted regime (substructure), the search strategy of 2015+2016 (1D) and 2017 (3D). Touch on triboson signatures to emphasise 3D

1 The Standard Model and Beyond

1.1 The Standard Model Lagrangian

1.1.1 Particles and fields

1.1.2 Electroweak theory

1.1.3 The Higgs Mechanism

1.1.4 Quantum Chromodynamics

1.2 Beyond Standard Model Physics

1.2.1 The hierarchy problem and the gravitational force

1.2.2 Theories of New Physics

Warped extra dimensions

Compositeness

1.2.3 Heavy Vector Triplet formalism

2 Experimental setup

2.1 The Large Hadron Collider

In March 1984, the European Organization for Nuclear Research CERN) and the European Committee for Future Accelerators (ECFA) held a workshop in Lausanne entitled "Large Hadron Collider in the LEP Tunnel". This is history's first written mention of the Large Hadron Collider (LHC) and the topic under discussion was exactly how and where to build a new type of high-energy collider, capable of bringing hadrons to collide rather than leptons. The LHC would be housed in a tunnel which, at the time, was under excavation to host the Large Electron-Positron Collider (LEP) designed to collide leptons with center-of-mass energies up to around 200 GeV. LEP was a circular collider with a circumference of 27 km and the tunnel hosting it was located roughly 100 meters underground on the border between France and Switzerland, at the outskirts of Geneva. The justification for building a machine like the LHC, was that once LEP got to maximum reach, a new and more powerful collider would be needed in its place in order to probe higher energies. While collisions of electrons with positrons provided exceptionally clean and precise measurements due to them being point particles, their lightness prevent them from being accelerated to higher energies. Collisions of hadrons, however, would allow for center-of-mass energies two orders of magnitude higher than that of LEP. Therefore, after running a while at two times the W mass (160 GeV) and reaching a maximum center-of-mass energy of 209 GeV, LEP was dismantled in 2000 in order to make room for the LHC.

The Large Hadron Collider started up in September 2008 and, while having the same 27-kilometer radius as the LEP collider, is capable of accelerating protons up to a center-of-mass energy of around 14 TeV, 70 times that of LEP. The accelerator consists of two oppositely going proton beams, isolated from each other and under ultrahigh vacuum, which are accelerated up to speeds close to that of the speed of light through radio frequency (RF) cavities, before being brought to collide at four different interaction points along the ring. These four collision points correspond to the location of the four LHC particle detectors; ATLAS, CMS, LHCb and ALICE. While ATLAS and CMS are general-purpose detectors built in order to study a large range of different physics processes, LHCb and ALICE are built for dedicated purposes; LHCb for b-physics processes and ALICE for heavy ion collision. A protons journey from gas to one of the LHC collision points is as follows: First, hydrogen nuclei are extracted from a small tank of compressed hydrogen gas and stripped of their electrons. The remaining protons are then injected into the LINAC2, a linear accelerator responsible for increasing the proton energy to about 50 MeV through RF cavities that push charged particles forward by switching from positive to negative electric fields. LINAC2 additionally divides the constant stream of particles into equally spaced "bunches" by careful tuning of the frequency of the field switch. The accelerated protons are then injected into the Proton Synchrotron Booster (PSB), where their energy is increased thirty folds more, to an energy of roughly 1.4 GeV. The two final acceleration stages before the protons reach the LHC ring are the Proton Synchrotron and Super Proton Synchrotron, eventually leaving the

protons with a total energy of 450 GeV. The protons are now ready for the final stage of their travel and are injected into the two beam pipes of the LHC in oppositely going direction. They are injected in trains of 144 bunches each (with an order of 10^{11} protons per bunch), where each bunch is roughly 7.5 meters apart (or 25 ns). There are some larger beam gaps present in each beam in order to give the beam dump and injection kickers sufficient time to reach full voltage, where the largest one, the beam abort gap, is roughly 3 ms or 900 m long. The ring is filled with proton bunches until these are equally distributed throughout the two rings, a process taking roughly 4 minutes. This is called a "fill". Here, the protons are accelerated to their maximum energy of 6.5 TeV, a process taking roughly 20 minutes, through eight RF cavities. These RF cavities are also responsible of keeping the proton bunches tightly bunched, ensuring maximum luminosity at the four collision points. A complete sketch of the CERN accelerator complex is shown in Figure 2.1.

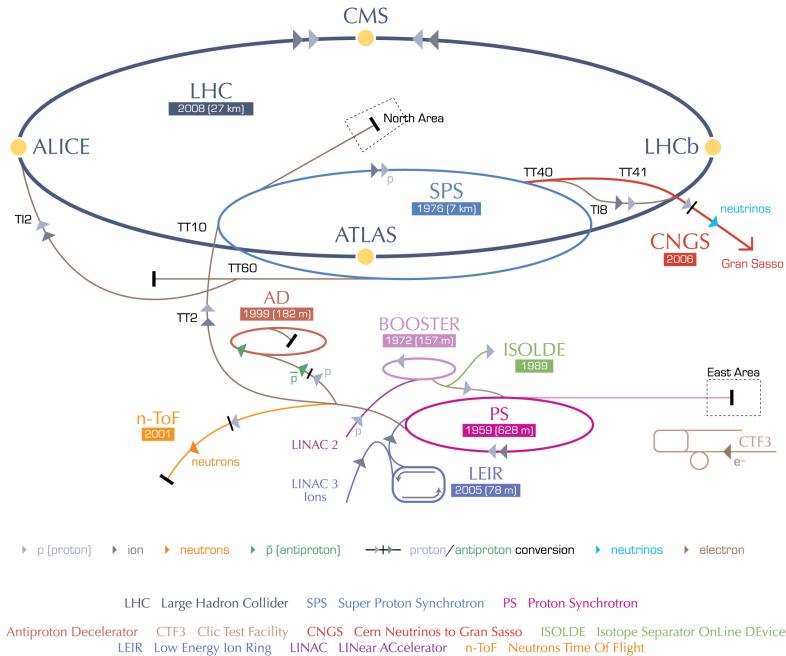


Figure 2.1: The Large Hadron Collider accelerator complex. The four collision points along the ring correspond to the location of the LHC particle detectors CMS, LHCb, ATLAS and ALICE [1].

After the beams have reached their maximum energy and are stably circulating in the LHC ring, they are brought to collide. The goal of such a collision, which occurs every 25 nano seconds, is that some of the protons will undergo an inelastic collision, allowing the quark/gluon constituents of each proton to interact with one another and produce new and interesting particles. The number of times such an interaction will take place inside a detector per area and time is quantified through the luminosity, \mathcal{L} , which is the proportionality factor between the number of observable events per second, and the cross section σ of the process you are interested in

$$\frac{dN_{events}}{dt} = \mathcal{L}\sigma. \quad (2.1)$$

The cross section is the probability that an event (like one which would produce new and interesting particles) will occur and is measured in barns, where 1 barn = 10^{-28} m². This proportionality factor should therefore be as high as possible. It depends only on parameters of the detector and can, in the case of LHC, be defined through the following accelerator quantities

$$\mathcal{L} = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi \epsilon_n \beta_*} F, \quad (2.2)$$

where N_b is the number of particles per bunch, n_b is the number of bunches, f_{rev} is their revolution frequency, γ_r is the relativistic gamma factor, ϵ_n is the transverse beam emittance (how confined the particles are in space and momentum), β_* is the beta function at the collision point (how narrow, or "squeezed", the beam is) and F is a reduction factor to account for a constellation where the beams do not collide heads-on but at slight crossing angles. From this, it becomes clear that the main goal of the LHC is to; maximize the number of particles (N_b, n_b), their frequency (f_{rev}) and their energy (γ_r), while at the same time ensuring the protons are packed together as tightly as possible (lower ϵ_n and β_*). Using the nominal values of the LHC, the peak luminosity is roughly $\mathcal{L} \sim 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$.

The peak luminosity of the LHC by the end of Run 2 in 2018 was grazing around $2.0 \cdot 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, corresponding to 2 times the nominal design luminosity.

To quantify the size and statistical power of a given LHC dataset, the integrated luminosity is used. This is the integral of the instantaneous luminosity over time and is defined as

$$\mathcal{L}_{int} = \int \mathcal{L} dt. \quad (2.3)$$

It is usually defined in units of inverse cross section, b⁻¹.

Despite the LHC starting up in 2008, there would be another year before data taking began. In March 2010, the LHC saw its first collision with a center-of-mass energy of 7 TeV, and continued running at this energy collecting around 5 inverse femtobarns of data by the end of 2011. In 2012, the energy was increased to 8 TeV and the LHC continued running until a planned long shutdown scheduled to begin in February 2013, collecting a total of $\sim 20 \text{ fb}^{-1}$ and discovering the Higgs boson. This marked the end of Run 1 and the beginning of a two-year maintenance project intended to prepare the LHC for running at a center-of-mass energy of 13 TeV; Run 2.

Run 2, and where this thesis begins, started in June 2015. With the accelerator now running at 90% of its nominal energy, and with a peak luminosity between 1-2 times the design luminosity, the LHC managed to collect an impressive $\sim 160 \text{ fb}^{-1}$ at this energy until its planned shutdown at the end of 2018. Some key LHC accelerator parameters that were in use for the datasets analyzed in this thesis, are quoted in Table 2.1

2.2 The CMS detector

The Compact Muon Solenoid (CMS) detector is true to its name; with a diameter of 15 meters and a weight of 14000 tons, it is 60 % smaller but two times heavier than its general purpose counterpart, the ATLAS detector. Its large weight is due to the CMS housing the world's largest and most powerful solenoid: A superconducting niobium titanium coil circulating 18500 Amps and capable of generating a magnetic field of 3.8 Tesla. Together with its corresponding iron return yoke, responsible for reflecting the escaping magnetic flux, it accounts for 90% of the total detector weight. The CMS detector is cylindrically symmetric and organized in the following way: closest to the beam pipe and at a radius of about 3

Parameter	Units	Nominal	2015	2016	2017
Energy	[TeV]	7.0	6.5	6.5	6.5
Bunch spacing	[ns]	25	25	25	25
Bunch intensity	$\times 10^{11}$ [protons/bunch]	1.15	1.15	1.15	1.2-1.45
Bunches per train		144	144	96	144
Total number of bunches		2808	2244	2220	2556
β^*	[cm]	55	80	40	27/25
Peak luminosity	$\times 10^{34}[\text{cm}^{-2}\text{s}^{-1}]$	1.0	0.5	1.4	2.0
Integrated luminosity			4.2	39.7	50.2

Table 2.1: Some key LHC detector parameters achieved during the first years of 13 TeV data taking

cm, the inner tracking system begins. It consists of an inner silicon pixel detector and an outer silicon strip tracker, stretching out to a radius of roughly 1.2 meters. Following the tracker are two calorimeter layers: the electromagnetic calorimeter (ECAL) consisting of lead tungstate scintillating crystals and responsible for measuring the energy of electromagnetically interacting particles, followed by the hermetic hadronic calorimeter (HCAL) measuring the energy of hadrons. Contrary to "standard" configurations for general purpose detectors, the CMS calorimeters are located inside the superconducting solenoid. This allows the detector to be rather compact, by reducing the necessary radius of the calorimeters, and additionally for the magnet to be strong enough (the magnetic field strength depends on the coil radius) to allow muon detectors to be located within the magnetic field so their momentum can be measured. The muon detectors are alternated with three layers of steel return yoke responsible for containing and reflecting the magnetic field and which only allows muons and weakly interacting particles to pass. A schematic overview of the CMS detector is shown in Figure 2.2. In the following, the different sub-detectors will be described in detail.

2.2.1 Coordinate system

To describe locations within the CMS detector, a Euclidian space coordinate system is used. Here, the positive z axis points along the beam pipe towards the west, the positive x axis points towards the center of the LHC ring, and the positive y axis upw towards the earths surface. Due to the cylindrical symmetry of the detector, polar coordinates are more convenient and most frequently encountered. In this scheme, the azimuthal angle ϕ is measured in the xy-plane, where $\phi = 0$ correspond to the positive x axis and $\phi = \pi/2$ correspond to the positive y axis. The polar angle θ is measured with respect to the z axis, $\theta = 0$ aligning with the positive and $\theta = \pi$ with the negative z axis. To define a particles angle with respect to the beam line, the pseudorapidity $\eta = -\ln \tan(\theta/2)$ is preferred over θ . This is due to the fact that particle production is approximately constant as a function of pseudorapidity and, more importantly, because differences in pseudorapidity are Lorentz invariant under boosts along the z-axis when assuming massless particles. To measure angular difference between particles in the detector, the variable $R = \sqrt{\eta^2 + \phi^2}$ is used, again Lorentz invariant under longitudinal boosts. A summary of the CMS coordinate system together with some example values are shown in Figure 2.3.

2.2.2 Tracking detectors

The CMS tracker is responsible for accurately reconstructing the momentum of charged particles and consists of two sub-detectors. Closest to the interaction point, and where the

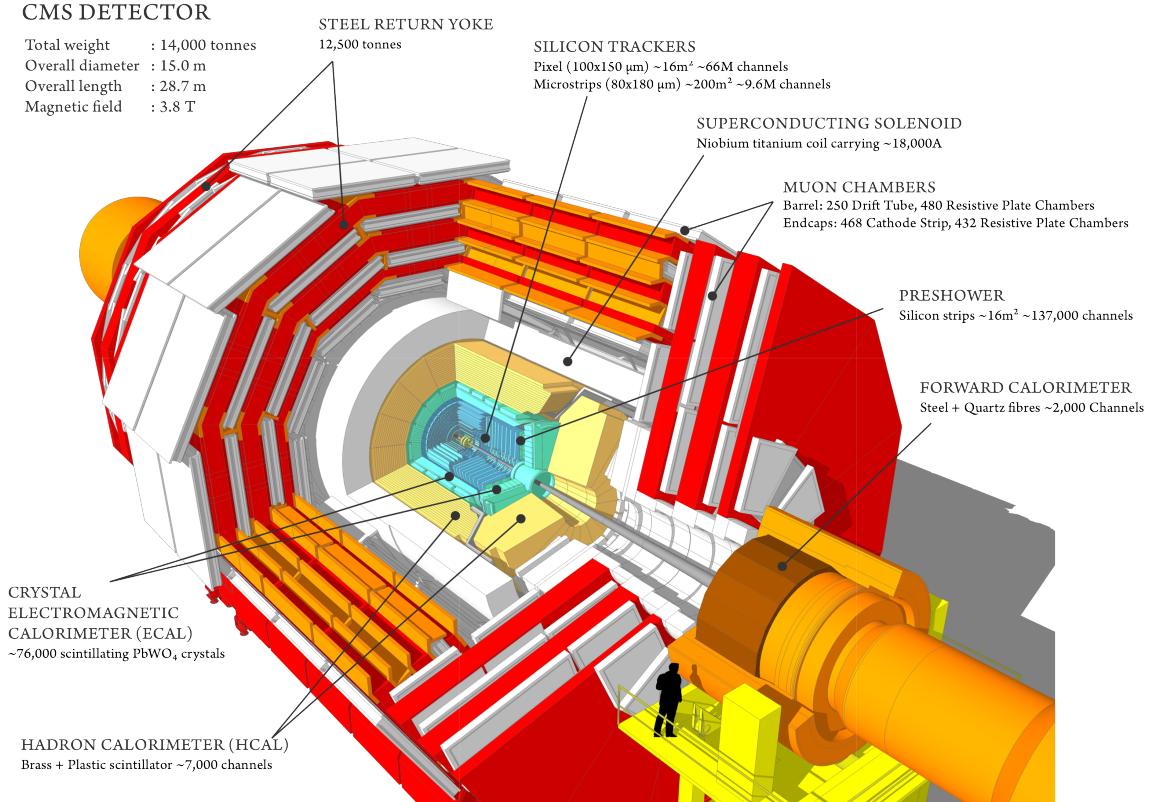


Figure 2.2: The CMS detector and its subsystems: The silicon tracker, electromagnetic and hadron calorimeters, the superconducting solenoid and the muon chambers inter-layered with the steel return yoke [2].

particle intensity is the highest, the silicon pixel detector is located. Upgraded in 2017, from the so-called Phase-0 to the Phase-1 layout, it is structured in four cylindrical barrel layers at radii 2.9, 6.8, 10.9 and 16.0 cm (the barrel pixel) and three disks in each of the forward regions placed at a distance from the nominal interaction point of 29.1, 39.6 and 51.6 cm (the forward pixel). A sketch of the current Phase-I pixel detector compared to the Phase-0 detector is shown in Figure 2.4. The sensors located closest to the beam pipe are subject to hit intensities of $\mathcal{O}(\text{MHz}/\text{mm}^2)$ which puts strict constraints on the maximum sensor size in order to minimize occupancy in the detector. The pixel sensors are $100 \mu\text{m} \times 150 \mu\text{m}$ with a thickness of $285 \mu\text{m}$, and when counting both barrel and pixel sensors, sum up to a total of 79 million. The pixel sensors are mounted on detector modules with 16 read-out chips each, where the type of read-out chip depends on how close the module is to the beam pipe: the inner layer uses read-out chips with a rate capability of $600 \text{ MHz}/\text{cm}^2$ while for the outer layers, read-out chips with a rate capability of up to $200 \text{ MHz}/\text{cm}^2$ are sufficient.

As the hit intensity reduces as you go further away from the beam pipe, the pixel sensors are replaced by silicon strip sensors, making up the second of the two tracker sub-systems, the silicon strip tracker. There are ten strip layers in total, stretching out to a radius of roughly 130 cm. These are divided into four sections: The inner barrel (TIB) with four strip layers, the two inner endcaps (TID) consisting of three disks each, the outer barrel (TOB) consisting of 6 cylindrical layers and the two endcaps (TEC) with 9 strip layers each. A schematic overview of the strip tracker layout is shown in Figure 2.5. The strips in the TIB and TID are 10 cm long, with a width of $80 \mu\text{m}$ and a thickness of $320 \mu\text{m}$. The TOB and TEC sections consist of slightly larger strips of $25 \text{ cm} \times 180 \mu\text{m}$ and a thickness of $500 \mu\text{m}$. The strip

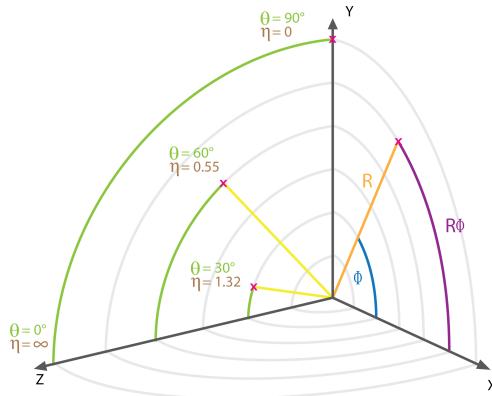


Figure 2.3: The CMS coordinate system [3]

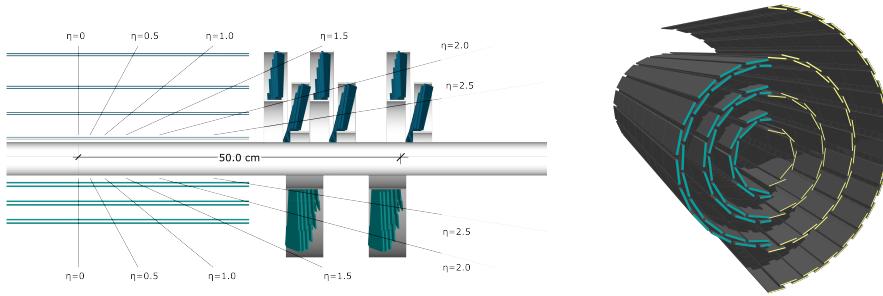


Figure 2.4: Left: The forward pixel detector layout before (bottom) and after (top) the Phase-I upgrade. Right: The barrel pixel detector before (left) and after (right) the Phase-I upgrade [4].

tracker has a total of 10 million detector strips and covers an area of $\sim 200 \text{ m}^2$. To prolong the silicon detector lifetime, the entire tracker (pixel and strip) is kept at a temperature of -20°C through a liquid cooling system. The tracker has a coverage up to $|\eta| < 2.6$ and a resolution of roughly $\sigma/p_T \approx 1.5 \times 10^{-5} p_T + 0.005$.

2.2.3 Electromagnetic calorimeter

Following the tracking detectors is the electromagnetic crystal calorimeter (ECAL). Consisting of 75 848 laterally segmented scintillating lead tungstate (PbWO_4) crystals, it was designed to have the highest possible photon energy and position resolution in order to resolve a Higgs boson decaying into two photons, the cleanest of the Higgs discovery channels. With a goal of a photon/electron energy resolution of 0.5% above 100 GeV, the choice of detector material for the ECAL has been its most crucial design feature. In order to withstand the high doses of radiation and the high magnetic field present within the detector, while at the same time generating well-defined signal responses within the 25 nanoseconds between particle collisions, an extremely dense and transparent material capable of producing fast and clean photon bursts when hit, is required. The choice eventually fell on metal-heavy lead tungstate crystals, each taking roughly two days to artificially grow (and a total of about ten years to grow all of them). With a density of $\delta = 8.28 \text{ g/cm}^3$ (slightly higher than for stainless steel), the crystals are compact enough to yield excellent performance without taking up too much volume, allowing the ECAL to sit within the CMS superconducting solenoid. The homogeneous medium allows for a better energy resolution as it minimizes sampling

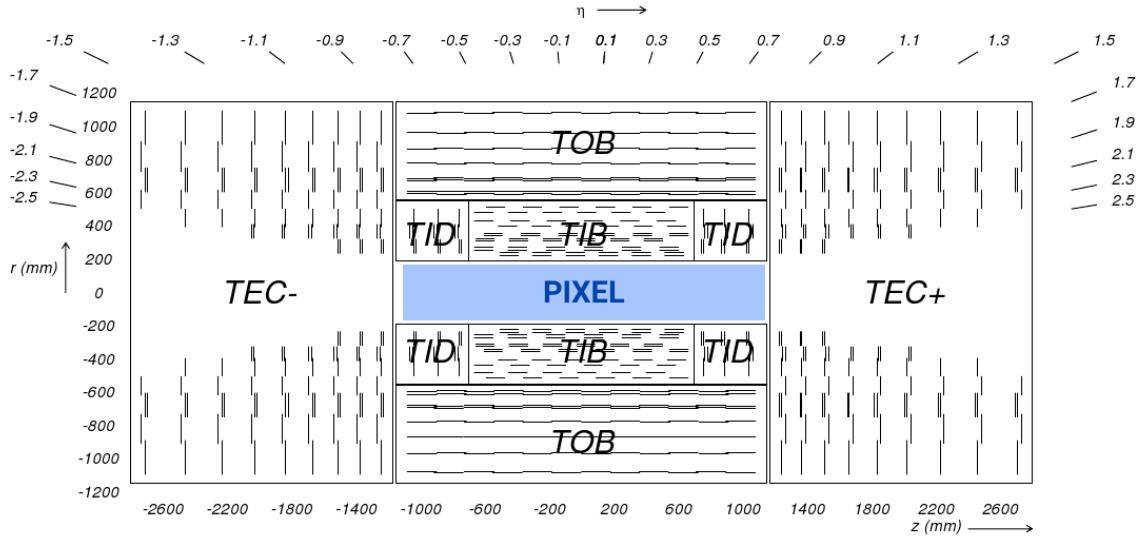


Figure 2.5: Schematic of the CMS silicon strip tracker and its four subsections: The inner barrel (TIB), inner endcaps (TID), the outer barrel (TOB) and the two endcaps (TEC) [5].

fluctuation effects and it additionally contains enough oxygen in crystalline form to make it highly transparent to their entire scintillation emission spectrum. With an extremely short radiation length and small Moliére radius ($X_0 = 0.85$ cm,

$$R_M = 2.19$$

cm), the required homogeneity, granularity and compactness is obtained while at the same time emitting 80% of generated light within the 25 ns timeframe required. The largest drawbacks with a lead tungstate detector is the low light yield (100 γ per MeV), requiring dedicated avalanche photodiodes to increase the gain, as well as a light yield which strongly depends on the temperature. The detector response to an incident electron changes by 3.8 ± 0.4 % per degree Celsius which requires the ECAL temperature to be kept stable around $18.180(5)$ °C, obtained through an intricate water cooling system. The ECAL is completely hermetic and sorted into a barrel part (EB), covering pseudorapidities up to $|\eta| < 1.48$, and two endcap parts (EE) extending the total coverage to $|\eta| < 3.0$ in order to match the tracker coverage of $|\eta| < 2.5$. In order to improve the γ/π_0 separation power, a pre-shower detector (ES) using lead absorbers and silicon sensors covers the forward region between $1.65 < |\eta| < 2.6$. The crystals in the barrel are organized into supermodules, each consisting of about 1700 crystals, while the endcap is divided into two half disks consisting of 3662 crystals each (so-called "Dees"). Each PbWO₄ crystal weighs around 1.5 kilogram and has a slightly tapered shape with a front face of 2.2×2.2 cm² in the barrel and 2.86×2.86 cm² in the endcaps. The crystals are 2.3 and 2.2 cm long in the barrel and endcaps, respectively. The total volume of the calorimeter including barrel and endcaps is 11 m² and weighs a total of 92 tons. The ECAL detector layout is illustrated in Figure 2.6.

Having no longitudinal segmentation, the ECAL relies on an accurate reconstruction of the event primary vertex, provided by the tracker, in order to reconstruct the photon angle correctly.

The obtained energy resolution of the ECAL can be parametrized in three parts: a stochastic, a noise and a constant term [6]. It is given as

$$\frac{\sigma E}{E} = \frac{2.8\%}{\sqrt{E}} \oplus \frac{0.128 \text{ GeV}}{E} \oplus 0.3\%$$

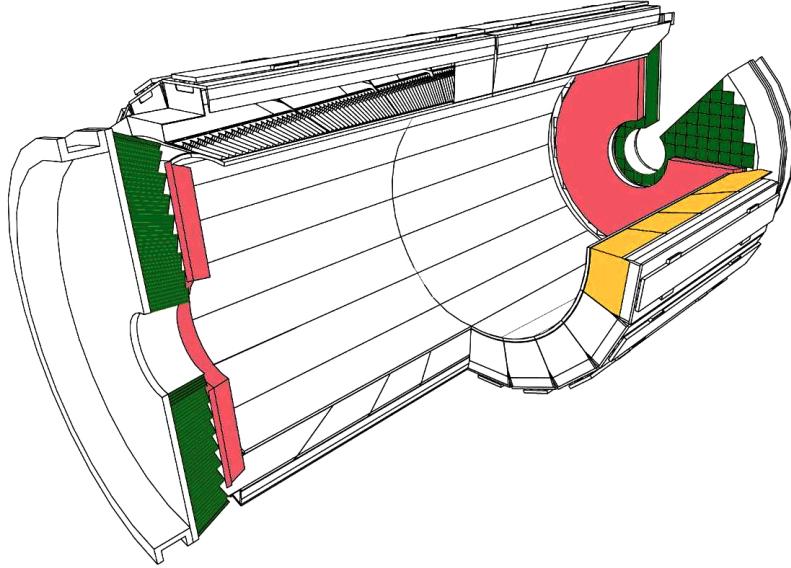


Figure 2.6: A schematic of the CMS electromagnetic calorimeter showing the barrel supermodules (yellow), the individual barrel crystals (black, top left), the endcap modules (green) and the pre-shower detectors (pink) [5].

where the constant values were estimated in an electron test beam. The constant term of 0.3% is dominated by the non-uniformity in longitudinal light collection [7], and one of the main goals of the detector design was to get this term below 1%. The energy resolution as a function of electron energy is shown in Figure 2.7.

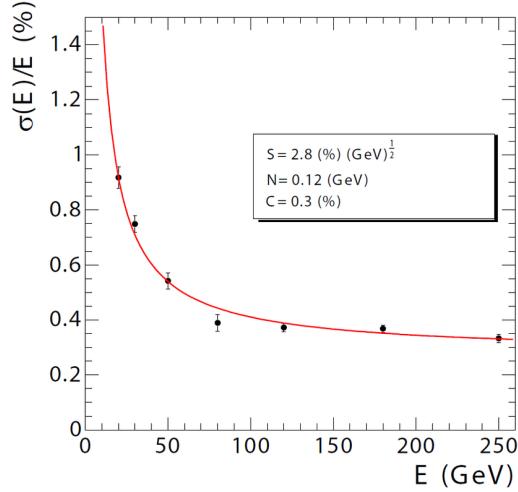


Figure 2.7: The ECAL energy resolution as a function of electron energy as measured in an electron test beam.

2.2.4 Hadron calorimeter

Hadrons are heavier than electrons and muons and usually traverses the ECAL without decaying. Instead they are stopped in the hadronic calorimeter (HCAL) by alternating layers of dense absorber material and plastic scintillators. The hadron calorimeter is split into two parts: the inner

2.2.5 Muon chambers

2.3 Trigger system: From collision to disk

3 Object reconstruction

3.1 The Particle Flow Algorithm

3.2 Jet reconstruction

3.3 Monte Carlo Simulation

4 Jet substructure

4.1 Jet substructure techniques

4.1.1 Jet grooming

4.1.2 Jet substructure

4.2 W-tagging algorithms

4.2.1 Performance

4.2.2 Validation

5 Diboson resonance searches in CMS

5.1 Searches for heavy resonances decaying to dibosons

5.1.1 Diboson resonance searches in Run 1: A small bump

5.1.2 Searches in the all-hadronic final state

5.2 Search I: First search for diboson resonances at 13 TeV

5.3 Search II: Developing a new vector boson tagger

5.4 Search III: A novel multi-dimensional search

6 LoLa: A novel Machine Learning W-tagger for future analyses

- 6.1 Machine Learning: The future of high-performance, low-latency taggers
- 6.2 LoLa: A Lorentz Invariance Based Deep Neural Network for W-tagging
 - 6.2.1 Inputs
 - 6.2.2 The Combination Layer
 - 6.2.3 The Lorentz Layer
- 6.3 Decorrelating from mass and p_T
- 6.4 Performance

7 Summary

Bibliography

- [1] CERN, “Accelerators and Schedules”, 2018 (accessed November 12, 2018),
<https://beams.web.cern.ch/content/accelerators-schedules>.
- [2] CMS, “CMS Detector Design”, 2018 (accessed November 21, 2018),
<https://cms.cern/news/cms-detector-design>.
- [3] T. Lenzi, “Development and Study of Different Muon Track Reconstruction Algorithms for the Level-1 Trigger for the CMS Muon Upgrade with GEM Detectors”, Master’s thesis, U. Brussels (main), 2013.
- [4] A. Dominguez et al., “CMS Technical Design Report for the Pixel Detector Upgrade”, Technical Report CERN-LHCC-2012-016. CMS-TDR-11, CERN, Sep, 2012.
- [5] CMS Collaboration, “The CMS Experiment at the CERN LHC”, *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.
- [6] P. Adzic et al., “Energy resolution of the barrel of the CMS electromagnetic calorimeter”, *JINST* **2** (2007) P04004, doi:10.1088/1748-0221/2/04/P04004.
- [7] C. Biino, “The CMS Electromagnetic Calorimeter: overview, lessons learned during Run 1 and future projections”, *Journal of Physics: Conference Series* **587** (2015), no. 1, 012001.