CMS Electron and Photon Performance at 13 TeV

Jonas Rembser on behalf of the CMS Collaboration
Laboratoire Leprince-Ringuet, Ecole Polytechnique / CNRS-IN2P3, Palaiseau, France
E-mail: jonas.rembser@cern.ch

Abstract. The Compact Muon Solenoid (CMS) detector is one of the two multi-purpose experiments at the Large Hadron Collider (LHC) and has a broad physics program. Many aspects of this program depend on our ability to trigger, reconstruct and identify events with final state electrons, positrons, and photons with the CMS detector with excellent efficiency and high resolution. We present the full process of electron and photon reconstruction in CMS, starting from tracker hits and energy deposits in the electromagnetic calorimeter, the methods to achieve the ultimate precision in Run II energy measurements, the identification strategies (based both on cut-based approach and on multivariate analysis) to discriminate prompt electrons and photons from background, and the methods to estimate the associated systematic uncertainties. Finally the performance on benchmark channels (such as $Z \rightarrow e^{+}e^{-}$) will be shown.

1. Introduction

A performant processing of electrons and photons is essential for maximizing the physics potential of the CMS experiment [1] at the CERN Large Hadron Collider. Besides the well known Higgs boson analyses in the diphoton [2] and four lepton [3] final states, the recent search for high mass resonances decaying to electron pairs [4] illustrates the importance of treating electromagnetic objects with a high efficiency over a wide energy range from a few GeV up to the TeV scale.

This paper gives an overview of electron and photon reconstruction and identification in CMS and discusses the performance on data taken in the last two years. The figures on 2017 performance are the first public electron performance results for this year. The 2016 data has been reprocessed with improved ECAL calibrations [5], hence the potential of Run 2 data after scrutinizing it for about a year is highlighted. Most plots in this paper are taken from a detector performance summary note published in May 2018 [6].

The central feature of the CMS apparatus is a superconducting solenoid of 6 m internal diameter, providing a magnetic field of 3.8 T. Within the solenoid volume are a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter (ECAL) [7], and a brass and scintillator hadron calorimeter (HCAL), each composed of a barrel and two endcap sections.
A more detailed description of the CMS detector, together with a definition of the coordinate system used and the relevant kinematic variables, can be found in Ref. [1].

The treatment of electrons and photons with the CMS detector relies primarily on the ECAL with its 75848 quasi-projective scintillating PbWO4 crystals, extending up to a pseudorapidity of $|\eta| = 3.0$ with a crack between the barrel and endcaps at $|\eta| = 1.479$. The crystals feature 25.8 (24.7) radiation lengths and a front face of $2.2 \times 2.2$ ($2.86 \times 2.86$) cm$^2$ in the barrel (endcaps). In the endcaps, the crystal calorimeter is complemented by a silicon preshower detector of about 3 radiation lengths. Since the footprint of an electromagnetic object can also comprise traces in the inner tracker, information from the silicon tracker is mandatory. The inner tracker encompasses a microstrip detector with 14 (12) layers in the barrel (endcaps), surrounding a highly granular pixel detector.

A pixel detector replacement became necessary between the 2016 and 2017 data taking periods [8]. The original one could not support the expectations-exceeding number of interactions per bunch crossing (pileup), both from a bandwidth and radiation perspective. The new pixel detector features an additional fourth layer of active modules in the barrel and a third disk per endcap, resulting in a four-hit coverage in the whole tracking region. The radius of the innermost layer has been reduced from 44 mm to 29 mm for improved vertex resolution. The material budget in the endcaps has been reduced by up to 50%. This lowers the number of converting photons and bremsstrahlung-emitting electrons, which are challenging for reconstruction algorithms. The tracking system extends up to $|\eta| = 2.5$, defining the fiducial region for electrons and photons.

After the successful first operational run between 2009 and 2013 (Run 1), the Large Hadron Collider is now back in data-taking and closing up the second operational period from 2015 to 2018 (Run 2) at $\sqrt{s} = 13$ TeV and an ever-increasing pileup. The last published papers on electron and photon performance with the CMS detector are dated back to 2015, summarizing the performance at $\sqrt{s} = 8$ TeV during Run 1 [9] [10]. Since then, the reconstruction algorithm was not changed as far as the underlying concepts are concerned. Some improvements were made to better keep track of the energy flow in the event, i.e. to avoid double counting of energy. This is essential for an accurate determination of the missing transverse energy in the event, especially at elevated pileup levels. Equally important is the steady refinement of the identification algorithms to reduce the number of jets misidentified as prompt electrons or photons, as well as alignments and calibration efforts to keep the residual differences between data and Monte Carlo (MC) simulated events at a minimum.
2. Electron and Photon Reconstruction

Figure 1 gives a conceptual overview of the electron and photon reconstruction algorithm taking electrons as an example. On the ECAL side, the energy deposits in the crystals get merged to clusters which topologically match the expectation from an individual particle impacting the calorimeter (single-particle-like clusters). The energy deposit in one crystal may be split among more than one single-particle-like cluster. These clusters get further combined into so-called “superclusters”, matching the pattern of an array of clusters expected from an electromagnetic object that might have radiated or converted before reaching the ECAL. In particular, that means the individual clusters are spread in the azimuthal direction due to the magnetic field. These superclusters seed the track building and fitting algorithm. Electrons suffer large radiative energy losses in material, which are not Gaussian distributed but follow rather the Bethe-Heitler model \cite{11}. Hence, electron tracks are reconstructed using a Gaussian Sum Filter (GSF) instead of the widespread Kalman Filter (KF) \cite{12}. The tracks are in turn used to refine the superclusters. Clusters matching extrapolated track tangents at tracker layers are merged, while conversions are recognized and fitted. The refined supercluster and electron track make up what is called a reconstructed electron. The GSF-based track reconstruction can be also seeded from a general track which is matched to a single-particle-like cluster, complementing the reconstruction at low energies. This tracker-driven seeding is only available for offline reconstruction and not at trigger level. As indicated in Figure 1, energy corrections via multivariate regression are inserted after most clustering steps, instead of just inserting one energy regression at the end of the reconstruction. This is because the intermediate reconstructed objects are used in other algorithms as well, like the CMS particle-flow algorithm \cite{13}.

The reconstruction efficiency for electrons at the $Z$ peak is about 96%, with a slight increase in the endcaps in 2017 due to the new pixel detector with reduced material budget. The improved ECAL calibrations in the recent 2016 data reconstruction were beneficial for the efficiency as well, on the full range of pseudorapidity. MC indicates that the new pixel detector not only yields a higher electron reconstruction efficiency, but also a fake rate about 30% lower at the $Z$ peak than in the previous year. This is attributed to the more robust quadruplet-based track seeding algorithm, made possible by the four-hit coverage which reduces the number of fake tracks in general. So far, the algorithm targeted a maximum efficiency at a fake rate which is still computationally tolerable. It is only in the so-called identification (ID) step where the background gets effectively rejected while keeping a signal efficiency which is tolerable for the analysis at hand. This step will be explained later in more detail after discussing the charge identification for electrons.

The reconstruction path of the High Level Trigger follows the lines of the offline reconstruction as close as possible to not complicate analysis work with differently-behaving online and offline reconstruction efficiencies. The picture for photons is similar, except that tracks can come from converted photons only.

3. Electron Charge Identification

The rates of correct electron charge identification were measured in 2016 data. Three different charge estimates are inferred from the GSF track curvature, from the curvature of the closest KF track, and from $\Delta \phi$ between the cluster and the GSF track extrapolated to the vertex. The default charge estimation for electron candidates is taken as the majority vote of these three estimates. For $Z$ electrons which pass a loose cut-based selection, this gives misidentification rates at the $10^{-3}$ level in the barrel and around 2% in the endcaps (Figure 2a), increasing with electron energy due to the tracks being less curved. A very high rate of correct charge assignment
4. Electron And Photon Identification

Everything described so far was common to almost all CMS analyses. However, each analysis has its own efficiency and fake rate requirements. This is why two different ID algorithms are implemented for both electrons and photons. The first is a sequential cut-based selection with several working points for general use. The second algorithm relies on Boosted Decision Trees (BDTs) [14], particularly meant for maximum separation down to low transverse momentum ($p_T$). The BDTs are trained in several kinematic bins. An overview of the performance of the identification step is given in Figure 3. The ID variables can be grouped in shower-shape, track, track-cluster matching, conversion identification, and isolation variables.

While the shower-shape variables rely on the energy deposits in the ECAL and HCAL, the track variables rely on the GSF track and the nearest KF track. An interesting example for a track variables would be $f_{\text{brem}} = 1 - p_{\text{in}}/p_{\text{out}}$, where $p_{\text{in}}$ and $p_{\text{out}}$ denote the momentum estimate at the beginning and at the end of the reconstructed track. Hence, the $f_{\text{brem}}$ variable measures the momentum fraction lost by bremsstrahlung in the tracker. Apart from being a powerful ID variable, it is an excellent tool to access the material budget in data and compare it to MC. Data/MC discrepancies in this variable during Run 1 hinted a mismodeling of the material budget, which has been corrected for in Run 2. The precise extrapolation of the electron
trajectory to the calorimeter with the GSF tracking algorithm and the little material between
the tracker and the ECAL result in powerful track-cluster matching variables, especially cuts
on $\Delta\eta$ can be tight as no magnetic field spreads out the energy clusters in this direction. The
employed conversion ID variables are the goodness of a potential conversion vertex fits and
the number of missing hits at the beginning of the track, separating prompt electrons from
non-prompt conversion electrons.

On the isolation side, the sequential cut IDs use the energy flow in a $\Delta R = 0.3$ cone around
the object, corrected by an area–median pileup subtraction scheme. The BDT algorithm for
electrons, which uses isolation variables since 2017, instead takes the underlying neutral hadron,
charged hadron and electromagnetic energy flow components separately, plus a pileup estimator.
This results in better performance compared to taking only the pileup-corrected isolation sum
as isolation input. Complex classifiers like BDTs benefit from lower level input variables, which
they combine in a way which is adequate to meet their target. Including isolation in the BDTs
significantly improves the ID efficiency at highly efficient working points ($\gtrsim 85\%$) compared to
the traditional approach where an isolation cut is applied on top of the BDT identification. Next
to the new pixel detector, which reduces the fake rate in the endcaps at reconstruction level,
the new isolation-inclusive BDT selection is ensuring that the fake rate for electrons is at the
same level in 2017 as it was in the year before, despite the increased pileup. It should be noted
that isolation variables depend on the physics of the event, e.g. they might be less appropriate
for studying electrons in boosted systems. Therefore, a flavor of the BDT algorithm without
isolation variables is available for these analyses as well.

5. Selection Efficiencies in Data versus Simulation

Selection efficiencies can be measured with the so-called tag and probe method using $Z \rightarrow e^+e^-$
events both in data and simulation. In this method, probes are taken from events which are
tagged by a well identified electron selected by an invariant mass constraint ($65 < m_{ee} < 115$
GeV). Only little background enters in this mass window around the $Z$ peak, hence it results in

Figure 3: ROC curves for the electron multivariate identification (Boosted Decision Trees) and
the cut-based selection working points for $p_T > 20$ (a) and low $p_T$ (b) \[6\]. Signal is from Drell-
Yan+Jets Monte Carlo. Background are reconstructed electrons from Drell-Yan+Jets Monte
Carlo which do not match a generated electron within a cone of size $\Delta R = 0.1$. The signal
efficiencies are not corrected for data/MC scale factors, which affect more the BDT selection.
Figure 4: $Z \rightarrow e^+e^-$ mass distribution (a) and fitted $Z$ mass versus $Z\ p_T$ (b) where the (sub-) leading electron satisfies $p_T > 25(20)$ GeV, pass the medium cut-based identification and the pair is of opposite sign. The trigger requires a 23 and a 12 GeV online electron and the data has been re-reconstructed with improved ECAL detector calibrations. The electron ECAL energy has been corrected post reconstruction to ensure good data/MC agreement at the $Z$ peak. The peak is fitted with the convolution of a Breit-Wigner and a Crystal Ball function in the range 81 to 101 GeV to obtain the mass scale shift ($\Delta M$) and resolution ($\sigma_{CB}$) [6].

a sample of mostly real electrons on which selection efficiencies can be studied. The systematic effects considered are the tag selection, MC generator differences, the shape of the background fitting function as well as the signal shape (analytic fit or template from Monte Carlo). For photons, the $Z \rightarrow e^+e^-$ sample is used analogously, the only difference being that the electrons are reconstructed as photons without applying an electron veto based on tracker information. The ratio of the identification efficiencies in data over MC (often called data/MC scale factors) are typically around 0.95 in the barrel and 0.90 in the endcaps and increasingly deviate from one as the selection is tightened. For photons, the scale factors typically deviate less from one than in the electron case, because the selection does not rely as much on the new pixel detector which is not modeled optimally at the time of writing. The data/MC scale factors are another case where the revised ECAL calibrations for 2016 data improved data/MC agreement. This is reflected in almost all calorimetric ID input variables, in particular more complex ones like the electromagnetic energy flow isolation.

The DC-DC converters providing low voltage for the new pixel detector modules started failing on October 2017. This affected up to 10 % of the channels during data taking and raised concerns about efficiency losses in the reconstruction of electromagnetic objects. As the effect is not present in the simulation, the efficiency loss would also show up in the data/MC scale factors for the identification step. The scale factor deviation from unity indicates the maximum possible effect. The difference between the electron and photon scale factors gives the scale of the issue, since the photon identification relies almost completely on calorimetric information. From comparing periods with and without the DC-DC converter issue, the effect is estimated to be a few percent efficiency loss at the identification step. This is thought to be due to degraded tracker resolution affecting the ECAL-Tracker matching cuts.
6. $Z \rightarrow e^+e^-$ Invariant Mass with Full Energy Corrections

To put the reconstruction and identification to the test, one can extract the $Z$ peak from data and Monte Carlo simulation as was done in Figure 4a for the recalibrated 2016 reconstruction. The resolution, obtained by fitting a Breit-Wigner convoluted with a Crystal Ball function, is approximately 1.8 GeV, which is very comparable with Run 1. Data and MC is agreeing in Figure 4a by construction, as the mass scale correction in data and the resolution smearing in MC is calibrated exactly on the $Z$ peak. Therefore, the good data/MC ratio in Figure 4a serves more as a validation of the calibration procedure. However, the scale and smearing are not calculated differential in $Z$ boson $p_T$, so comparing the fitted $Z$ mass as a function of the $Z$ boson $p_T$ in data and simulation is an example of a benchmark where data and MC do not agree by construction. This is done in Figure 4b, which shows an excellent stability for the mass scale agreement versus the $p_T$ of the $Z$ boson.

7. Conclusion

The first CMS electron performance results for the initial reconstruction of data taken in 2017 were shown. The electron and photon performance in recalibrated 2016 data was discussed, which demonstrated the high-level benefits of a good understanding of low-level detector calibrations. Differences between data and simulation are reduced significantly by the recalibration. The agreement between data and simulation in 2017 is comparable to the previous year, despite challenging data taking conditions with a new pixel detector. Overall, the electron and photon performance in Run 2 is at the same level as in Run 1, thanks to advanced reconstruction algorithms which integrate well with the CMS particle-flow algorithm, excellent Monte Carlo simulations where the material budget of the inner detector is well modeled and sophisticated identification algorithms to ensure good background separation down to low transverse momentum.

References

[1] CMS Collaboration, The CMS Experiment at the CERN LHC, JINST 3, S08004 (2008). [doi:10.1088/1748-0221/3/08/S08004]
[2] CMS Collaboration, Measurements of Higgs boson properties in the diphoton decay channel in proton-proton collisions at $\sqrt{s} = 13$ TeV, CERN-EP-2018-060, [arXiv:1804.02716]
[3] CMS Collaboration, Measurements of properties of the Higgs boson in the four-lepton final state at $\sqrt{s} = 13$ TeV, CMS-PAS-HIG-18-001.
[4] CMS Collaboration, Search for high mass resonances in dielectron final state, CMS-PAS-EXO-18-006.
[5] A. Zghiche, Performance of the CMS electromagnetic calorimeter during the LHC Run II, these proceedings.
[6] CMS Collaboration, Electron and Photon performance in CMS with the full 2017 data sample and additional 2016 highlights for the CALOR 2018 Conference, CERN-CLS-18-017.
[7] CMS Collaboration, The CMS electromagnetic calorimeter project, CERN-LHCC-97-033.
[8] CMS Collaboration, CMS Technical Design Report for the Pixel Detector Upgrade, CERN-LHCC-2012-016, [doi:10.2172/1151650]
[9] CMS Collaboration, Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV, [doi:10.1088/1748-0221/10/06/P06005] [arXiv:1502.02701]
[10] CMS Collaboration, Performance of Photon Reconstruction and Identification with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV, JINST 110, P08010 (2015), [doi:10.1088/1748-0221/10/08/P08010] [arXiv:1502.02702]
[11] W. Adam et al., Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC, CERN-CMS-NOTE-2005-001.
[12] S. Baffioni et al., Electron reconstruction in CMS, Eur. Phys. J. C49, 1099–1116 (2007). [doi:10.1140/epjc/s10052-006-0175-5]
[13] CMS Collaboration, Particle-flow reconstruction and global event description with the CMS detector, JINST 12, P10003 (2017). [doi:10.1088/1748-0221/12/10/P10003] [arXiv:1706.04965]
[14] Tianqi Chen, Carlos Guestrin, XGBoost: A Scalable Tree Boosting System, ArXiv e-prints (2016), [doi:10.1145/2939672.2939785] [arXiv:1603.02754]