

Higgs to 4 lepton mass spectrum using CMS Open Data

N. Z. Jomhari, A. Geiser

DPG T 15.7
DESY Hamburg

March 19, 2018

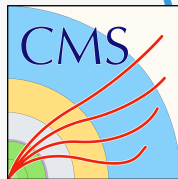
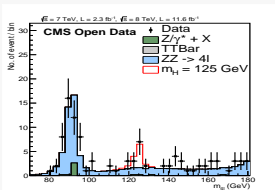
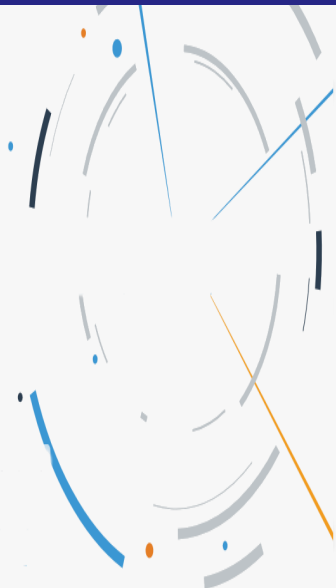


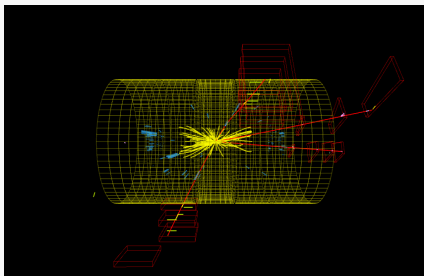
Table of content

- Introduction
- How do CMS detect particles?
- **CMS Open Data overview**
- **Higgs to 4 lepton mass spectrum**
- Motivation of Higgs example
- The challenge
- Original datasets
- Object & event selection
- **How can YOU reproduce Higgs?**
- Conclusion



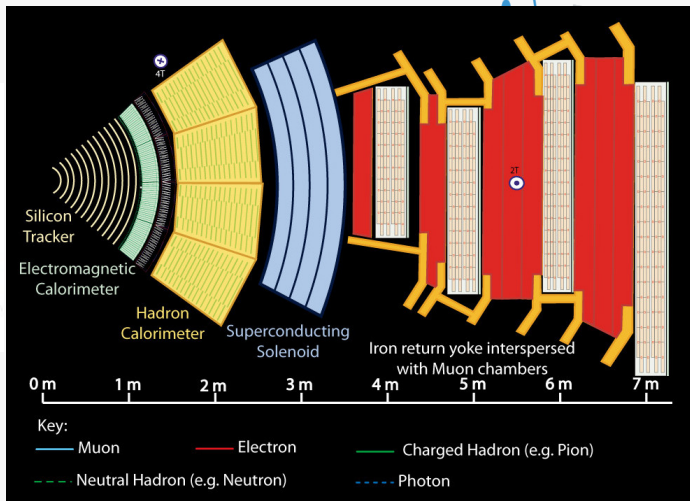
Introduction

- In 2012, Higgs boson; the last piece of Standard Model was discovered at CERN
- Its discovery was crucial because Higgs mechanism explained how W and Z boson acquire mass
- It can decay to 4 charge lepton through intermediate Z bosons and this decay channel is the cleanest experimentally



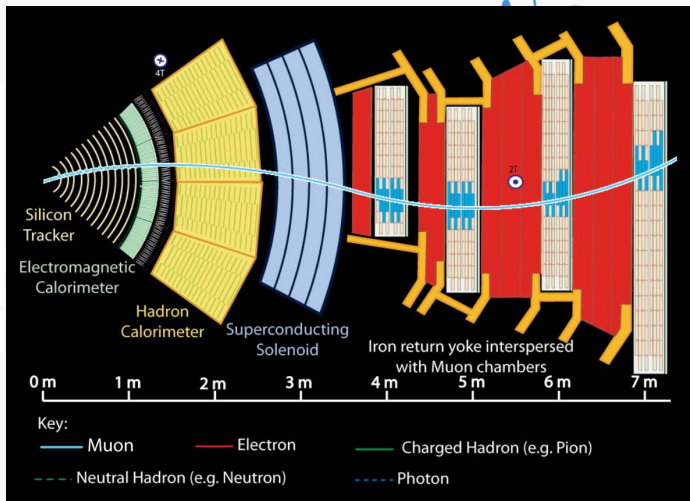
- Compact Muon Solenoid (CMS) experiment is one of the experiment that confirmed the existence of Higgs boson

How do CMS detect particles?



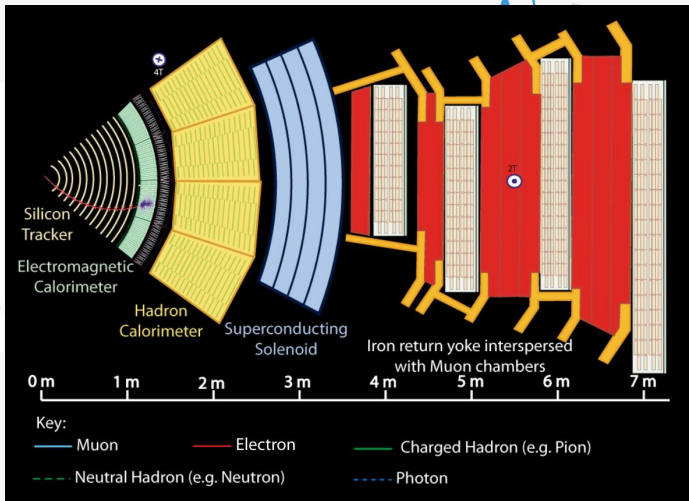
Transverse view of CMS detector

How do CMS detect particles?



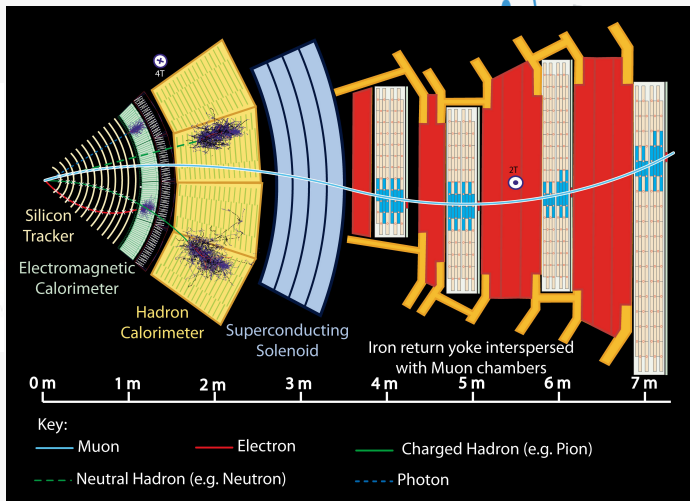
Transverse view of CMS detector

How do CMS detect particles?



Transverse view of CMS detector

How do CMS detect particles?



Transverse view of CMS detector

CMS Open Data overview

- What is CMS Open Data?
 - CMS Open Data are original data used by CMS members that are released to the public via CERN Open Data portal: <http://opendata.cern.ch/docs/about-cms> (see also: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/CMSPublicData>)
 - CMS made public around half of the Run1 data collected by the CMS detector at the Large Hadron Collider (LHC)
- Purpose?
 - **Research application**: to encourage people to **conduct and publish their own analysis** using original preserved data
 - **Educational application**: available to school pupils and university students for outreach and educational purposes using simplified data
- This talk focus on **research application**

CERN Open Data portal

opendata
CERN

About ▾

Explore more than **1 petabyte**
of open data from particle physics!

Start typing...

Search

search examples: [collision datasets](#), [keywords:education](#), [energy:7TeV](#)

Explore

[datasets](#)
[software](#)
[environments](#)
[documentation](#)

Focus on

[ATLAS](#)
[ALICE](#)
[CMS](#)
[LHCb](#)

▽ Get started ▾

CERN Open Data homepage:

<http://opendata.cern.ch/>

CERN Open Data portal: CMS Experiment

Learn

Discover the world of open data from particle physics

Welcome to our updated portal
CMS Guide to education use of CMS Open Data

Improving educational content with high school teachers: A field report from our summer students

Glossary

more

Visualise

Explore detector events and run basic histogramming

CMS Event Display
CMS Histograms

News

Analyse

Run your own physics analyses, start virtual machines

CMS Guide to research use of CMS Open Data

ATLAS Higgs Machine Learning Challenge

Getting Started with LHCb Open Data

Getting Started with ALICE Open Data

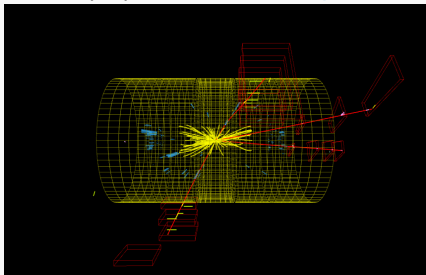
more

CERN Open Data homepage:

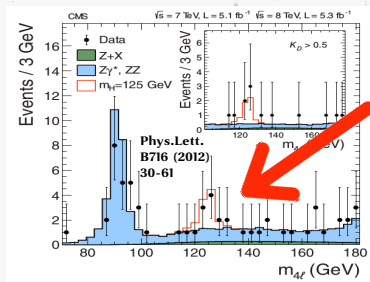
<http://opendata.cern.ch/>

Higgs to 4 lepton mass spectrum

- Higgs to 4 lepton is one of the several examples available in CMS Open Data. This example is strongly simplified analysis but similar to the one in the published paper
- Motivation:
 - To **approximately reproduce** $H \rightarrow ZZ \rightarrow 4l$ ($4e$, 4μ & $2\mu 2e$) mass spectrum using opendata source and compare with published paper [Phys. Lett. B716 \(2012\) 30-61](#)
 - To **provide example code** for Higgs search in CMS for educational and research purposes



CMS event display of Higgs $\rightarrow 4\mu$ event



Higgs $\rightarrow 4l$ mass spectrum

Higgs

The challenge

- Here is the list of challenges we met in order to setup this Higgs example:
 - How to pick up the right objects in the data and their documentation
 - How to know if there are additional selections, corrections etc.
 - Takes long time to complete a full analysis

Original dataset and Monte Carlo (MC) samples, 2012 data

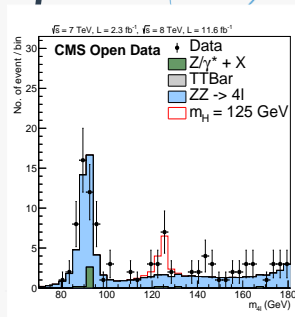
- Here are the lists of original dataset and MC samples in Analysis Object Data (AOD) format, same as used by CMS physicists

- **For data:** Muon and electron dataset ($L = 11.6 \text{ fb}^{-1}$)

- /DoubleMuParked (for 4μ , $2\mu 2e$)
- /DoubleElectron (for $4e$)

- **For MC:** To generate the background distribution

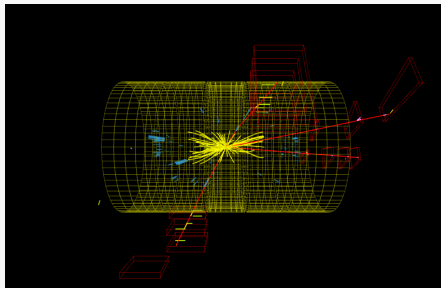
- /ZZTo4mu_8TeV-powheg-pythia6
- /ZZTo4e_8TeV-powheg-pythia6
- /ZZTo2e2mu_8TeV-powheg-pythia6
- /SMHiggsToZZTo4L_M-125_8TeV-powheg15-JHUGenV3-pythia6
- /DYjetsToLL_M-50_TuneZ2Star_8TeV-madgraph-tarball-tauola-tauPolarOff
- /DYjetsToLL_M-10to50_HT-200to400_TuneZ2star_8TeV-madgraph-tauola
- /DYjetsToLL_M-10to50_HT-400toInf_TuneZ2star_8TeV-madgraph-tauola
- /TTbar_8TeV-Madspin_aMCatNLO-herwig



Similar for 2011 data

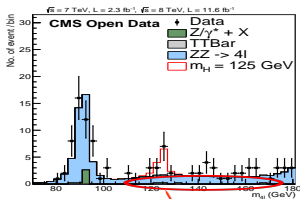
Object and Event Selection

- Muons and electrons selection
 - Muons and electrons are isolated
 - Transverse momentum, $p_T^\mu > 5 \text{ GeV}$, $p_T^e > 7 \text{ GeV}$
 - Geometrical acceptance, $|\eta_\mu| < 2.4$, $|\eta_e| < 2.5$
- Muons and electrons come from the same vertex
- Ensure the total electrical charge and lepton number of $4l$ is 0

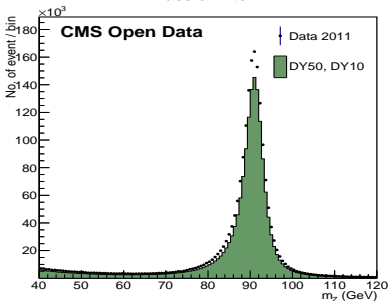


Data driven background normalization

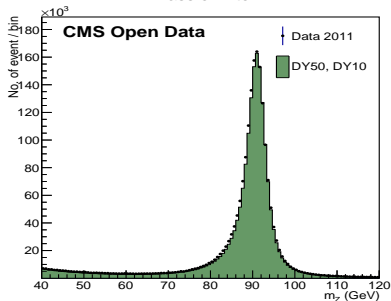
The MC is normalized to data



mass of Z to 2l



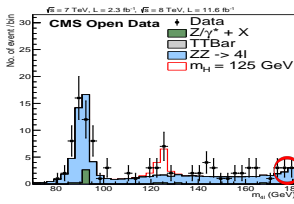
mass of Z to 2l



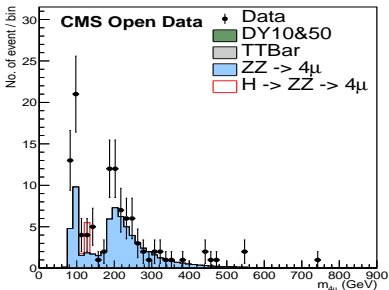
(From left) Mass of $Z \rightarrow 2l$ before and after the normalization

Data driven background normalization

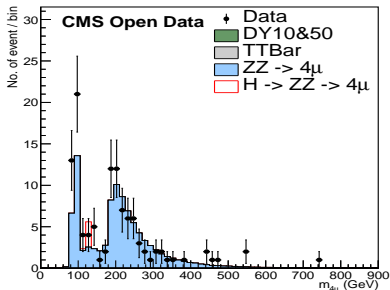
The MC is normalized to data



mass of 4μ 2011 & 2012

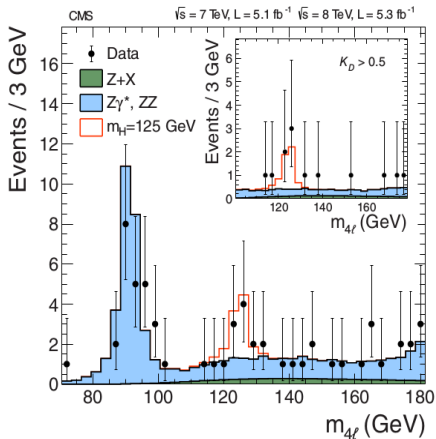


mass of 4μ 2011 & 2012

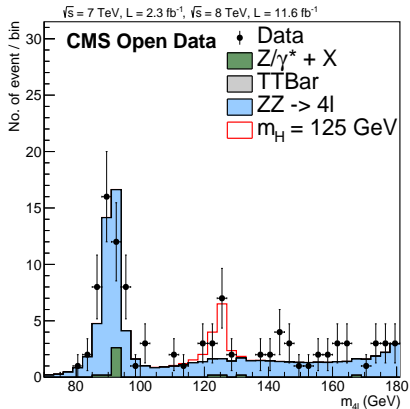


(From left) Mass of 4μ before and after the normalization

4 lepton mass spectrum



Mass of $4l$ from published paper:
[Phys. Lett. B716 \(2012\) 30-61](#)



Mass of $4l$ from CMS Open Data. The Higgs contribution is normalized to the data in the signal region

How can YOU reproduce Higgs mass spectrum?

Step 1

CMS Guide to research use of CMS Open Data

Documentation Guide

If you are interested in step-by-step instructions to start working with CMS Open Data, please consult these pages:

- [Install Virtual Machine](#)
- [Get started with CMS Open Data](#)

← Install Virtual Machine

**↖ Get started with
CMS Open Data**

How can YOU reproduce Higgs mass spectrum?

Step 2

open**data**
CERN

Search

About ▾

Software x Analysis x
CMS x

Filter by type

- Dataset 878
- Documentation 37
- Environment 13
- News 5
- Software 25
 - Analysis 10
 - Framework 3
 - Tool 7
 - Validation 5
- Supplementaries 2642

Filter by experiment

- ALICE 4
- ATLAS 1
- CMS 10
- LHCb 1

Filter by year

- 2010 4
- 2011 4
- 2011-2012 2

Sort by: Best match ▾ asc. ▾ Display: detailed ▾ 20 results ▾

Found 10 results. < 1 >

Analysis code for extracting the trigger information from the CMS 2010 data

This is an example of the C++ code (and its configuration in Python) needed to extract the trigger information from CMS Open/Legacy data. These instructions are valid to work with 2010 data. Snipp...

[Software](#) [Analysis](#) [CMS](#)

Higgs-to-four-lepton analysis example using 2011-2012 data

This research level example is a strongly simplified reimplementaion of parts of the original CMS Higgs to four lepton analysis published in ...

[Software](#) [Analysis](#) [CMS](#)

Two-lepton/four-lepton analysis example of CMS 2011 open data

Example code for a simple analysis on Z decays to two leptons and ZZ decays to four leptons, of CMS open data...

[Software](#) [Analysis](#) [CMS](#)

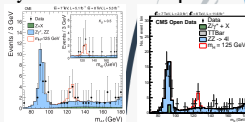
Higgs example

How can YOU reproduce Higgs mass spectrum?

- There are **four levels** of increasing complexity for this example:

(1) **Compare** the provided final output plot

- Takes \sim seconds

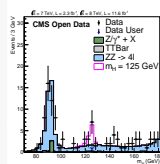


(2) **Reproduce** the final output plot from the predefined histogram files using a root macro

- Takes \sim few minutes to \sim hours depending on setup and proficiency

(3) **Produce** a root data input file from original data and MC files for one Higgs signal candidate

- Takes \sim 1 hour if Virtual Machine is already installed



(4) **Reproduce** the full example analysis

- takes $>$ 1 months on single CPU with fast internet connection

Conclusion

- CMS has **publicly** released 2010, 2011 and 2012 datasets and their corresponding MC, \sim half of the respective Run1 datasets
- **ANYONE** (public, external and internal researchers) can use CMS Open Data. e.g. reconstruct part of Higgs discovery!
- For **research purposes**, several analysis and validation examples are provided. (For **educational purposes**, there are a lot of fun activities too!)
- The provided higgs example is a simplified analysis. It was meant as a guide for people to **start conduct and publish their own analysis**
- **Feedback** from users are always welcome :)

opendata.support@cern.ch

The end

backup



CMS Open Data additional information

- CMS Open Data for research: AOD format (CMS root)
 - 1st release: 28 TB of reconstructed 2010 7 TeV pp collision data in Nov. 2014
 - 2nd release: 130 TB of 2011 7 TeV pp collision data and >200 TB of corresponding MC data in April 2016
 - 3rd release: 8 TeV pp data + MC (~ 2 PB) in Dec 2017
- CMS data preservation, re-use and open access policy:
<http://opendata.cern.ch/record/411>

Original dataset and Monte Carlo samples for 2011

- Here are the list of original dataset and MC samples in Analysis Object Data (AOD) format, same as used by CMS physicists
- **For data:** */Run2011A-12Oct2013-v1/AOD ($L = 2.3 \text{ fb}^{-1}$)
 - /DoubleMu (for 4μ , $2\mu 2e$)
 - /DoubleElectron (for $4e$)
- **For MC:** */Summer11LegDR-PU_S13_START53_LV6-v1/AODSIM
 - /ZZTo4mu_mll4_7TeV-powheg-pythia6
 - /ZZTo4e_mll4_7TeV-powheg-pythia6
 - /ZZTo2e2mu_mll4_7TeV-powheg-pythia6
 - /SMHiggsToZZTo4L_M-125_7TeV-powheg15-JHUgenV3-pythia6
 - /DYJetsToLL_M-50_7TeV-madgraph-pythia6-tauola
 - /DYJetsToLL_M-10To50_TuneZ2_7TeV-pythia6
 - /TTTo2L2Nu2B_7TeV-powheg-pythia6

Object and event selection

- Muon and electron selection
 - Relative isolation, $\text{reIso}_{\mu,e} < 0.4$ with $\Delta R = 0.4$
 - Transverse momentum, $p_T^\mu > 5 \text{ GeV}$, $p_T^e > 7 \text{ GeV}$
 - Pseudorapidity, $|\eta_\mu| < 2.4$, $|\eta_e| < 2.5$
- Transverse impact parameter w.r.t. primary vertex, $|d_{xy}| < 0.5 \text{ cm}$
- Longitudinal impact parameter w.r.t. primary vertex, $|dz| < 1 \text{ cm}$
- 3D impact parameter significance, $|SIP_{3D}| < 4$
- No. of $\mu, e \geq 4$ for case 4μ and $4e$
- No. of $\mu, e \geq 2$ for case $2\mu 2e$
- For Z_a^1 , $p_T^{l_1} > 20 \text{ GeV}$ and $p_T^{l_2} > 10 \text{ GeV}$
- $40 < m_{z_a} < 120 \text{ GeV}$
- $12 < m_{z_b} < 120 \text{ GeV}$
- $m_{4l} > 70 \text{ GeV}$
- Total charge and lepton number of $4l$ is neutral

¹ Z_a = The lepton pair that closest to Z mass