

PhD Defense

V. Khristenko¹

¹Department of Physics, The University of Iowa

PhD Defense, 2017

Outline

1 Search for Standard Model Higgs Boson with CMS

2 Calibration of CMS Hadron Forward Calorimeter

3 Simulations of Modern Calorimeter Systems

Where we stand

- 5 years at The University of Iowa as PhD Student
 - Test Beams at Fermilab and CERN
- 3 years at CERN working on Compact Muon Solenoid Experiment
 - CMS 2015 Achievement Award
 - HCAL Operations Deputy Coordinator
 - HCAL DQM Group Lead.
- Received CERN IT Fellowship to work as Big Data Engineer.

Outline

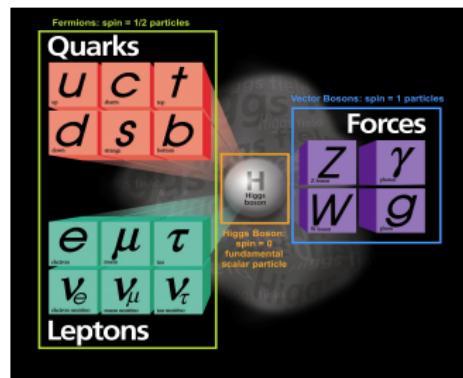
1 Search for Standard Model Higgs Boson with CMS

2 Calibration of CMS Hadron Forward Calorimeter

3 Simulations of Modern Calorimeter Systems

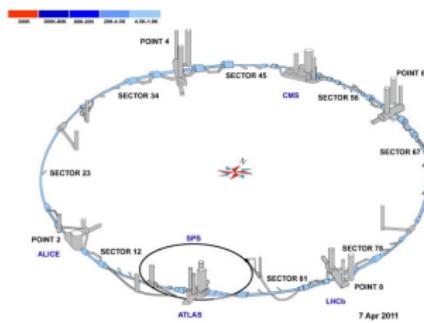
Introduction

- Standard Model - Gauge Quantum Field Theory
- Global Gauge Invariance is assumed - common symmetries (e.g. translation, rotation)
- Local Gauge Invariance
 $SU(3) \times SU(2) \times U(1)$ for internal symmetries
 - Massless Bosons
 - From experiment $W^{+/-}/Z$ are known to be massive
- Higgs Mechanism and Spontaneous Symmetry Breaking
 - Break the Local Gauge Invariance by selecting a particular ground state (minimum of potential energy)
 - Generates mass terms for W, Z
 - Generates mass terms for fermions thru Yukawa coupling to Higgs Field



Large Hadron Collider

- LHC - Particle Accelerating Complex
- Accelerates/Collides pp, Pb-p, Pb-Pb
- Current maximum momentum per beam is 6.5TeV/c ($\sqrt{s} = 13\text{TeV}$) for proton beam
- “Data Factory” one of its kind
 - Even without recording collisions



Compact Muon Solenoid

CMS DETECTOR

Total weight : 14,000 tonnes
Overall diameter : 15.0 m
Overall length : 28.7 m
Magnetic field : 3.8 T

STEEL RETURN YOKE
12,500 tonnes

SILICON TRACKERS
Pixel (100x150 μm) $\sim 16\text{m}^2 \sim 66\text{M}$ channels
Microstrips (80x180 μm) $\sim 200\text{m}^2 \sim 9.6\text{M}$ channels

SUPERCONDUCTING SOLENOID
Niobium titanium coil carrying $\sim 18,000\text{A}$

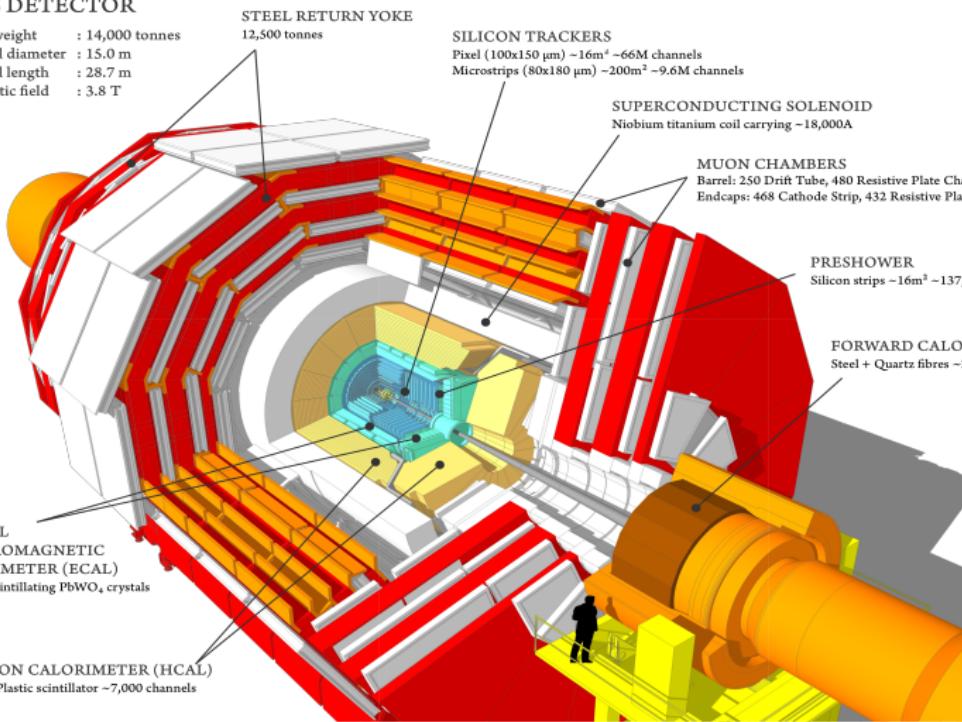
MUON CHAMBERS
Barrel: 250 Drift Tube, 480 Resistive Plate Chambers
Endcaps: 468 Cathode Strip, 432 Resistive Plate Chambers

PRESHOWER
Silicon strips $\sim 16\text{m}^2 \sim 137,000$ channels

FORWARD CALORIMETER
Steel + Quartz fibres $\sim 2,000$ Channels

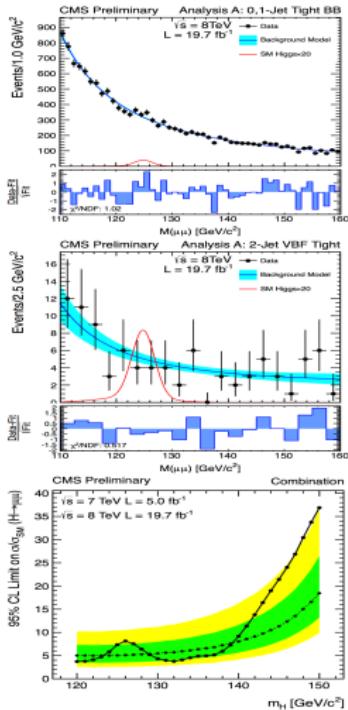
CRYSTAL
ELECTROMAGNETIC
CALORIMETER (ECAL)
 $\sim 76,000$ scintillating PbWO_4 crystals

HADRON CALORIMETER (HCAL)
Brass + Plastic scintillator $\sim 7,000$ channels



SM Higgs Search at CMS in Run I

- Run I proton-proton collisions at $\sqrt{s} = 7 \text{ TeV}$ ($L = 5 \text{ fb}^{-1}$) and $\sqrt{s} = 8 \text{ TeV}$ ($L = 19.7 \text{ fb}^{-1}$)
- CMS made observations of Higgs Boson
 - $H \rightarrow \gamma\gamma$
 - $H \rightarrow ZZ \rightarrow 4l$
 - $H \rightarrow \tau\tau$
 - no observation for $H \rightarrow \mu\mu$
- 1.1 σ excess in significance for mass 125 GeV was observed - compatible with statistical fluctuation
- 95% Confidence Level (CL) exclusion limits were set on 125 GeV Higgs Production Cross-Section.
 - for 7 TeV $\rightarrow \sigma = 12.8 \times \text{SM(Exp.)}$ and $\sigma = 19.0 \times \text{SM(Obs.)}$
 - for 8 TeV $\rightarrow \sigma = 5.6 \times \text{SM(Exp.)}$ and $\sigma = 6.9 \times \text{SM(Obs.)}$



Problem Statement

Remark

The primary objective of this work is to investigate the $\mu^+ \mu^-$ Higgs decay mode with 35.9 fb^{-1} of data collected in proton-proton collisions at $\sqrt{s} = 13 \text{ TeV}$ by Compact Muon Solenoid (CMS) Experiment during the second operational phase of LHC (2016)

Analysis Workflow

- Collect the data!
- Data Preprocesssing/Skimming
 - Apply basic selection criteria (e.g. at least 2 muons in event)
 - Reduce the amount of data by orders of magnitude.
- Event Selections
 - Apply muon corrections
 - Apply various selections
 - Choose objects of interest
 - Compare Data with MC simulations and establish the validity of simulated models.
 - If matches - good, does not - why?! Other models... ?
- Event Categorization
 - Grouping events by #jets, $p_t^{\mu\mu}$, BDT score, ...
- Compute Systematics
- Statistical Analysis
 - Build Signal/Background Models
 - Extract the 95% CL Exclusion Limits

Datasets and Configuration: General

- Data

- Use 36.9 fb^{-1} of data collected over 2016
- SingleMuon Primary Dataset

- Monte Carlo

- Higgs Signal
 - VBF, ggFusion, $W^{+/-}H$, Zh production processes
- Backgrounds
 - Drell-Yan + Jets
 - $t\bar{t}$ + Jets
 - ...
- PileUp reweighting for 69.2 mb

Datasets and Configuration: Data

Table: Datasets for proton-proton collisions recorded at $\sqrt{s} = 13$ TeV by CMS at LHC in 2016.

Datasets	Int. Luminosity (fb^{-1})
/SingleMuon/Run2016B-03Feb2017_ver2-v2/MINIAOD	5.788
/SingleMuon/Run2016C-03Feb2017-v1/MINIAOD	2.573
/SingleMuon/Run2016D-03Feb2017-v1/MINIAOD	4.248
/SingleMuon/Run2016E-03Feb2017-v1/MINIAOD	4.009
/SingleMuon/Run2016F-03Feb2017-v1/MINIAOD	3.102
/SingleMuon/Run2016G-03Feb2017-v1/MINIAOD	7.540
/SingleMuon/Run2016H-03Feb2017_ver2(3)-v1/MINIAOD	8.606

Datasets and Configuration: SM Higgs Signal

Table: Standard Model 125 GeV Higgs Boson Signal Datasets for 13 TeV. Dataset names for 120/130 GeV are omitted for brevity. Moriond 2017 conditions are used (omitted the conditions specification for brevity).

Datasets	σ (pb)
/GluGlu_HToMuMu_M125_13TeV_powheg_pythia8	48.58
/VBF_HToMuMu_M125_13TeV_powheg_pythia8	3.782
/WMinusH_HToMuMu_M125_13TeV_powheg_pythia8	0.5331
/WPlusH_HToMuMu_M125_13TeV_powheg_pythia8	0.851
/ZH_HToMuMu_M125_13TeV_powheg_pythia8	0.8839

Datasets and Configuration: Backgrounds

Table: Background Datasets. Moriond 2017 conditions have been used (omitted the conditions specification for brevity).

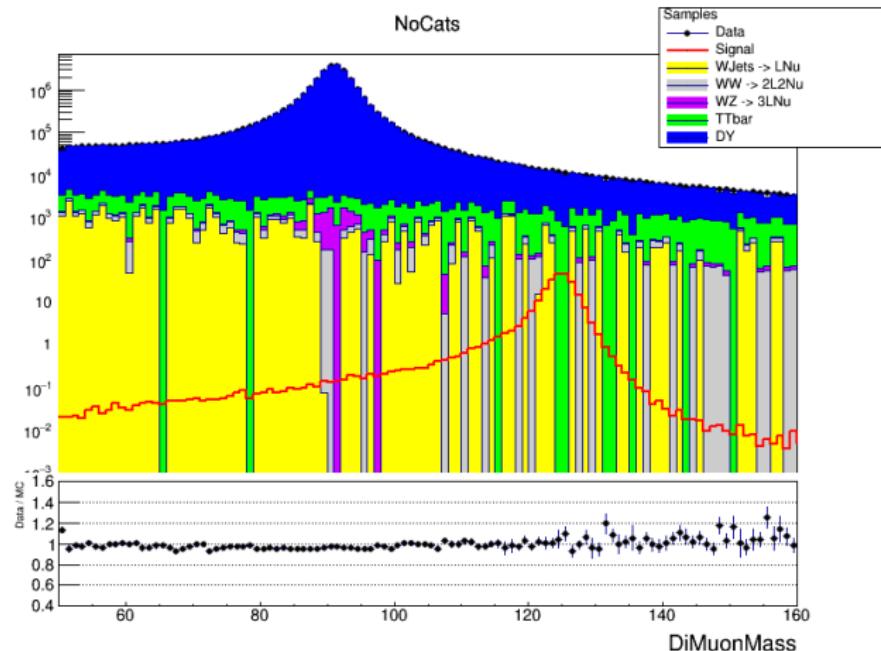
Dataset	σ (pb)
/DYJetsToLL_M-50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	5765
/ST_tW_top_5f_NoFullyHadronicDecays_13TeV-powheg_TuneCUETP8M1	35.85
/TTJets_DiLept_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	85.656
/WJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	61526.7
/WWTo2L2Nu_13TeV-powheg-herwigpp	10.481
/WZTo3LNu_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	4.712

Event Selections

- Primary Vertex
 - at least 1 Primary Vertex with
 - $|z_{PV}| < 24$ - shift along the beam axis w.r.t. nominal CMS center
 - $ndf > 4$ - degrees of freedom
- HLT - HLT_IsoMu24 or HLT_IsoTkMu24 to fire
 - Event must have at least 1 μ with $p_t \geq 24\text{GeV}$
- Jets
 - $p_t > 30\text{ GeV} \&\& |\eta| < 4.7 \&\& \Delta R_\mu < 0.4$
- Muons
 - 2 oppositely charged muons
 - Each Muon
 - Rochester Muon Corrections
 - Medium Muon Id
 - $p_t > 10\text{ GeV} \&\& |\eta| < 2.4 \&\& I_{rel}^{PF} < 0.25$
 - At least 1 muon to match HLT
 - $p_t > 26\text{ GeV} \&\& |\eta| < 2.4 \&\& \Delta R_{HLT} < 0.1$
- Passed basic event selections - identify this as the **TopCategory** - combination of all.

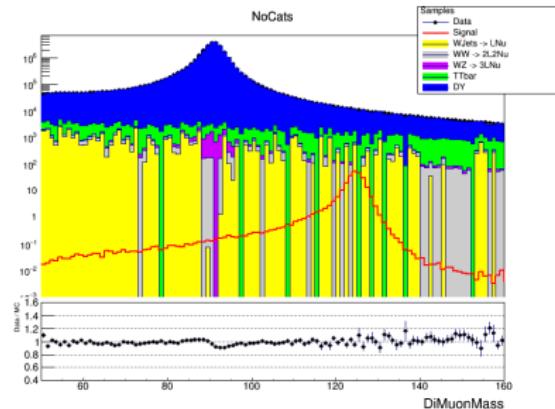
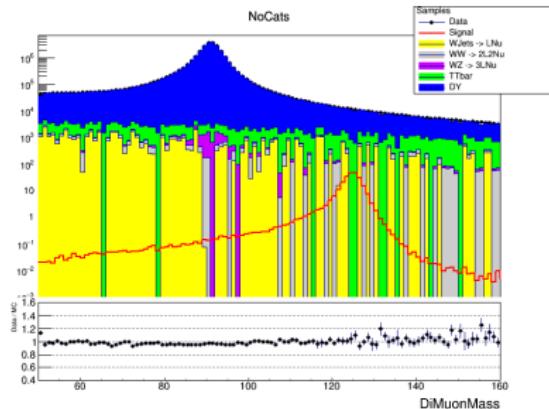
Top Category: Mass Distribution

Dimuon Mass ($m_{\mu\mu}$) distribution with Rochester Muon Corrections:

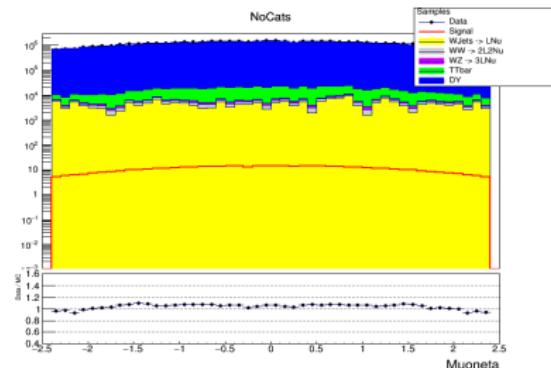
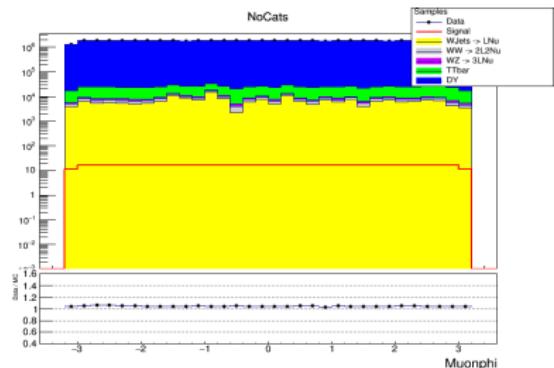
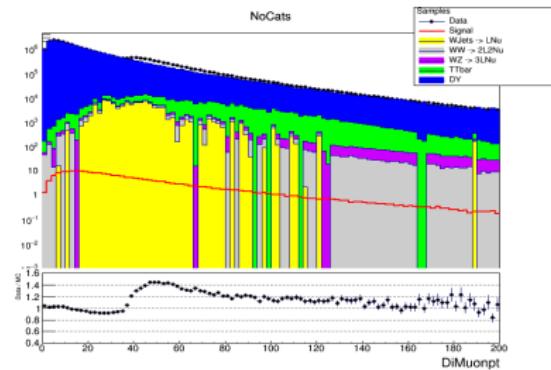
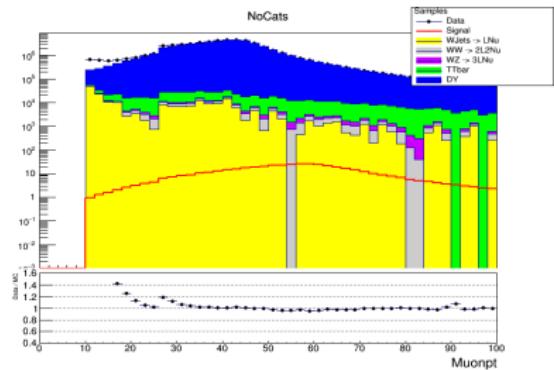


Top Category: Corrected vs Uncorrected

Corrected (Rochester) vs Uncorrected:



Top Category: Kinematic Distributions



Event Categorization

- Grouping events using certain criteria
- 2 different procedures
- **Baseline Categorization** is based on Run I categorization selections
- **Greedy Categorization**: 2-step procedure
 - Binary Classification: Engineer a new feature, BDT score, using Boosted Decision Tree (BDT)
 - Greedy Event Grouping: Use 2 features ($\max(\eta_{\mu_1}, \eta_{\mu_2})$ and BDT score) to split events into different categories.

Event Categorization: Baseline

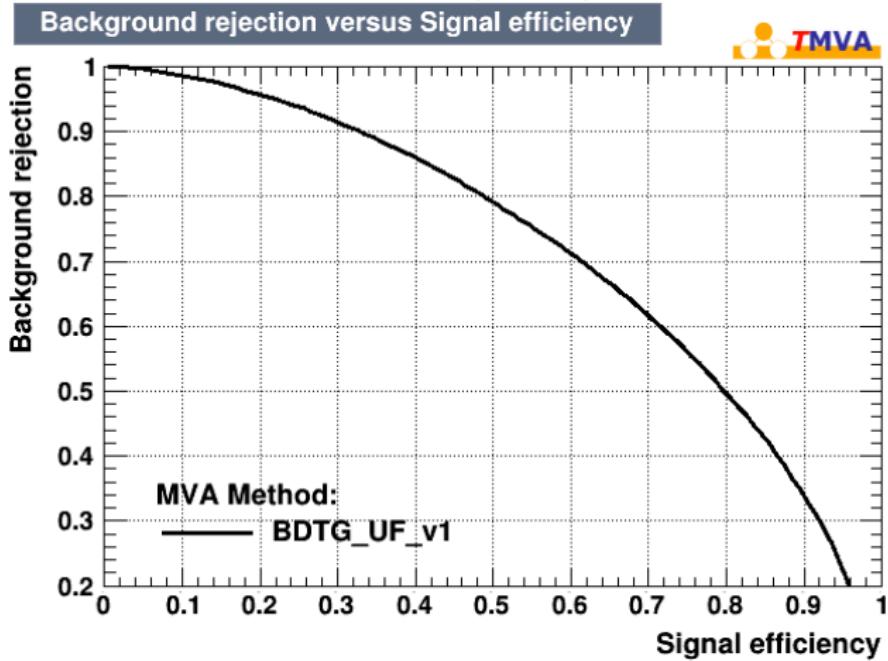
- Require only $m_{\mu\mu}$ in [110, 160]
- **2Jet Category**
 - # Jets ≥ 2
 - $p_t^{j_1} > 40 \text{ GeV} \text{ && } p_t^{j_2} > 30 \text{ GeV} \text{ && } p_t^{\text{MET}} < 40 \text{ GeV}$
 - if not satisfied, go to **01Jet** Category part
 - passed, then if $m_{jj} > 650 \text{ GeV} \text{ && } |d\eta| > 3.5 \rightarrow \text{VBFTight}$ Category
 - else if $m_{dijet} > 250 \text{ GeV} \text{ && } p_t^{\mu\mu} > 50 \text{ GeV} \rightarrow \text{ggFTight}$ Category
 - else $\rightarrow \text{ggFLoose}$ Category
- **01Jet Category**
 - # Jets ≤ 1
 - if $p_t^{\mu\mu} \geq 25 \text{ GeV} \rightarrow \text{01JetsTight} \rightarrow \text{Geometry Categorization}$
 - else $\rightarrow \text{01JetsLoose} \rightarrow \text{Geometry Categorization}$

Event Categorization: Greedy Step1 - Binary Classification

- Require only $m_{\mu\mu}$ in [110, 160] GeV mass range
- Train a Boosted Decision Tree to perform a binary classification: signal / background.
- Training / Cross-Validation / Testing using half of Signal and all of Background events
- BDT Features are various event and kinematic variables: $p_t^{\mu\mu}$, $\eta_{\mu\mu}$, # jets, ...
 - No mass dependence among the BDT features!
- BDT score (continuous value [-1, 1]) is preserved - do not actually select the "turning point" (for binary classification).

Event Categorization: Greedy Step1 - BDT ROC

BDT Receiver Operating Curve:



Event Categorization: Greedy Step2 - Event Splitting

Define a Significance, S_c , for a histogram:

$$S_c^2 = \sum_{c,i} N_{c,i}^{S2} / N_{c,i}^B \quad (1)$$

Use 2 features: $\max(\eta_{\mu_1}, \eta_{\mu_2})$ and BDT score

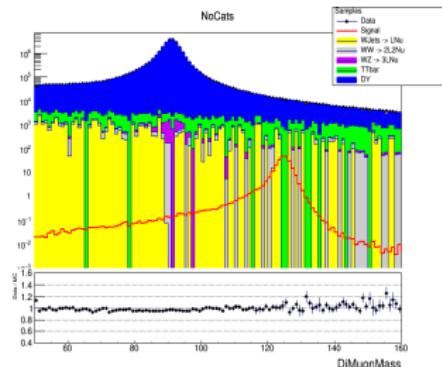
Define a Gain, G, of a split:

$$G = S_{left}^2 + S_{right}^2 - S_{node}^2 \quad (2)$$

- For a given histogram, greedily scan through the features
- Pick the split that maximizes Gain
- Proceed recursively.

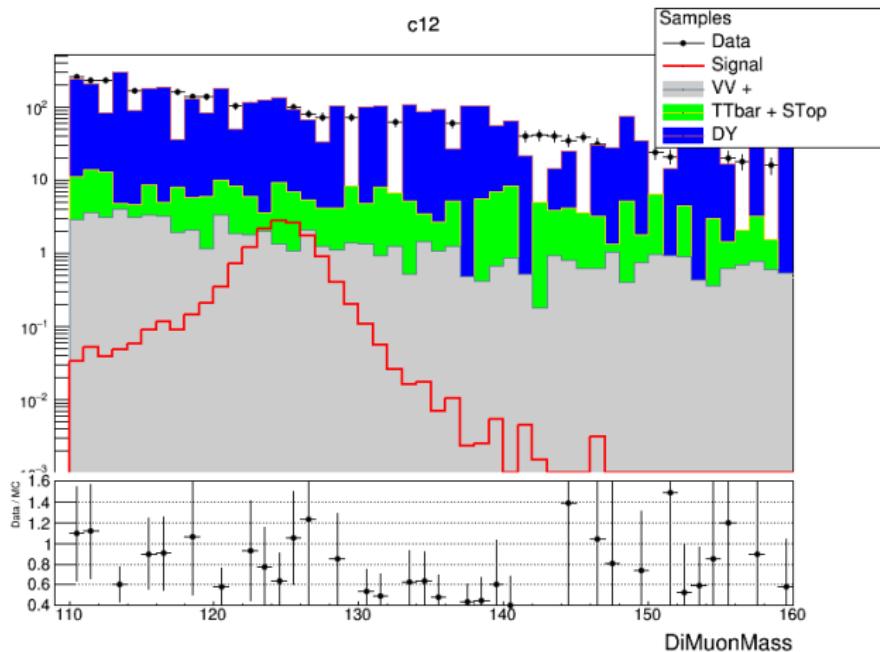
The result is 13 categories (15 for Baseline) that are obtained via a greedy procedure!

Top Category Mass Distribution:

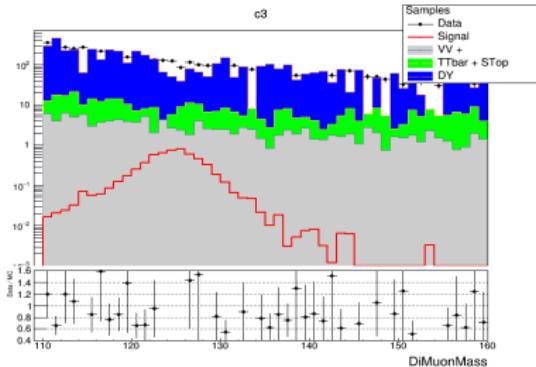
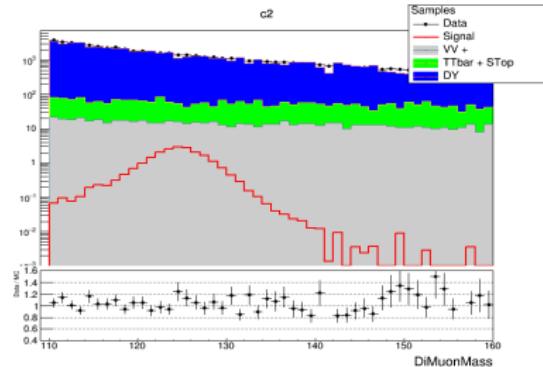
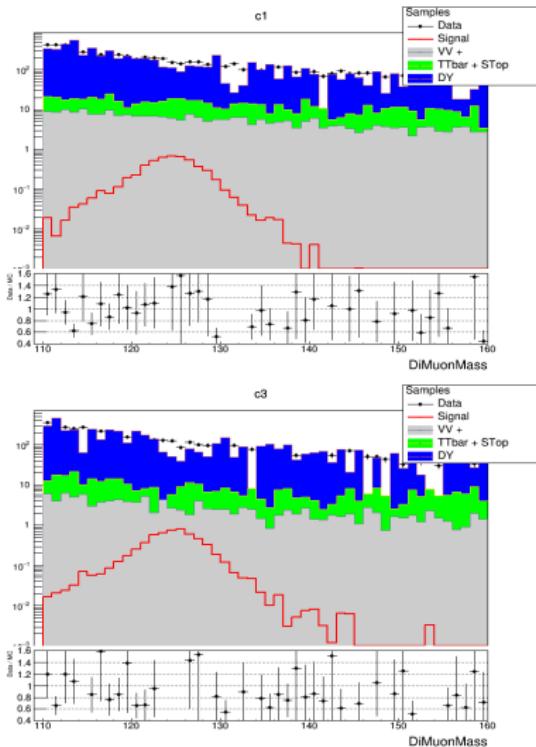
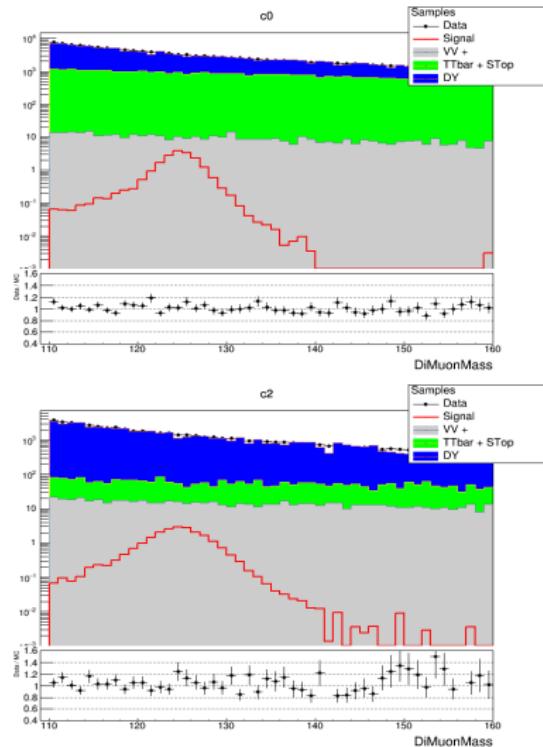


Event Categorization: Greedy - Mass Distributions

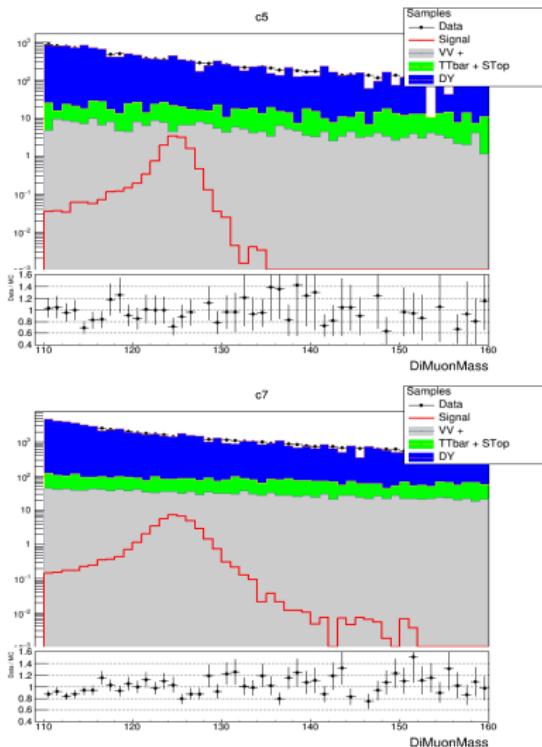
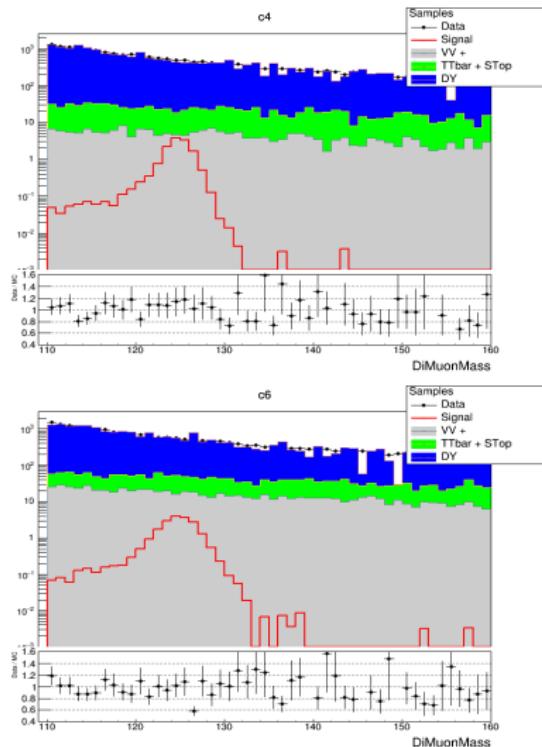
Most Sensitive Category "c12":



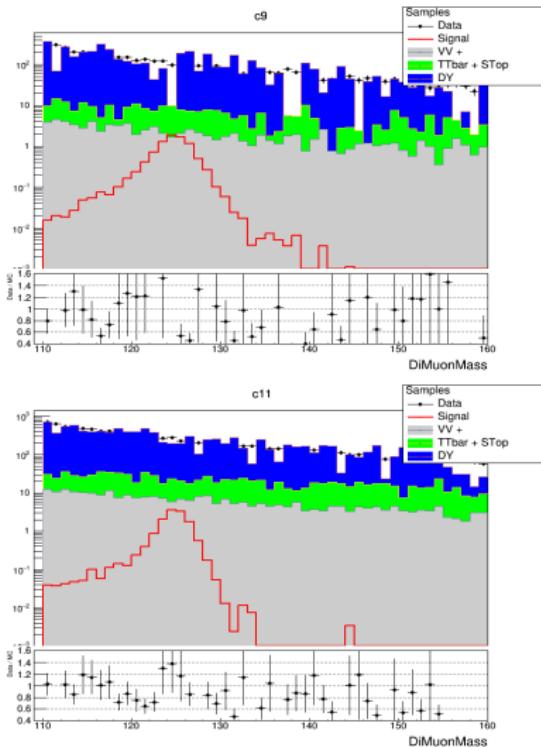
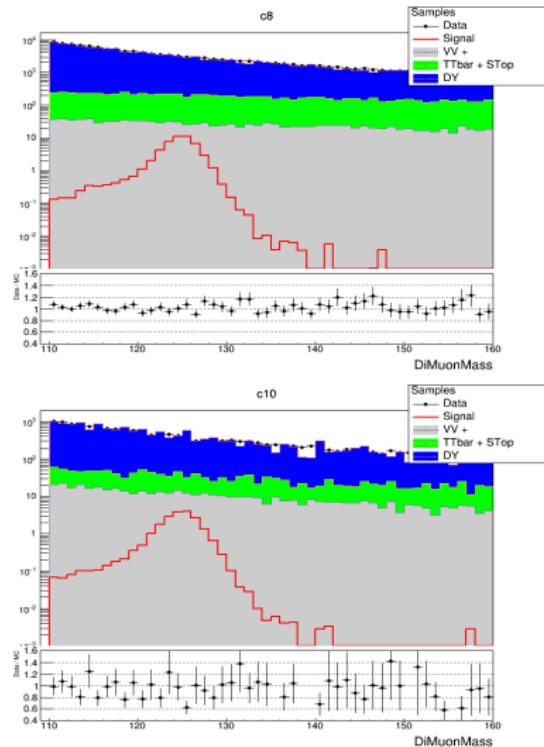
Event Categorization: Greedy - Mass Distributions



Event Categorization: Greedy - Mass Distributions



Event Categorization: Greedy - Mass Distributions



Signal Modeling

Background Modeling

Systematic Uncertainties

Conclusions

Outline

1 Search for Standard Model Higgs Boson with CMS

2 Calibration of CMS Hadron Forward Calorimeter

3 Simulations of Modern Calorimeter Systems

Hadron Forward

CMS Calorimeter

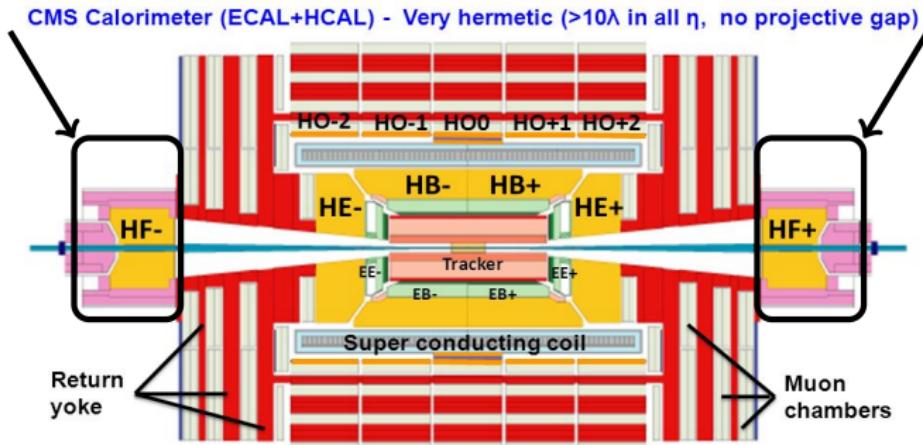
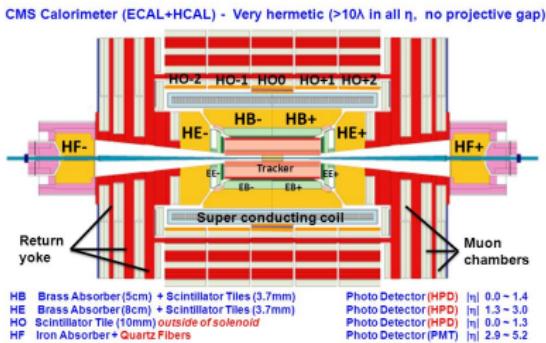


Photo Detector (HPD)	$ \eta $	0.0 ~ 1.4
Photo Detector (HPD)	$ \eta $	1.3 ~ 3.0
Photo Detector (HPD)	$ \eta $	0.0 ~ 1.3
Photo Detector (PMT)	$ \eta $	2.9 ~ 5.2

Description of Hadron Forward

- 165 cm Steel $\rightarrow 10\lambda$, Hadron Calorimeter.
 - Acts as Electromagnetic as well, as there is no ECAL in front.
- Covers large pseudorapidity range, $3 \leq |\eta| \leq 5$
- Active Medium: Quartz Fibers - **Radiation Hard**
- Light, generated via Cherenkov Radiation from showering particles, captured with fibers and read out with PMTs.
- 2 types of fibers per tower: Long and Short
 - Long \rightarrow catches Electromagnetic component
 - Short (22cm shorter) \rightarrow catches only Hadronically developed showers.

CMS Calorimeter

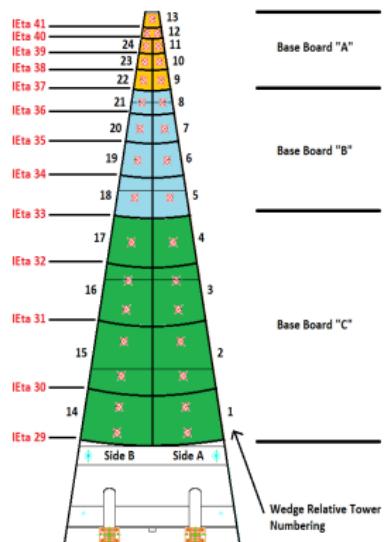


Description of Sourcing

Why Sourcing

Given that our system is located 100m underground, and weighs like a tank, our options are limited.

- To establish Absolute Energy Scale (up to 10%)
- Co-60 γ -source
- 3 sourcing campaigns
 - HFM 2013 - old HF PMTs
 - HFM/P 2014 - new PMTs
- Every Tower is sourced separately (7-8 mins per tube) $\rightarrow 24 \times 36 \times 7 \rightarrow 4\text{days}$ per mode
- Asynchronous data-taking mode (w.r.t. the source)
- 2 data-taking modes
 - 1TS with OV1
 - 2TS with OV2



HF Wedges



Problem statement

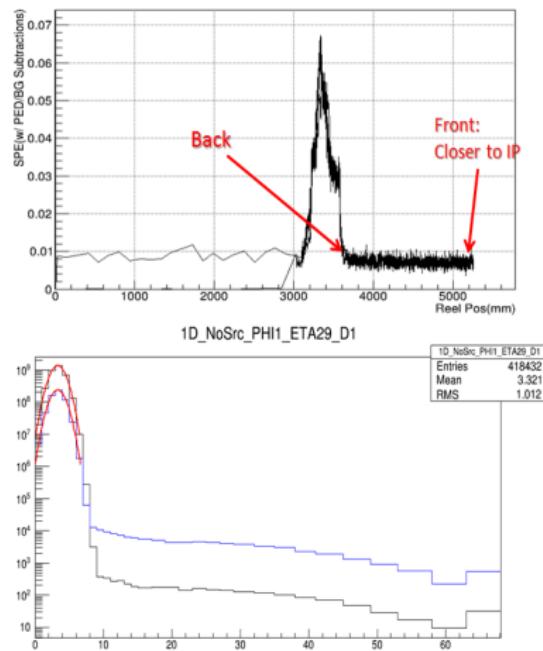
- During Long Shutdown of LHC, new PhotoMultiplier Tubes have been installed on HF.
- General: Transfer the Hadron Forward Energy Calibration used in Run I to be used with new Hardware in Run II.
- Step 1: Extract HF Energy Deposition from the Source using Run I calibration and old PMTs.
- Step 2: Derive the Calibration Coefficients for Run II. Provide Systematics Evaluation. Apply Corrections.

General Description of the Analysis

- Both steps 1&2 proceed in parallel at first - processing all of the data: various modes and voltages.
- Signal Reconstruction - obtain amount of charge collected from the source
- Resolve Issues during the actual sourcing: Channel or Tube swaps, ...
- For step 1: obtain the energy deposition of the source
- For step 2: Use step1 results to compute the calibration coefficient.
- Evaluate the systematics
- Apply correction for the Magnetic Field.

Signal Reconstruction

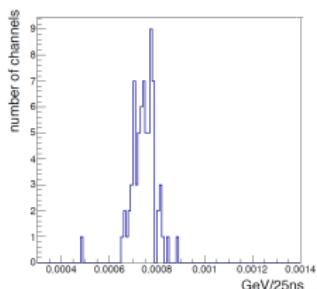
- Signal is reconstructed within $[Back + 300\text{mm}; Front - 300\text{mm}]$
- Add up all the RAW Charge Histograms
 - For Signal - channels in a tower that contains the source
 - For Background - channels outside of the tower being sourced. Geometric Isolation is applied
- Compute the Charge (Source Signal) for every channel
- Correct the charge by the geometry containment factor



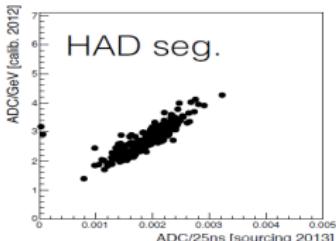
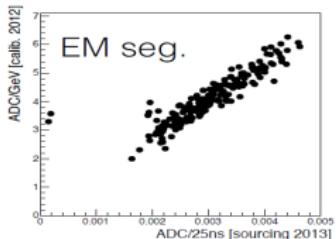
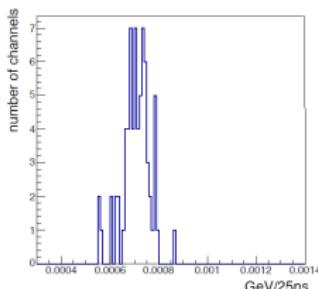
Signal Reconstruction

- The Source Signal (Charge) has been computed
- Extract the Run I Calibration Coefficients
 - Outlier channels have been excluded from further analysis
 - Only channels for $i\eta < 35$ considered (to minimize the raddam effects for higher $i\eta$ towers)
- Energy Deposition is computed separately for H and EM channels.

EM



HAD



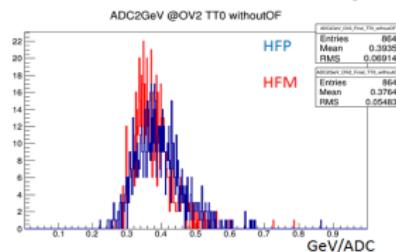
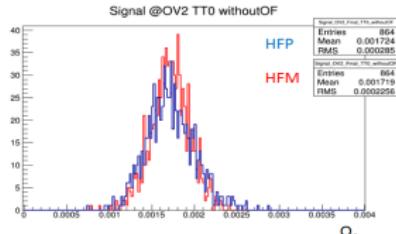
$$\langle E \rangle_c = \langle Q \rangle_c^{\text{Geom}} \times CC_c^{\text{Run I}}$$

Signal Reconstruction

$$\langle Q \rangle_c^{Geom, OV2} = \frac{\langle Q \rangle_c^{Geom}}{nTS} \times \frac{GAIN^{OV2}}{GAIN^{OV1, OV1+100}}$$

$$CC_c^{RunII} = \tau \times \frac{\langle E \rangle^{2013}_{Geom, OV2}}{\langle Q \rangle_c^{Geom, OV2}}$$

- The Source Signal (Charge) has been computed
- Signal is then further tweaked by applying PMT gain ratios and integration window correction.
- Finally Calibration Coefficient is computed.
 - Correct for the source radioactivity drop between sourcing campaigns.



Issues observed during Sourcing

Remark

During the Sourcing Campaigns, a few problems were observed, which, on one hand, made the process more complicated, on the other, it showed the extreme usefulness of the sourcing as the baseline detector validation after reassembly.

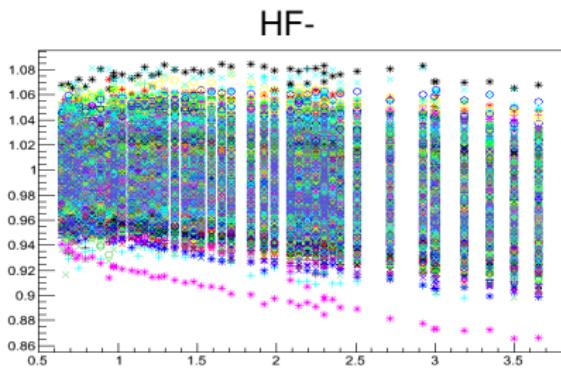
- Channel Swaps - Cables on the QIE side were not connected as expected.
 - No other data-taking mode is able to resolve this problem!
- Source Tube Swaps.
- Source Driver Error - source didn't penetrate the tube or stopped in between.

Systematic Uncertainties

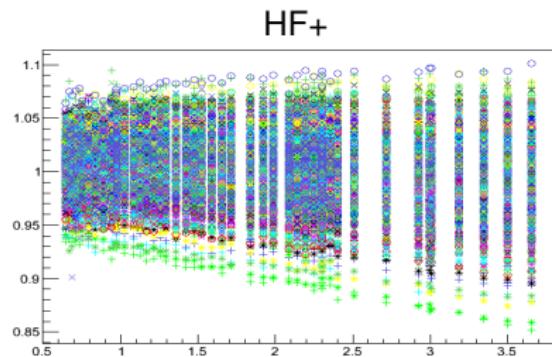
- 1TS vs 2TS - since data-taking has been performed for the whole detector using different integration windows we can estimate the difference
- Transversal Uniformity Tubes A&B - certain towers have 2 source tubes, therefore they were sourced twice and the difference can be estimated.
- Overflow - can estimate the lower bound only, but number of events in the tail is < 1% w.r.t. total
- Longitudinal Uniformity - can use parts of the source path within the tower to estimate the difference in the signal along the path.
- Results have been cross-checked via 2 completely analyses.

Remark

During the commissioning period, before proton-proton collisions, dependency of the PMT response on the magnetic field was evaluated. Calibration Coefficients are then further corrected.



B field(Tesla)



B field(Tesla)

Conclusions

- A brief description of the Calibration Procedure and Results have been provided
- Details on the study of systematics have been omitted.
- This project has been fully completed.

Remark

Calibration Coefficients derived in this project have been successfully used during Run II data-taking campaign!

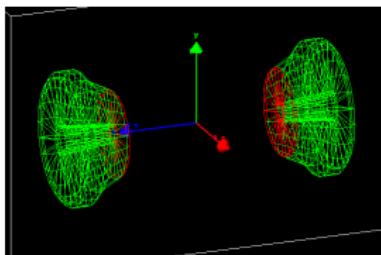
Outline

1 Search for Standard Model Higgs Boson with CMS

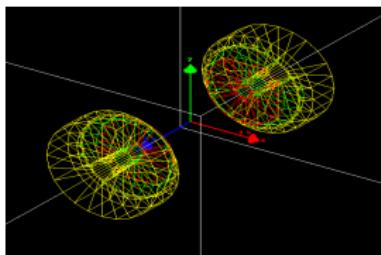
2 Calibration of CMS Hadron Forward Calorimeter

3 Simulations of Modern Calorimeter Systems

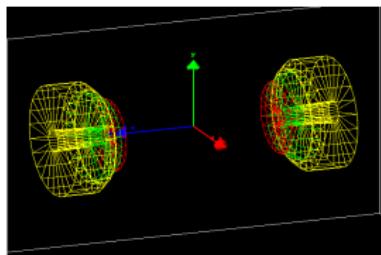
Geant4 based Systems



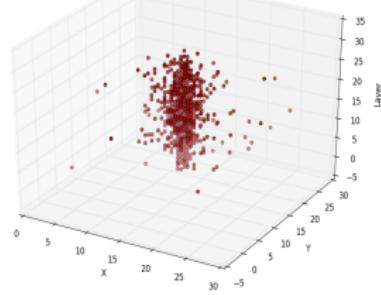
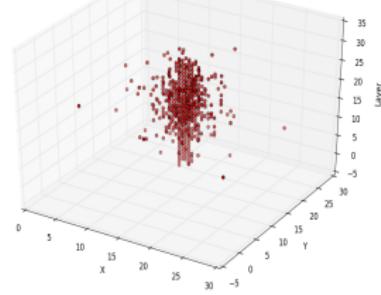
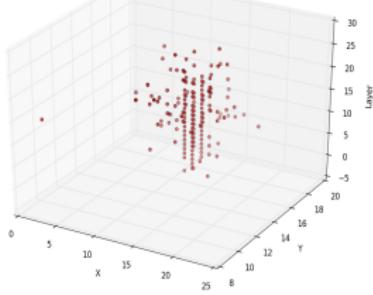
CALICE-like SiW 8GeV e- Hits



CALICE-like SiW 50GeV e- Hits



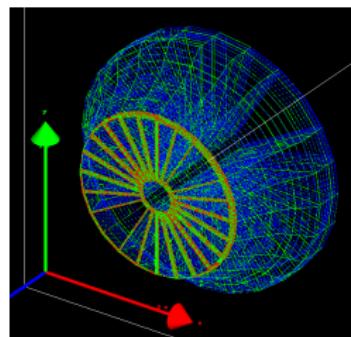
CALICE-like SiW 60GeV e- Hits



Geant4 based Systems

- Experimental High Energy Physics *revolves* around detecting particles, measuring deposited energy, identifying position, ...
- Geant4 is the HEP library that provides such functionality:
 - Build and Visualize the Geometry of the System.
 - Define the properties of various materials
 - Specify the Physics of interest or invent your own.
 - Provides an engine to carry out the actual simulation of the experiment.
 - Allows one to define special regions (Sensitive Detectors) mapped to geometry and responsible for readout.
- Geant4 provides means to create “The Matrix” for Experimental HEP

Full scale custom built
CMS Phase 2 Upgrade
Endcap.



Problem Statement

- Simulate small size Shashlik. Calibrate and establish energy resolution properties
- Simulate small size Hadron Endcap. Calibrate and establish energy resolution properties.
- Simulate small size CALICE-like SiW or High Granularity Calorimeter like system. Calibrate and establish energy resolution properties for both: EM and Hadronic parts.
- Scale up both Shashlik+HE and CALICE-like systems to the proposed CMS Phase 2 dimensions and estimate the “Particle Gun” reconstruction efficiency.
- Use CMS VBF Higgs MC GEN-level samples with 140 PileUp (PU) to estimate Higgs Reconstruction properties

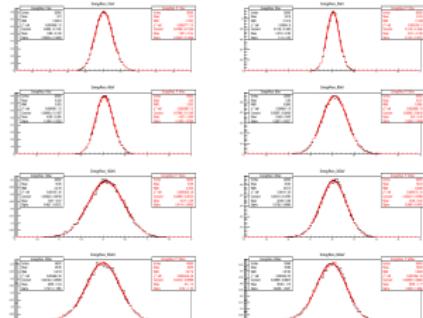
Problem Statement

- Simulate small size Shashlik. Calibrate and establish energy resolution properties
- Simulate small size Hadron Endcap. Calibrate and establish energy resolution properties.
- Simulate small size CALICE-like SiW or High Granularity Calorimeter like system. Calibrate and establish energy resolution properties for both: EM and Hadronic parts.
- Scale up both Shashlik+HE and CALICE-like systems to the proposed CMS Phase 2 dimensions and estimate the “Particle Gun” reconstruction efficiency.
- Use CMS VBF Higgs MC GEN-level samples with 140 PileUp (PU) to estimate Higgs Reconstruction properties

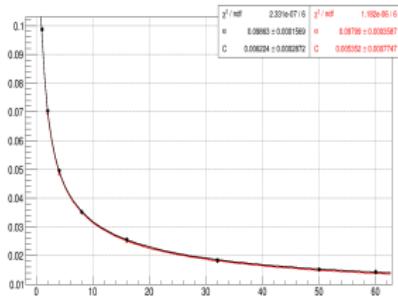
Shashlik

- 25 – 30 X_0 Electromagnetic Calorimeter with **no Longitudinal Segmentation**
- Alternating layers of W (tungsten) and LYSO plastic scintillator.
- Try both: Geant4 Scintillation mechanism and **parametrized scintillation response**.
- No simulation of photodetectors.
- Use e^- beam with 8 different energies.

Reconstructed Energy

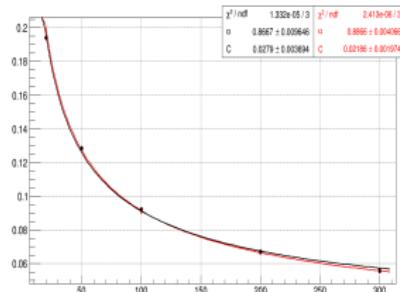
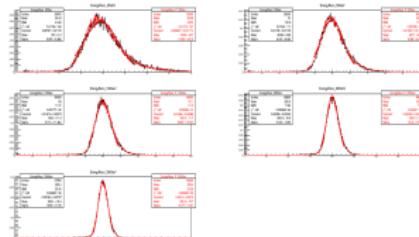


Energy Resolution



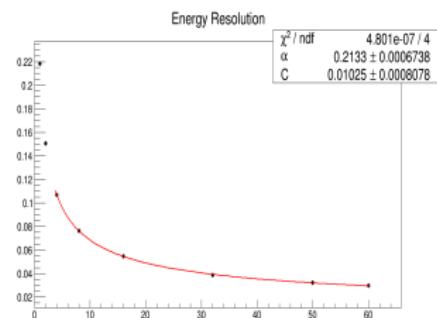
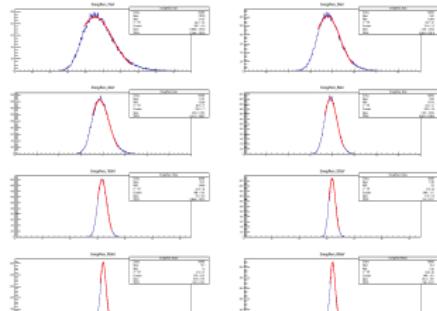
Hadron Endcap

- approx. 10λ Hadronic Calorimeter
- Alternating layers of Brass and SCSN-81 plastic scintillator
- Both G4Scintillation and **parametrized scintillation response** simulated.
- No simulation of photodetectors
- Use π^- beam with 5 different energies.
- Results are consistent with **CMS NOTE 2008/010**



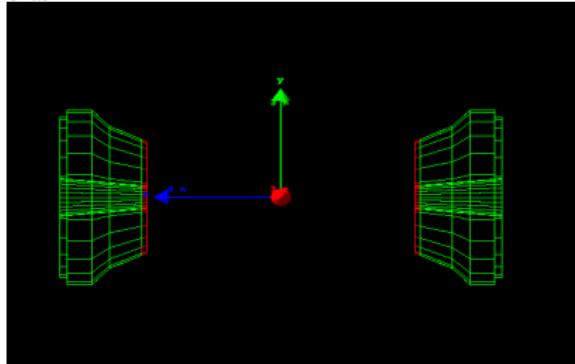
CALICE-like SiW/CMS High Granularity Calorimeter

- 24 X_0 Electromagnetic Calorimeter with **Longitudinal Segmentation**
- Alternating layers of Lead (Pb)/Si/PCB(G10)
 - Tungsten (W) and Lead (Pb) were tried. W is used for CALICE-like and Pb was used for HGC.
 - PCB and G10 are materials of the electronics boards that are “sandwiched in”.
- 80e/hole pairs per 1um - a simple assumption on the number of electron-hole pairs generated by the traversing charged track in Si used.

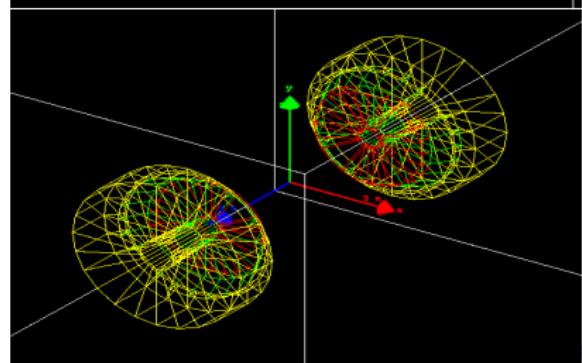
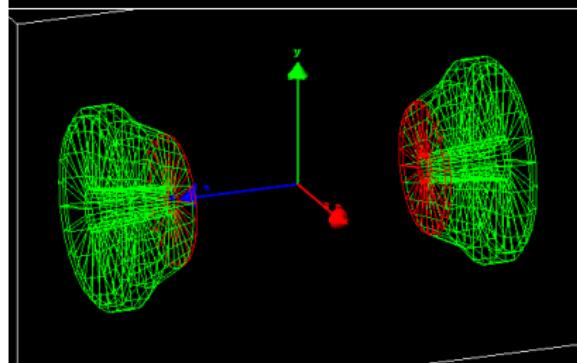
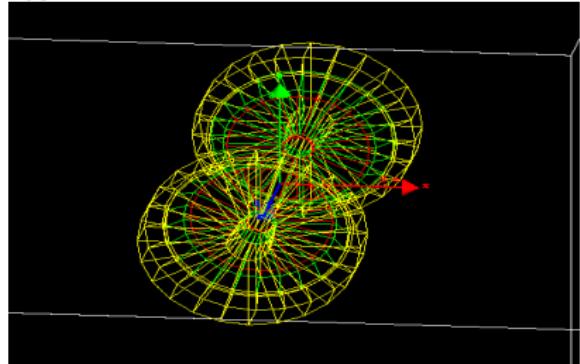


Full Scale CMS Phase 2 Options

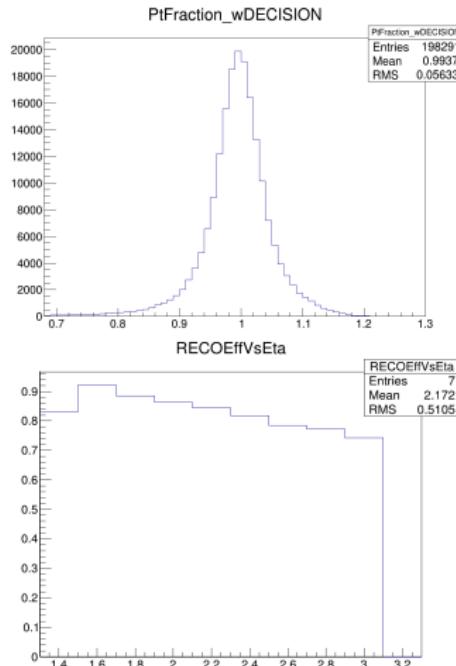
Shashlik



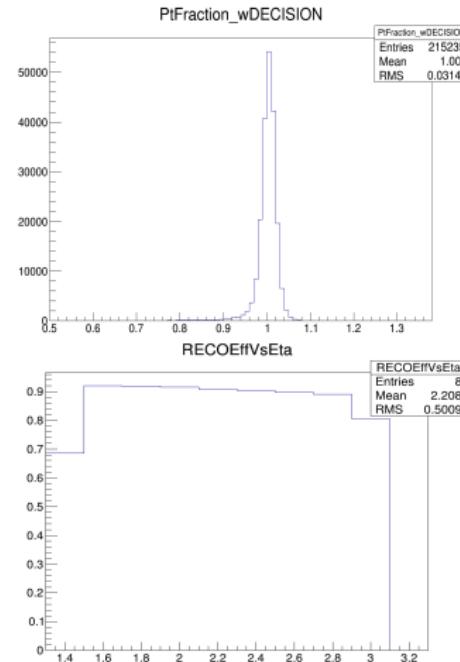
HGC



Full CMS scale Phase 2 with Particle Gun



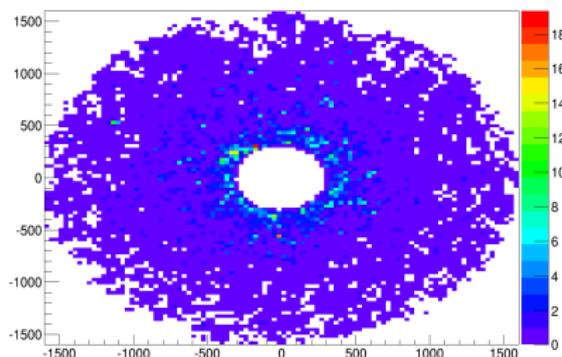
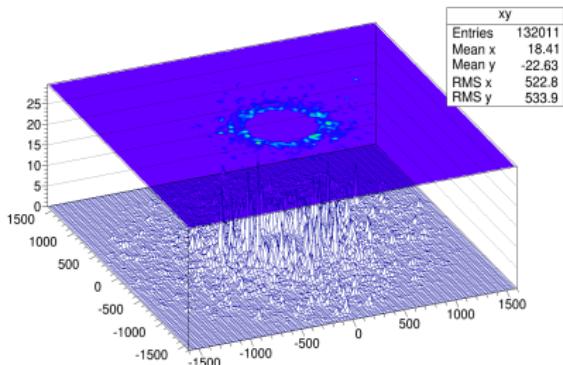
HGC. 3×3 clusterization RECO. Each layer is seeded individually



Shashlik/HE. 3×3 clusterization RECO.

Custom built full CMS scale Phase 2 with 140PU

- Added 140 Min Bias events for 1 VBF Higgs (to $\gamma\gamma$)
- Very time and memory consuming
- Event Displays: 3D and XY-plane projection



Summary

- Results for **3 projects** have been presented!
- **Search for SM-like Higgs Boson** in proton-proton Collisions at $\sqrt{s} = 13\text{TeV}$ with CMS Experiment.
- **Calibration of Hadron Forward Calorimeter** for LHC Run II.
- **Modern Calorimeter Systems Simulations** with Geant4.