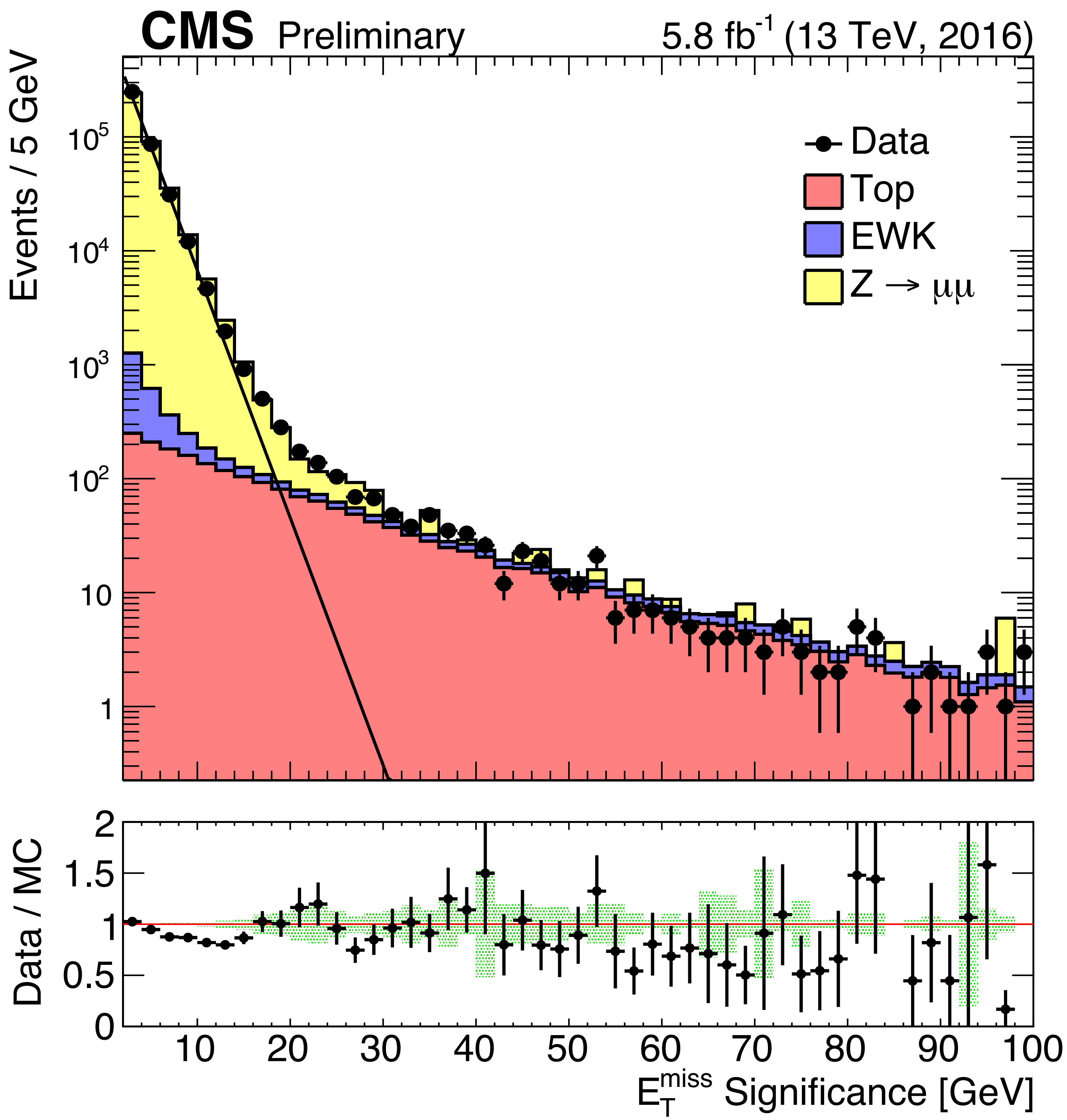
# Detector Simulation in CMS

India
June, 2018

Sunanda Banerjee

# Introduction




- One key ingredient in data analysis is an understanding of the detector performance and estimation of efficiency, (ir)reducible backgrounds, systematic uncertainties, ….
- The most important tool for these is simulation of signals of physics events in the CMS detector

- The reliability of MC prediction depends on
  - Goodness of Physics models of the event generator
  - Realistic description of the CMS detector and quality of models used in evaluating propagation of particles through the detector

# CMS Simulation Program

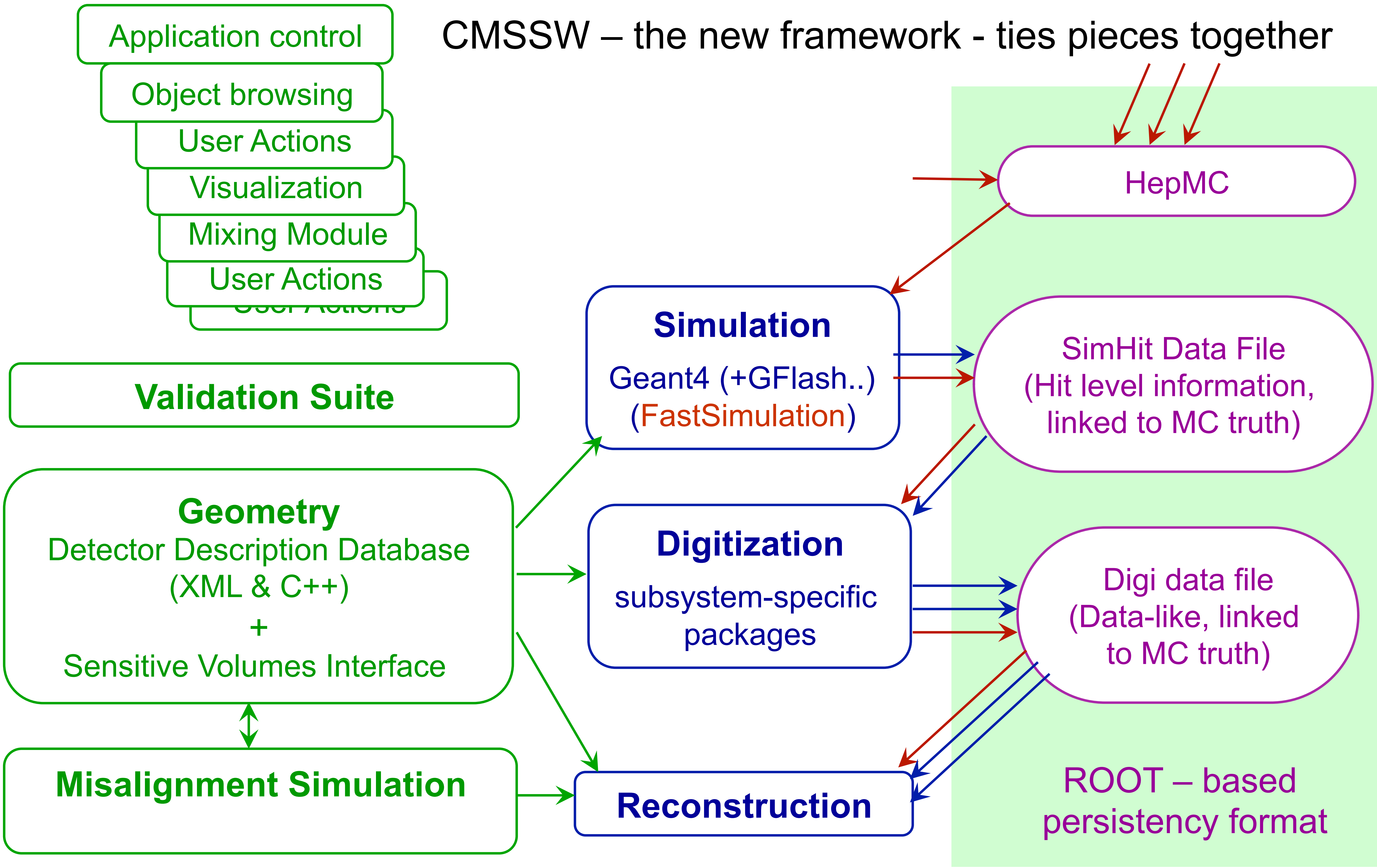
- Though in operation for a number of years, it's a live system – goals, requirements, tools evolve throughout the lifetime of the experiment

- CMS simulation program is based on the Geant4 toolkit. Geant4 provides the following functionalities:
  - Physics processes: electro-magnetic and hadronic interactions.
  - Tools for detector geometry and sensitive  element response.
  - Interfaces for tuning and monitoring particle tracking

- + CMS offline framework and Event Data Model which:
  - manage application control at run time
  - rely on the concept of event processing module (EDProducer)
  - provide interface to common tools (generators, magnetic field, MC truth handling, infrastructure for hits, event mixing, digitization, …)
  - ensure provenance tracking and event immutability

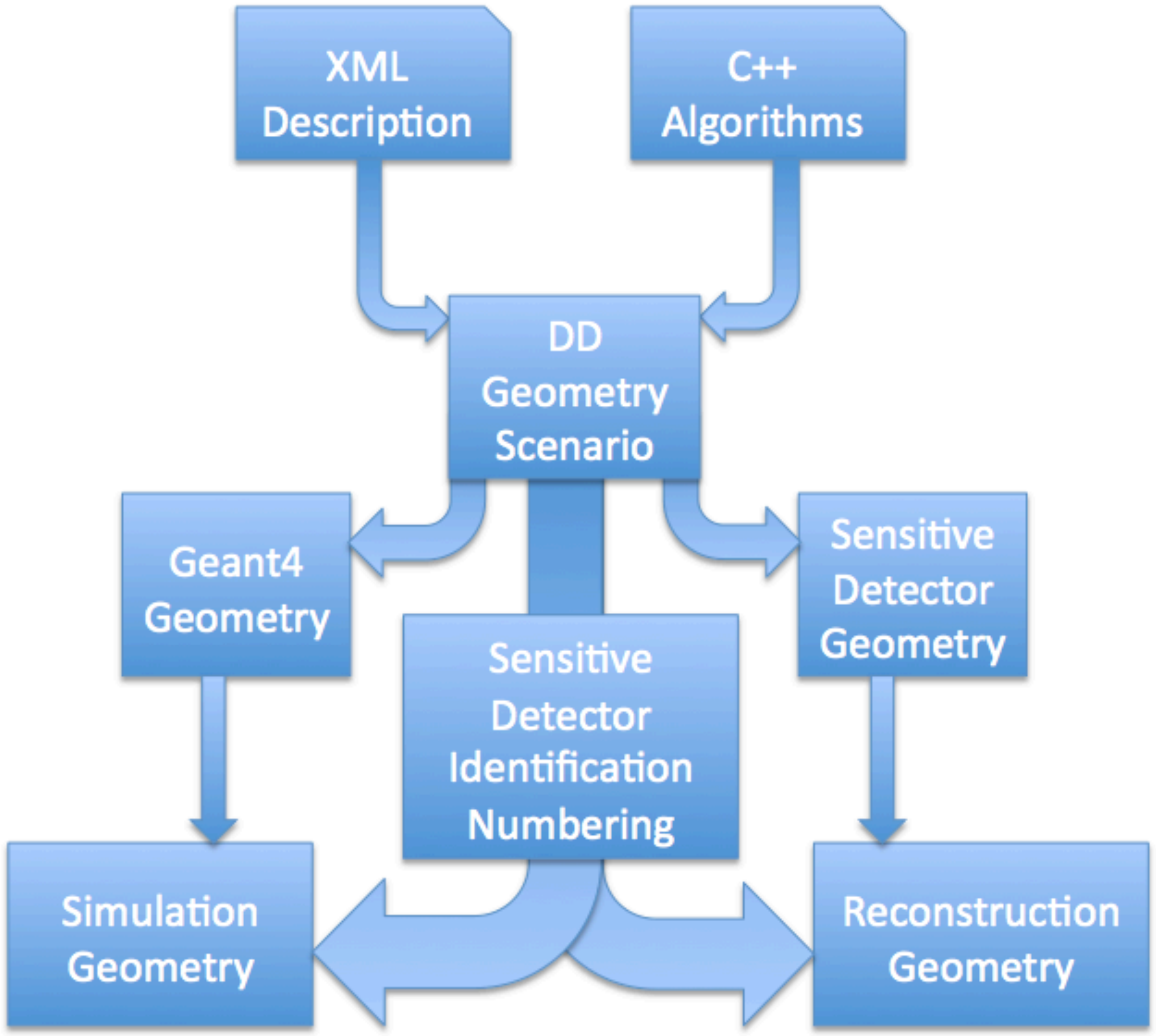
- Emphasis given to simulation program
  - Agreement with the data through most reliable physics models
  - Robust, performant and adapted to modern technology
  - Extended to describe newer detectors (changes in current design and an eye for the future)

# Simulation Software – CMS Solution



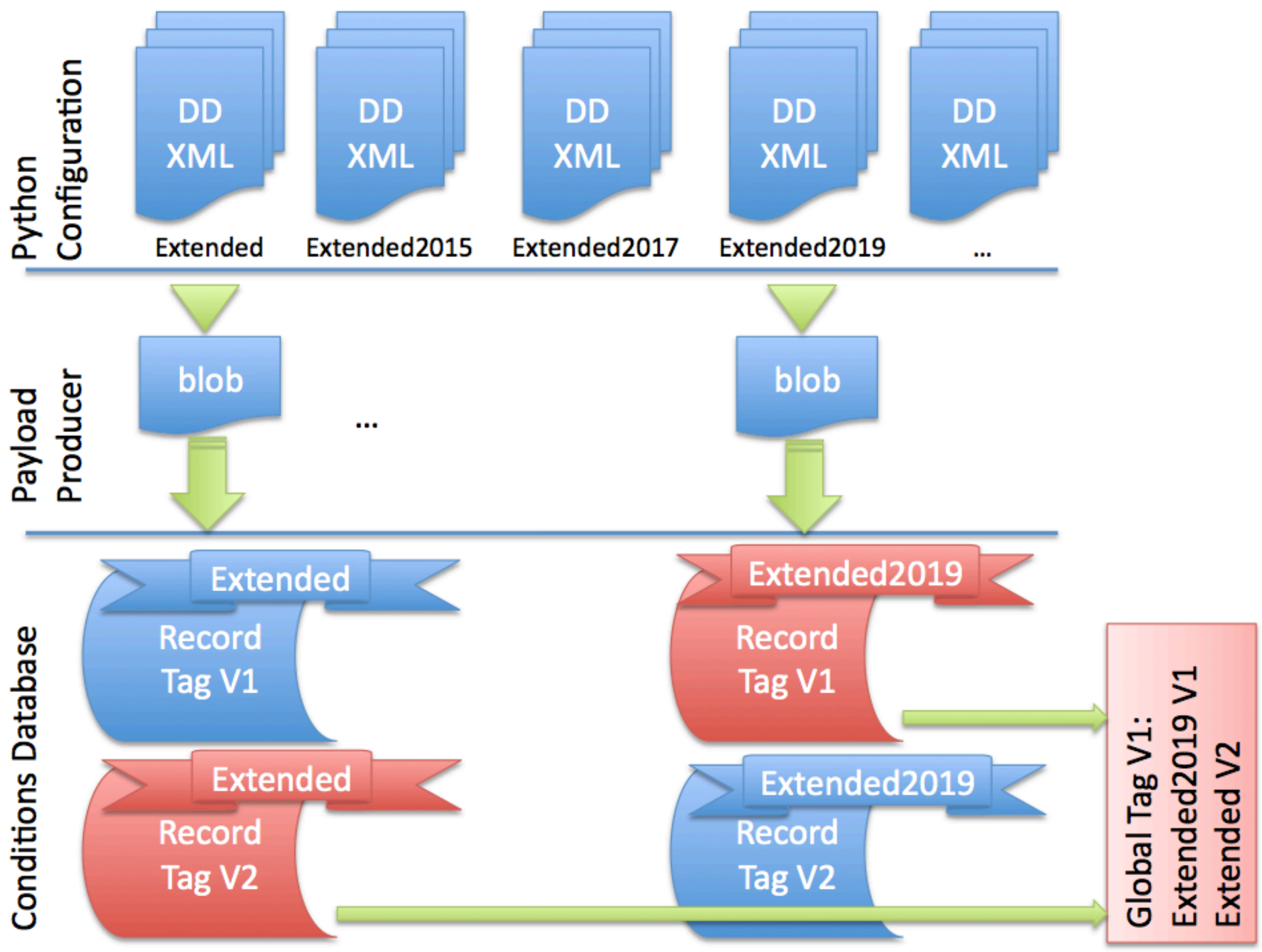
# Source of Geometry

- CMS geometry is currently defined through Detector Description Language (DDL)
  - Contains all static information: geometry, material, sub-detector specific constants
  - Accessed by a single interface in all applications: simulation, reconstruction, analysis
  - Realized through parameters defined using XML description and some specific C++ algorithms which utilizes some symmetry properties in hierarchical positioning
  - DDL is also used for CMS magnetic field geometry

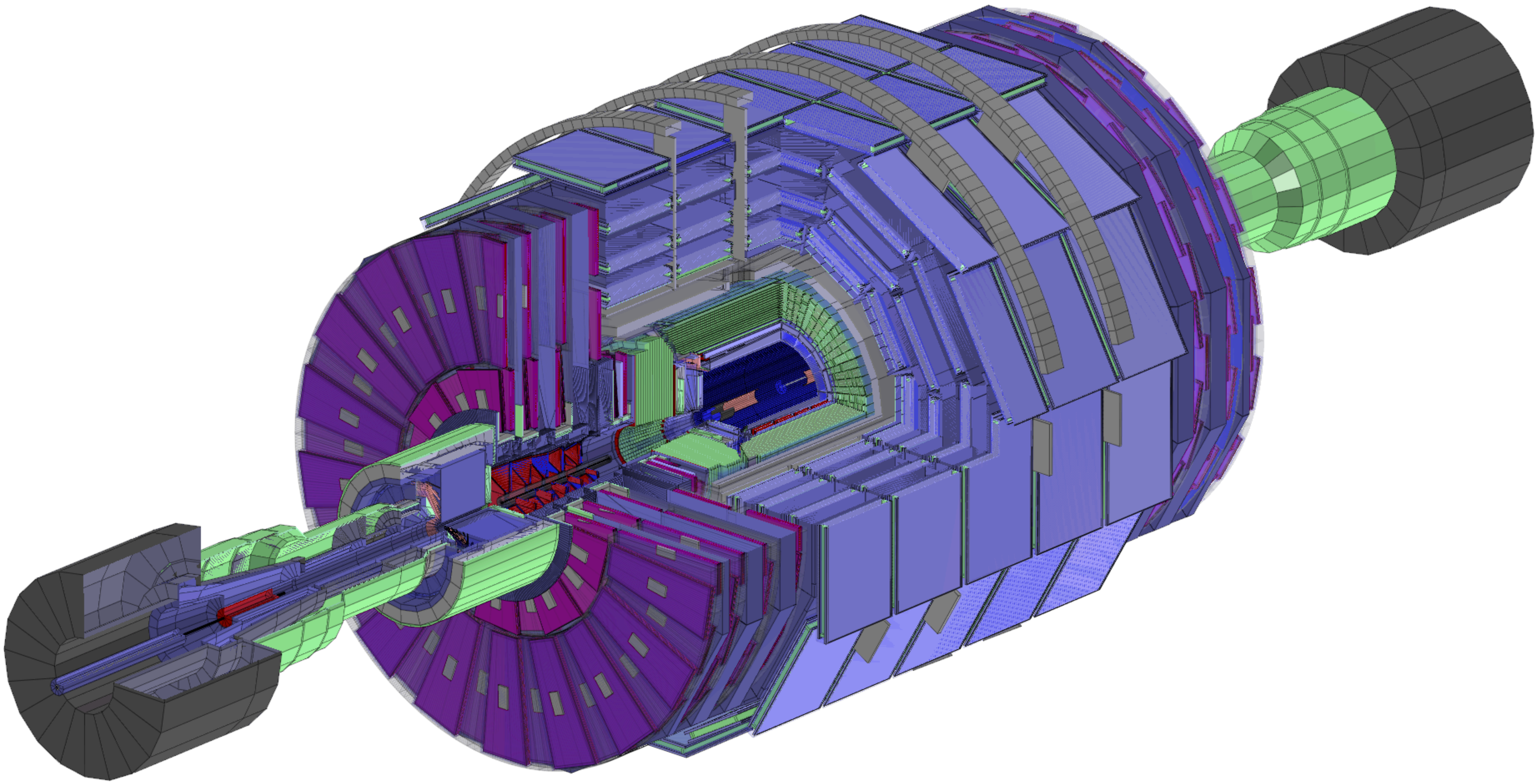


# Views of Geometry

- A set of XML files which includes the parameters as well as reference of the C++ algorithms define a scenario
- Several such scenarios exist and each of them are put in the condition database as a blob
- Application programs can access either of the two sources
  - Production jobs always access the geometry files in the condition data base
  - XML files are used for the development of new geometries



# CMS Geometry



- Emphasis has always been to define sensitive part of the detector (and its proximities) in great detail
- For tracking detectors all heavy spots are defined accurately and the rest using average material definition
- Material budget is verified by (1) checking the weight as defined against the real weight of the component; (2) radiography of the detector through studies of interaction and conversion vertices

# Run 2 (2017) Geometry

- Utilizes 3944 solids in 4257 Logical Volumes among 2.1 million touchable. A relatively new shape CutTubs used from an algorithm
- Summary of solids used in touchables:

| | Standard | Reflected |
|---|---|---|
| Box | 998k | 434k |
| Tube | 93k | 1.4k |
| Trapezoid | 238k | 149k |
| Cone | 1.9k | 0 |
| Polycone | 328 | 32 |
| Polyhedra | 2.5k | 0 |
| Torus | 160 | 32 |
| PseudoTrapezoid | 96 | 0 |
| TruncatedTube | 92 | 0 |
| UnionSolid | 134k | 0 |
| SubtractionSolid | 50k | 702 |
| IntersectionSolid | 1.1k | 0 |

- Overlap summary:

| Tracker | ECAL | PreSh. | HCAL | Muon | Infra.St. | Total |
|---|---|---|---|---|---|---|
| 0 | 47 | 26 | 0 | 73 | 236 | 382 |

# Phase 2 Geometry (one proposed scenario)

- Utilizes 5177 solids in 5372 Logical Volumes among 21.9 million touchable
- Summary of solids used in touchables:

| | Standard | Reflected |
|---|---|---|
| Box | 743k | 429k |
| Tube | 49k | 1k |
| Trapezoid | 160k | 141k |
| Cone | 1.9k | 0 |
| Polycone | 176 | 32 |
| Polyhedra | 16.9M | 0 |
| PseudoTrapezoid | 96 | 0 |
| TruncatedTube | 92 | 0 |
| UnionSolid | 131k | 0 |
| SubtractionSolid | 3.4M | 594 |
| IntersectionSolid | 70 | 0 |

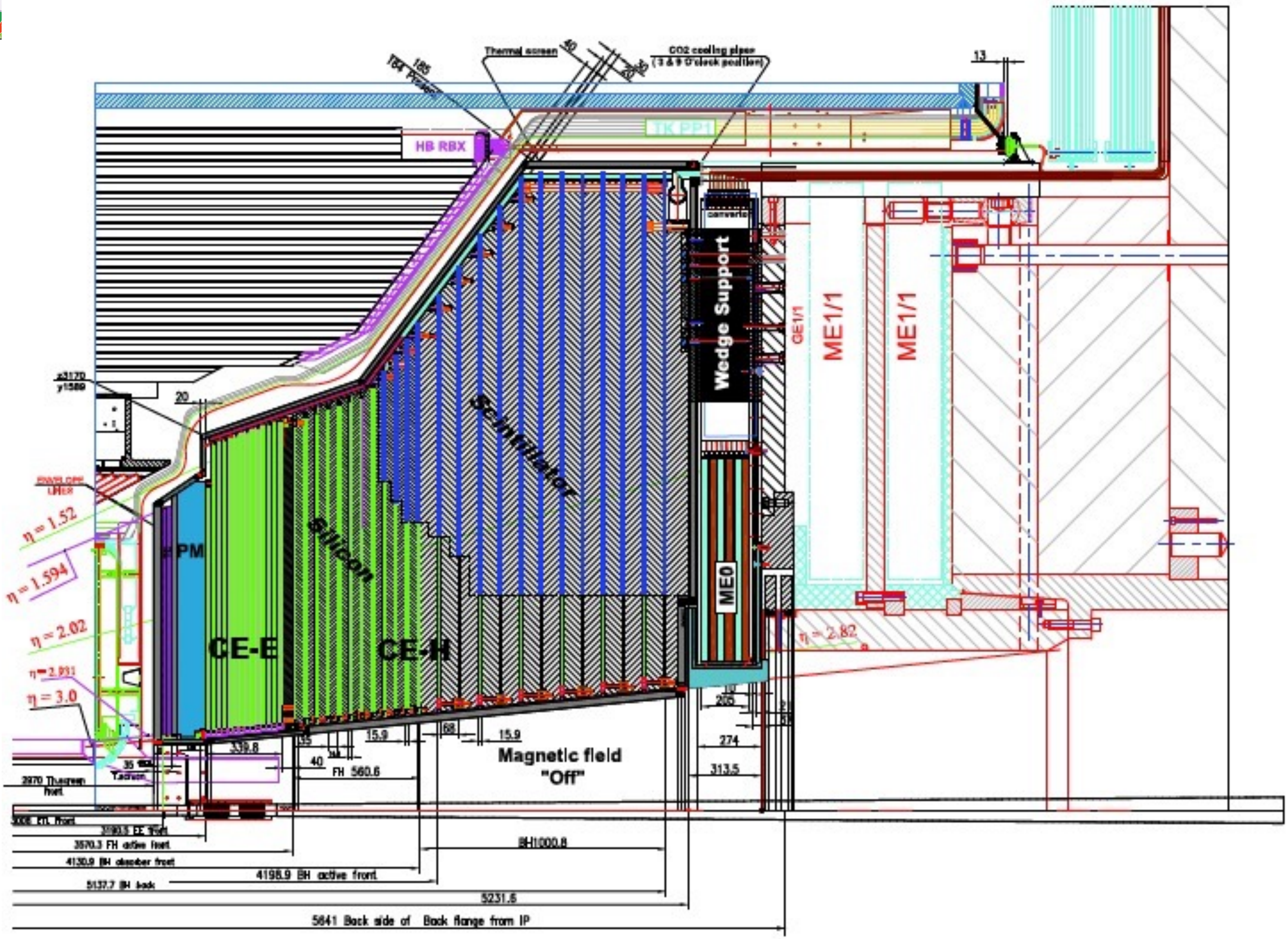
- Overlap summary:

| Tracker | ECAL | PreSh. | HCAL | Muon | Infra.St. | Total |
|---|---|---|---|---|---|---|
| 0 | 51 | 0 | 0 | 81 | 224 | 356 |

# Future of CMS Geometry

- CMS is fighting to remove the overlaps.
  - Some of the overlaps may be artifact of the method used in overlap detection (seen for ECAL barrel - subtracted polyhedra)
  - Some overlaps in non-critical area (absorber structure in return yoke)
- CPU profiler for the 2017 geometry shows Polycone (~8%) and Polyhedra (~4%) have sizable contribution to overall CPU usage
- CMS would require
  - More performant polycone as a standalone shape or taken from VecGeom
  - Same for Polyhedra shape
- Phase 2 CMS detector sees
  - Number of touchables increase by an order of magnitude
  - Substantial increase in number of polyhedra and subtraction solids
- These increases are due to the high granularity calorimeter utilizing hexagonal wafers with hexagonal and half hexagonal cells. This type of detector could be used in many other applications.

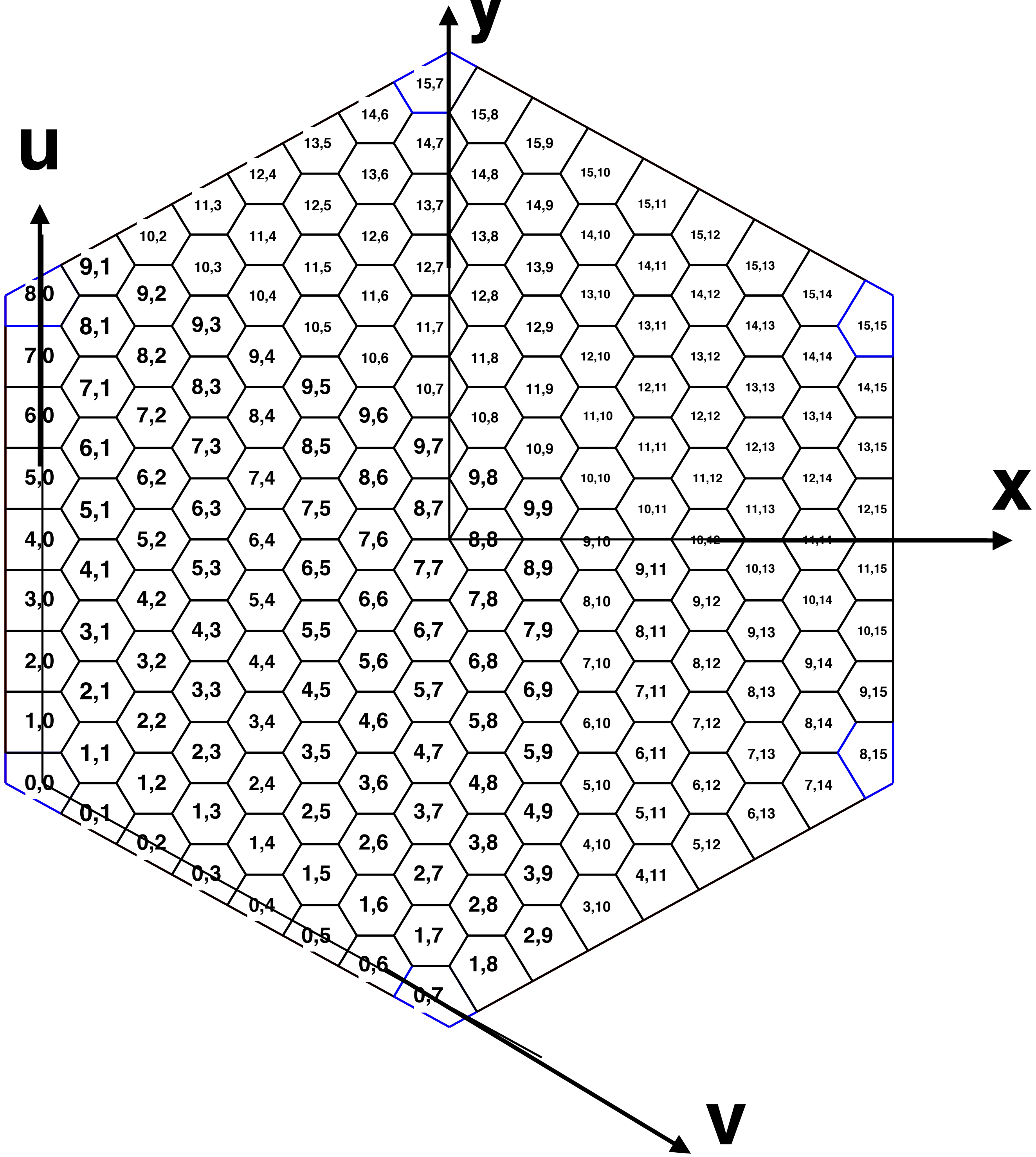
# The New Layout of Endcap

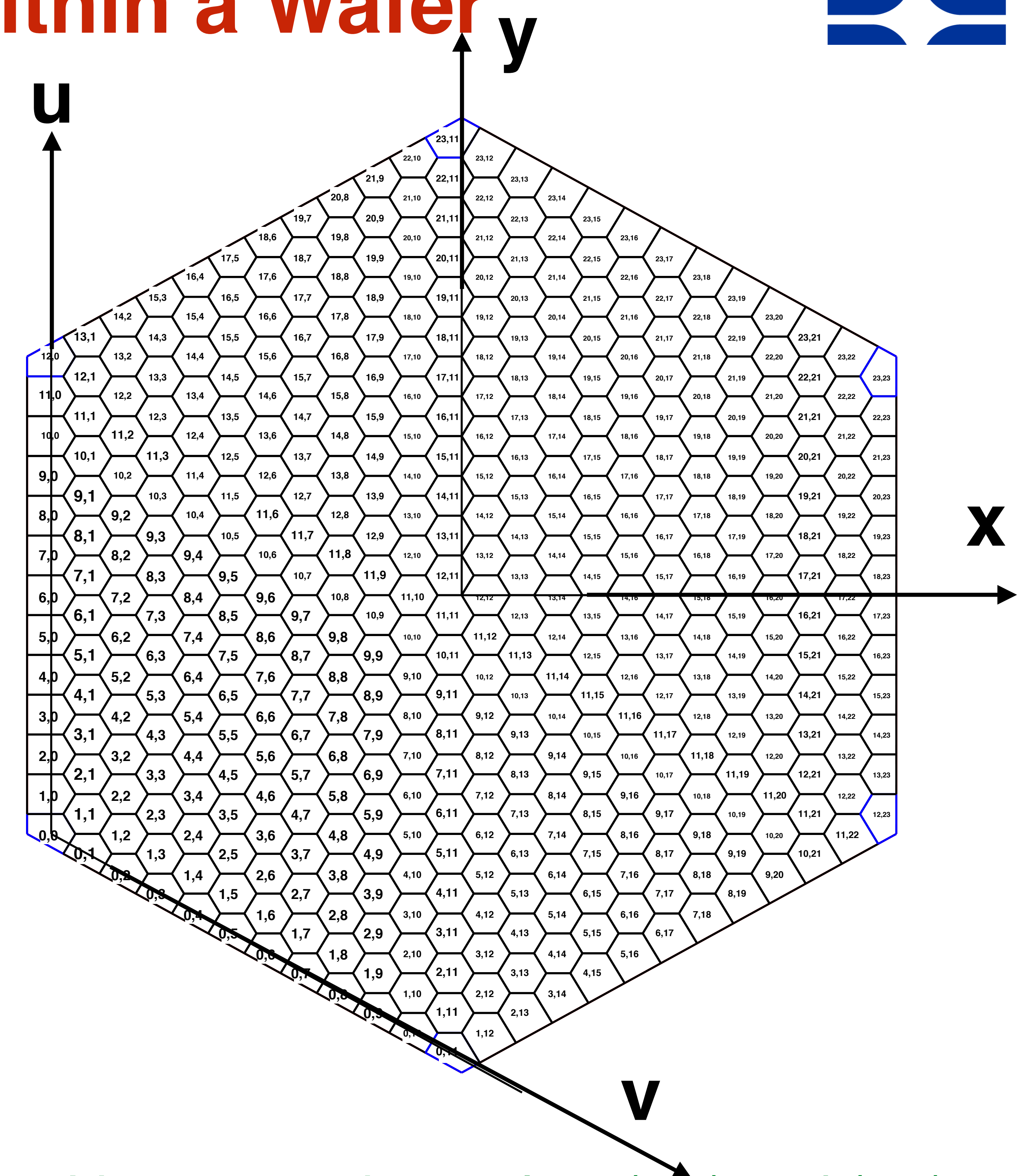


# Tasks to be Undertaken for HGCAL

- Get rid of "BH as part of HCAL" and make all 52 layers of HGCAL part of a single detector description
- Update longitudinal structure to match the TDR description
- Describe the "mixed scintillator and silicon" layers
- Restructure the detector numbering scheme
- Provide navigation tools
- Update silicon wafer size (6" to 8") and introduce inter-wafer gaps, and mouse bites
- Using the max wafer parameters describe the cell layout; inclusion of "calibration cells"
- Introduce final "incomplete wafers" geometry at inner and outer boundaries
- Correct wafer thicknesses (120, 200, 300 microns); separate silicon physical thickness and active thickness
- Provide more realistic active area coverage at the inner/outer boundary and allow flexibility to adjust the coverage
- Introduce inter-cassette gaps
- Describe mechanical structures on the inner and outer boundaries (support cone, thermal shield etc.)

# Cell Definition within a Wafer



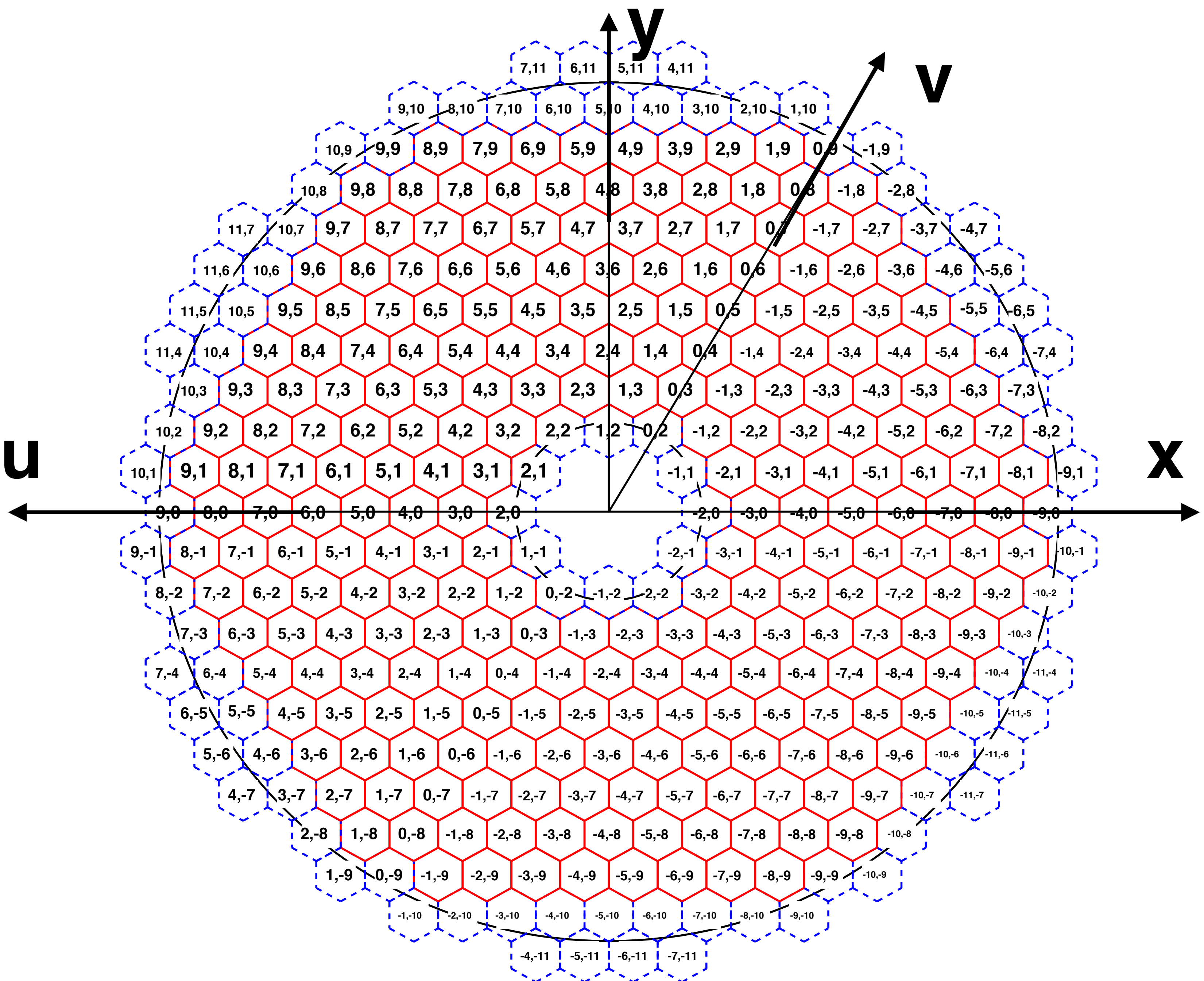




- Cells are numbered by u,v coordinates and can be transformed to row,column indices by trivial linear expressions.

- Linear equations relate (u,v) and (x,y)
- Maximum number of cells along u/v:
  - 16 for coarse cells
  - 24 for fine cells

# Wafers in a Layer

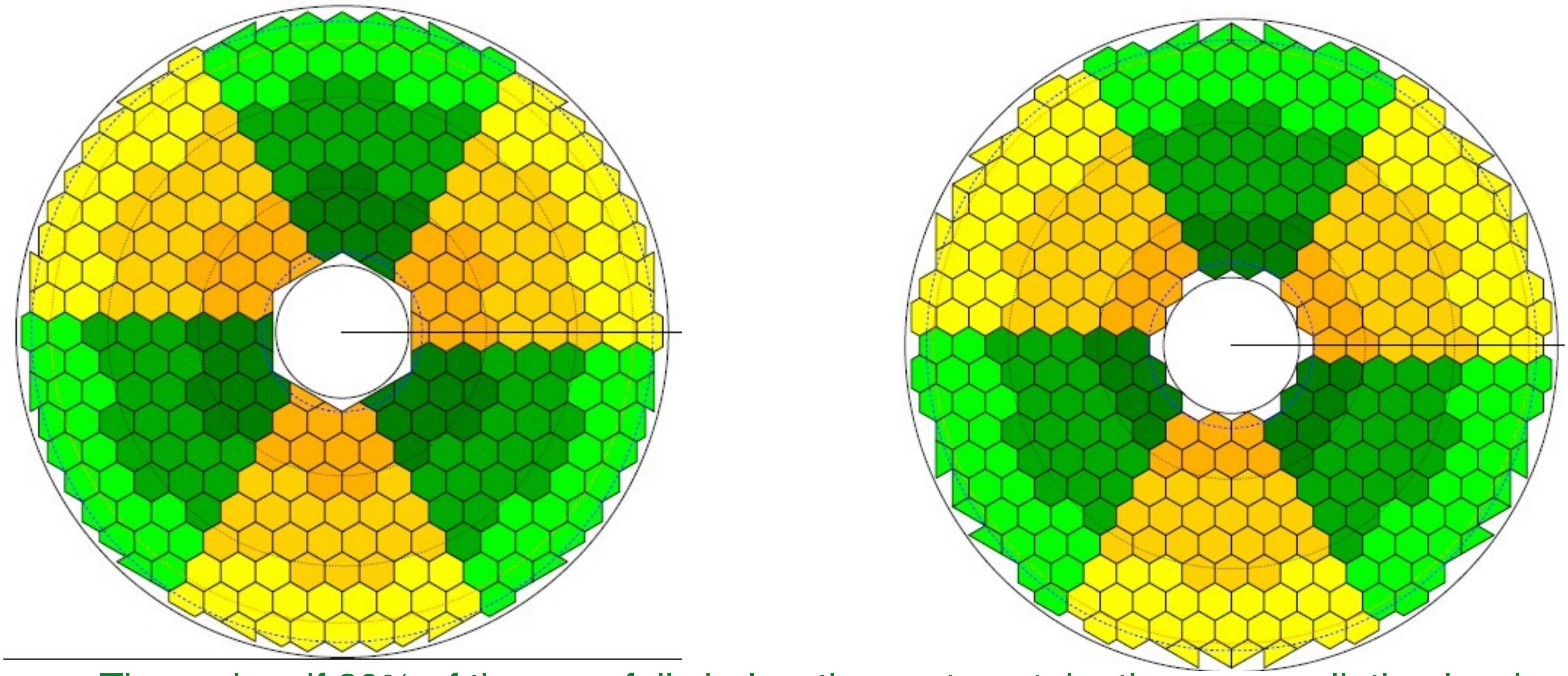
- Positioning wafers in a plane is done using full wafers only
  - Real wafers where all 6 corners are contained within the boundary
  - Virtual wafers where at least 1 corner is contained but not all 6




- The center is defined on the beam axis
- There is a simple linear relation to go from (u,v) to (x,y) scheme using

# Wafer Type

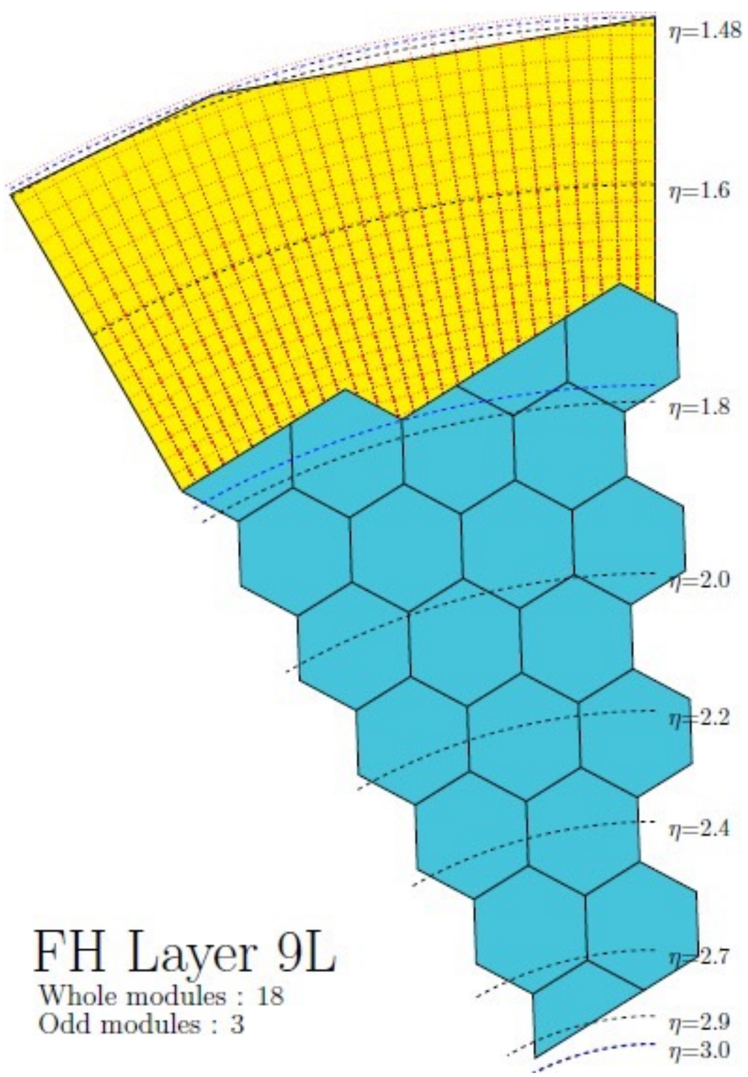
- Three different wafer types are there to handle different radiation environment:
- The boundaries for the 3 types depend on |z| position of the layer and the boundary is defined by a theoretical radius value (parametrized by 4th order polynomial)



- The recipe: if 20% of the area falls below the contour, take the more radiation hard option

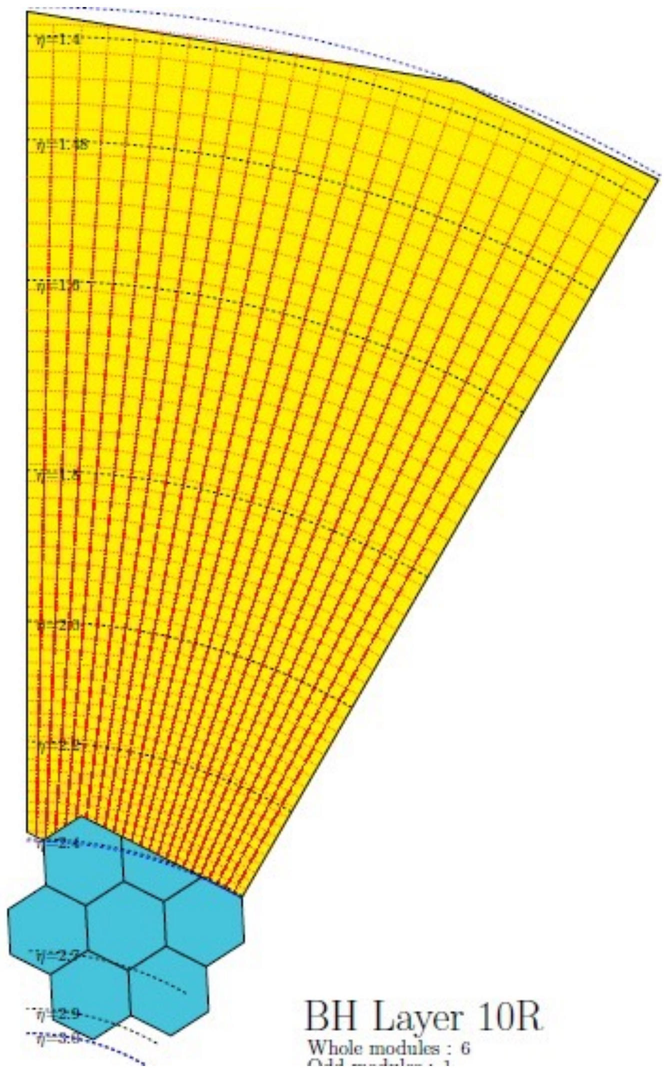
# The Mixed Section

- Divide a mixed section into two separate rings one containing wafers and the other scintillators
  - Bottom ring contain wafers with base plates, PCB's, .. and wafers are defined as real + virtual ones
  - The top ring contains scintillator, cables, connectors, PCB, …
    - The scintillator layer is defined as a disk and inner structure will be done at SIM level



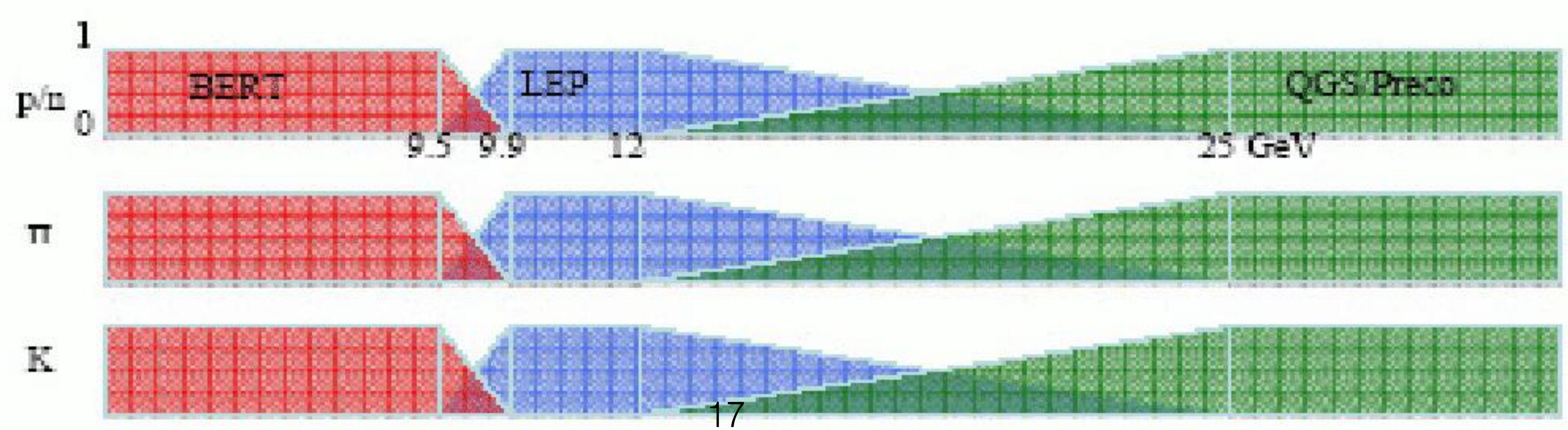

Two possible ways to define the partition:

- Use the theoretical radius obtained from a consideration of the radiation level
- Use the maximum extension of the silicon portion from a consideration of protecting scintillator part using similar consideration of fractional area



# Interface with Geant4

- Core application = framework-based Event Data Producer with a customized interface between Geant4 and CMS Event Data Model
- Geometry is available to either simulation or reconstruction via the framework EventSetup;
- Sensitive detectors associated with geometrical volumes through XML configuration files at run time
- Magnetic field based on dedicated geometry of magnetic volumes; provided by independent subsystem via EventSetup; field selection, propagation tuning configurable at run time
- Variety of lists (LHEP, QGSP_BERT, QGSP_FTFP_BERT, FTFP_BERT,…) for modeling physics processes; run-time selection of physics list and production cuts, activation/tailoring of individual processes;
- Variety of Physics event generators (particle guns, Pythia, Herwig,…); generator information stored in HepMC format and interfaced to G4Event
- User actions allow access to Geant4 objects at any stage (run, event, track, step); used for tuning, diagnostics, custom bookkeeping
- Monte Carlo truth record with decay/interaction history of the generator's particles and selected tracks from Geant4 simulation



# Physics Lists in CMS

- CMS used the physics list in the past for its Monte Carlo production
  - QGSP_FTFP_BERT_EML (with Geant4 versions 9.4.p02, 9.6.p02)
- CMS moved to multithreading mode from beginning of Run2 (2015)
  - QGSP_FTFP_BERT_EML (with Geant4 version 10.0.p02)
- CMS moved to a new physics list for its 2017 MC production
  - FTFP_BERT_EMM (with Geant4 version 10.2.p02)
- and again moved to a new Geant4 version (10.4) for 2018 production
- FTFP_BERT is the recommended physics list from Geant4 collaboration (J.Allison *et al*. NIM A506, 2003, 250; NIM A835, 2016,186)
- The list QGSP_FTFP_BERT combines QGSP, FTFP, Bertini Cascade models for π/K/p/n with a fixed validity region:
  - Bertini Cascade valid at ≤ 8 GeV
  - FTFP valid between 6 and 25 GeV
  - QGSP valid at ≥ 12 GeV
- The list FTFP_BERT uses FTFP and Bertini Cascade models:
  - Bertini Cascade valid at ≤ 5 GeV
  - FTFP valid at ≥ 4 GeV
- EML, EMM specify the physics models for electromagnetic processes
  - EML utilizes simplified multiple scattering model for all detectors
  - EMM uses the detailed multiple scattering model for HCAL and HGCAL and the simplified one for other detectors (handling of multiple scattering is critical for sampling calorimeter)

# Task of the Digitizer

- Convert the energy deposit record in the sensitive detectors (tracking type and calorimetric type) to signals as expected from the real detectors
  - Two distinct types of actions required for the two detector types
  - Also hit information are different for tracking type and calorimetric type detectors
    - For calorimetric detectors, information exist for all individual readout unit in finite time slices (~1 ns). However the energy deposits are stored as cumulative total if the source is the same particle or one of its shower product.
    - For tracking detector, every crossing with the sensitive detector is stored as a hit and the sensitive detector does not have finer most cell identification (e.g. wafer rather than strip or pixel). Here store all information like entry, exit point, energy deposit, timing, ….
- Include the effect of pile-ups which could be in time (multiple interactions during the bunch crossing of interest) or out of time (due to interactions from previous or later bunch crossings which will interfere because of finite response time of the electronics)
  - These effects are common to all detector types and are handled using common interface

# Digitization of Tracking Detectors

- The digitizer tries to take care of the following effects
  - Convert energy deposit to charge taking care of electrons produced, its fluctuation, drift in electric field (effect of Lorentz angle), …
  - Charge diffusion which may cause signal in multiple readout units
  - Smearing of the charge
  - Addition of noise
  - Take care of several effects
    - non-linearities and thresholds
    - miscalibration
    - noisy read out cells
    - inefficiencies and dead cells
    - saturation
    - aging and radiation damage
    - pulse shape
- Final output has digital information for every readout channel (with appropriate zero suppression algorithm)
  - a unique cell identifier
  - ADC and/or TDC information
- An external object to link the digit to simulation hit

# Digitization of Calorimetric Detectors

- The digitizer tries to take care of the following effects (all calorimeters use either scintillation or Cernekov photons as primary signal)
  - Converts hit energy to photo electrons
  - Do photo statistics
  - Convert PE's to analog signal
  - Smear the charge
  - Add noise (due to photo-transducers)
  - Take care of non-linearity effects of the photo-transducers
  - Simulate the electronics taking care of
    - pulse shape
    - electronic noise
    - time slew effects
    - noisy or dead cells
    - radiation damage
- Again store digital information for every readout channel (with appropriate readout option)
  - Unique identifier of the readout channel
  - ADC (and TDC) information

# Mixing Module

- Task of the mixing module is to add a given number of minimum bias events to the signal event to mimic in-time and out-of-time pike-up effects
- To get a coherent software scenario, the digitization process has to happen after the mixing is completed → mix the SimHit information
  - For high luminosity operation average number of in-time pile-ups is rather large and for many detectors (barrel muon in particular) one need to consider a large number of bunch crossings
  - The mixing scenario was revised even during Run 1 operation to optimize the mixing stage
    - Each detector is called once for each event to be added to accumulate the hits
    - A final call is make when for the event the accumulation is completed and the digitization can be made
  - For high pileup runs a new approach is made using “premixed” events
    - A digitized sample for a certain running period is made with a given pileup scenario from a set of minimum bias events including in-time and out-of-time scenario
    - This “premixed sample” is used in the mixing module
    - The raw format of CMS hits is extended to ensure sufficient precision for making sums of small pulse heights in the Digi step

# Data Mixer

- Instead of mixing a large number of minimum bias simulated events with a signal MC event, events from real collision is considered for the mixing process
- A random collision event represents the PU conditions of the machine at a given time. It includes properties of boys in-time and out-of-time PU interactions.
- CMS collects zero bias trigger events at the rate of 1 Hz and can be increased to 10 Hz if required
- To make appropriate usage of this one need to worry about
  - zero suppression effect in the detectors
  - mis-alignment in the detectors
    - simulation uses perfectly aligned detectors and effect of mis-alignment need to be taken care of by mapping each misaligned cell to a perfectly aligned cell
      - may lead to some edge effect in certain detectors
- The data mixer approach is tried for calorimetric objects
  - relative variation of this approach w.r.t. full simulation approach (without pre-mixed events) is significantly smaller (by a factor 5) than jet energy resolution

# Recent Changes in Geant4 Version

- Geant4.10.2.p02 was chosen for 2017 over version 10.0.p02 from the following considerations:
  - It has full implementation of Multithreading mode
  - It is C++11 compliant; can be used with newest compilers
  - It includes many technical improvements and fixes
  - Its is faster than 10.0 and requires less memory in general G4 benchmarks
  - It has improved physics predictions
- Choose physics list FTFP_BERT over QGSP_FTFP_BERT
  - Geant4 collaboration considers it to be most appropriate for LHC experiments (J. Allison et al., NIM A835, 186, 2016)
- Geant4.10.4 is chosen for 2018 over version 10.2.p02 from the following considerations:
  - It has better compatibility with Multithreading operation
  - A new Geometry package (VecGeom) can be integrated which provides better performance as well as robustness
  - Better tools for tracking (propagation in B-field and smarter approximations)

- Change of Geant4 version and Physics list needs more validation effort
  - Restore test beam simulation code within CMSSW (using the same noise model at all beam energies and beam types)
  - Restore isolated hadron analysis

# Integration of Geant4+VecGeom with CMSSW

- CMS has been interested in using the improved geometry classes from VecGeom integrated with Geant4 in CMS simulation software. Almost all the shapes required by CMS come from VecGeom. CMS participated in validating VecGeom and making it usable
- There have been a series of developments of VecGeom which have been tested using the CMSSW framework
  - The first results were reported using the 10.4.beta version integrated with the VecGeom version (v.00.04.00)
    - There was an issue due to segmentation violations after tracks with direction along z-axis getting stuck
      - This was due to inadequacy in one of the methods of finding distance to in/out
  - The first fix using the VecGeom branch raman/vecgeom-450 was tested also with 10.4.beta version of Geant4
    - There was no crash but several reports of stuck tracks for 2016 and 2018 geometries
      - Larger statistics of events are generated (3000 minimum bias and 1500 t-tbar) to see the type of volumes which cause this
      - The volumes concerned were only Polycone's and Polyhedra's
        - At least one of the concerned volume (mother or daughter) has one z-value where $R_{In} = R_{Out}$
  - A second set of corrections taken from the main branch (November 2) was tested with 10.3.Ref10 version of Geant4
    - This cured the stuck track issues for Polyhedra's
  - The branch raman/vecgeom-453 was tested with 10.3.Ref11 version of Geant4
    - This branch caused more stuck tracks and also segmentation violation due to lack of recovery from stuck tracks
  - The next set of corrections were taken from the main branch (November 29) and was tested with 10.4.Cand01 version of Geant4
  - The tagged version v00.05.00 is tested with the release version of Geant4.10.4

# Tests Performed

- Two different CMS geometries are used for testing CMSSW with Geant4-VecGeom
  - 2016 Geometry with 100k single particles (mixture of π/K/p) and their anti-particles with momentum between 1-20 GeV
  - 2018 Geometry for 8 different data sets
    - 3000 events of single muons in the barrel and similar number in the endcap
    - 3000 events of single electrons in the barrel and the same in the endcap
    - 3000 events of single pions in the barrel and the same in the endcap
    - 3000 minimum bias events at center-of-mass energy of 14 TeV
    - 1500 t-tbar events at 14 TeV

## Summary of Stuck Tracks in different versions

| | v00.04.00 | raman450 | Nov 2 | raman453 | Nov 29 | v00.05.00 |
|---|---|---|---|---|---|---|
| Barrel μ | 0/3000 | 0/3000 | 0/3000 | 65/3000 | 0/3000 | 0/3000 |
| Endcap μ | 0/3000 | 0/3000 | 0/3000 | 66/3000 | 0/3000 | 0/3000 |
| Barrel e | 0/3000 | 0/3000 | 0/3000 | 41/88 | 0/3000 | 0/3000 |
| Endcap e | 1/3000 | 1/3000 | 1/3000 | 151/3000 | 0/3000 | 0/3000 |
| Barrel π | 0/3000 | 0/3000 | 0/3000 | 733/3000 | 0/3000 | 0/3000 |
| Endcap π | 1/3000 | 1/3000 | 1/3000 | 54/1828 | 0/3000 | 0/3000 |
| MinBias | 29/500 | 194/3000 | 177/3000 | 78/7 | 1/3000 | 0/3000 |
| t-tbar | 39/110 | 413/1500 | 408/1500 | 53/1 | 0/1500 | 0/1500 |
| single had | 585/173k | 306/100k | 334/100k | 485/912 | 0/100k | 0/100k |

# Comparison between Native/VecGeom Versions

- "Stuck Track" warnings are not fatal if Geant4 can recover with a small push
- Both "Native" and "VecGeom" versions report "stuck track". The most recent "VecGeom" version provides acceptable rate.
  - "Single particle" sample uses 2016 Geometry while all others come from 2018 Geometry

| Version Used | Single Particle | 50 GeV pions in the barrel | Top Pair |
|---|---|---|---|
| Native | 1/100000 | 1/3000 | 2/1500 |
| VecGeom | 0/100000 | 0/3000 | 0/1500 |

- Volumes giving rise to "Stuck Track" warnings
  - Native version in 2018 Geometry:
    - PixelBarrelLayer0CoolantHalf: Tubs
    - PixelBarrelLayer3CoolantHalf: Tubs
    - ESPM: Subtraction Solid
  - Native version in 2016 Geometry:
    - ESPM: Subtraction Solid

# Performance of the Scalar Version

- The CPU and memory performance of CMSSW with Geant4.10.4 using "native" and "VecGeom" geometry versions:

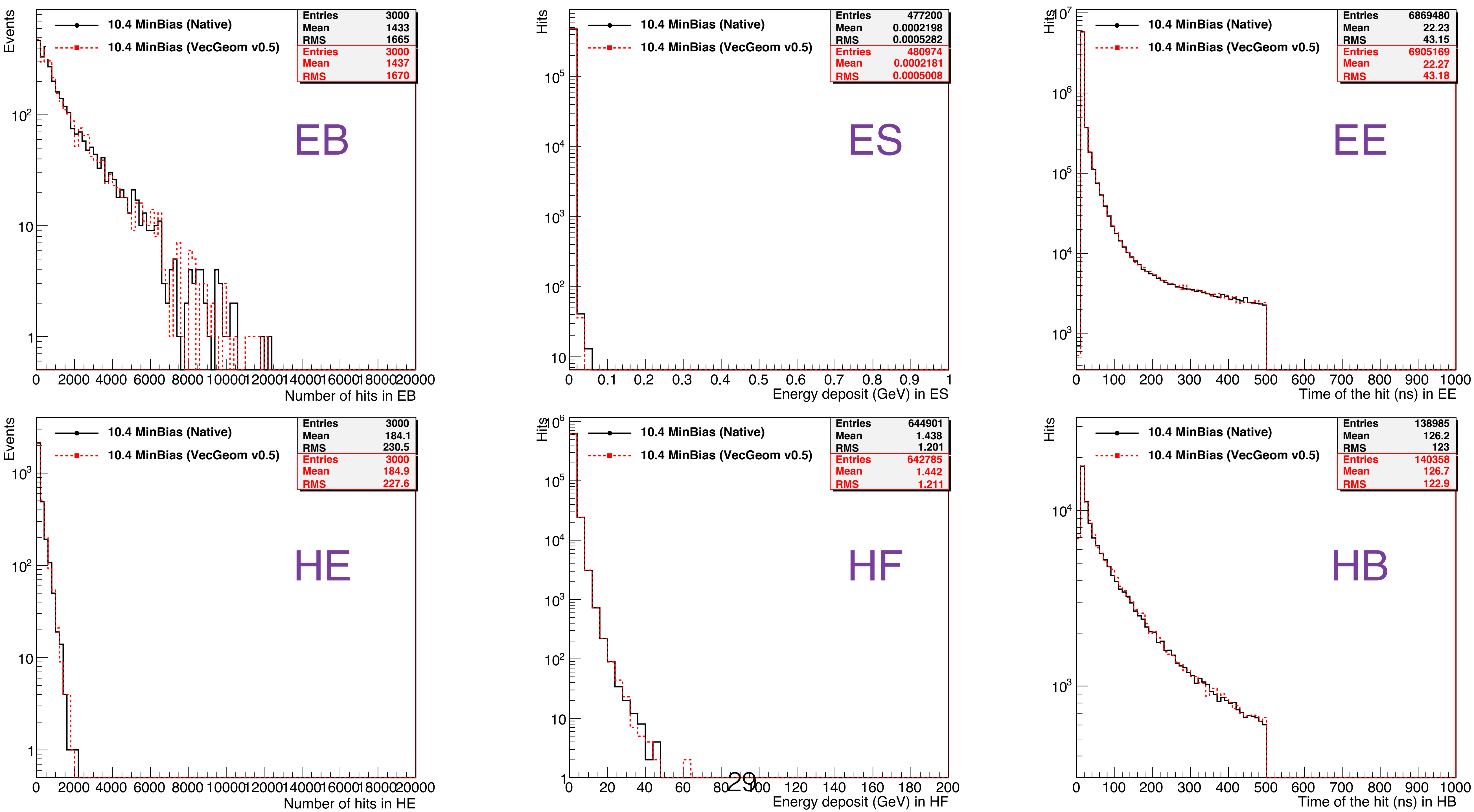
| | RSS (Native) (GB) | CPU (VecGeom) (wrt Native) | RSS (VecGeom) (GB) |
|---|---|---|---|
| Muon (Barrel) | 0.55 | 0.926 | 0.60 |
| Muon (Endcap) | 0.55 | 0.953 | 0.60 |
| Pion (Barrel) | 0.59 | 0.930 | 0.63 |
| Pion (Endcap) | 0.58 | 0.903 | 0.63 |
| Elec (Barrel) | 0.52 | 0.974 | 0.53 |
| Elec (Endcap) | 0.57 | 0.944 | 0.61 |
| Minimum Bias | 0.65 | 0.870 | 0.67 |
| t-tbar | 0.69 | 0.926 | 0.69 |
| Single Hadrons | 0.61 | 0.872 | 0.71 |

- 7-13% improvement in CPU performance with similar memory usage
- Usage of CLHEP version 2.4.0 does not change the CPU performance

# Quality of Simulation

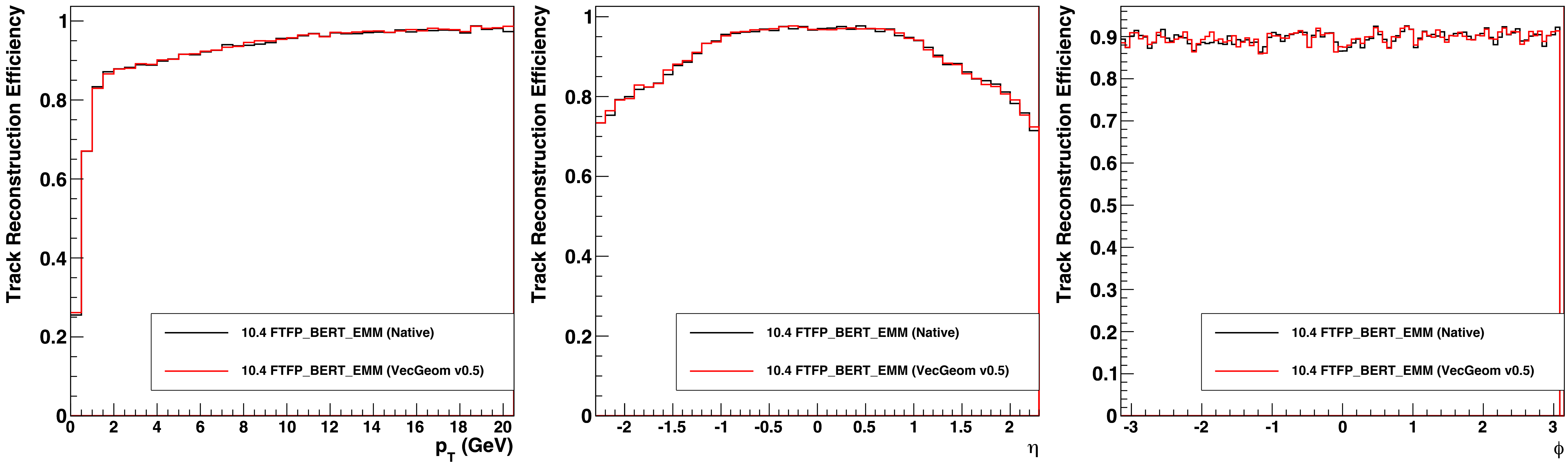
- Compared some of the basic calorimetric quantities in 3000 minimum bias events from
  - version using native Geant4 geometry routines
  - version using Geant4 + VecGeom package

Distributions from the two sets of simulation in terms of number of hits, energy deposit in calorimeter, hit timing match rather well



# Track Reconstruction Efficiency

- Generate single particle events with a mixture of pions, kaons, protons and anti-protons
  - Do a complete simulation going all the way up to reconstruction
  - Study track reconstruction efficiency as a function of p, $p_T$, η and φ




- There is no difference observed among the versions using native Geant4 geometry package or using the VecGeom package

# Further Improvements in Simulation

- With the Geant4.10.4 version there could be 2 possible additional improvements in the SIM step:
  - Use of a new stepper G4DormandPrince745 instead of G4ClassicalRK4
  - Use of "Smart Track" idea - namely treating charged particles below a certain energy with less precision during tracking in B-field
- Both these options were available (not as default) in CMSSW since version beginning of 2018
- Some of these new features are tested and compared with the default in CMS for
  - Computing performance
  - Physics performance
- Four different scenarios are used
  - Default (stepper: G4ClassicalRK4 and energy Threshold: 0 MeV)
  - New Stepper (stepper: G4DormandPrince745 and energy Threshold: 0 MeV)
  - Smart Track (stepper: G4ClassicalRK4 and energy Threshold: 2 MeV)
  - Smart Track with new Stepper (stepper: G4DormandPrince745 and energy threshold: 2 MeV)

# Performance Comparison

- The CPU and memory performance of CMSSW with the 4 different versions:

| | RSS (Default) (GB) | CPU (Stepper) (wrt default) | RSS (Stepper) (GB) | CPU (Smart) (wrt default) | RSS (Smart) (GB) | CPU (Smart +Stepper) (wrt default) | RSS (Smart +Stepper) (GB) |
|---|---|---|---|---|---|---|---|
| Muon (Barrel) | 0.59 | 0.973 | 0.54 | 0.986 | 0.53 | 0.932 | 0.53 |
| Muon (Endcap) | 0.52 | 1.000 | 0.53 | 0.978 | 0.54 | 0.923 | 0.60 |
| Pion (Barrel) | 0.55 | 1.028 | 0.55 | 1.004 | 0.56 | 0.992 | 0.62 |
| Pion (Endcap) | 0.62 | 1.035 | 0.56 | 0.981 | 0.55 | 0.979 | 0.55 |
| Elec (Barrel) | 0.52 | 1.005 | 0.52 | 0.977 | 0.54 | 1.003 | 0.59 |
| Elec (Endcap) | 0.54 | 1.017 | 0.53 | 0.977 | 0.53 | 0.984 | 0.59 |
| Minimum Bias | 0.68 | 1.006 | 0.69 | 0.982 | 0.70 | 0.940 | 0.74 |
| t-tbar | 0.69 | 1.013 | 0.70 | 0.982 | 0.69 | 0.952 | 0.70 |

Possible improvement in performance by 5-6%

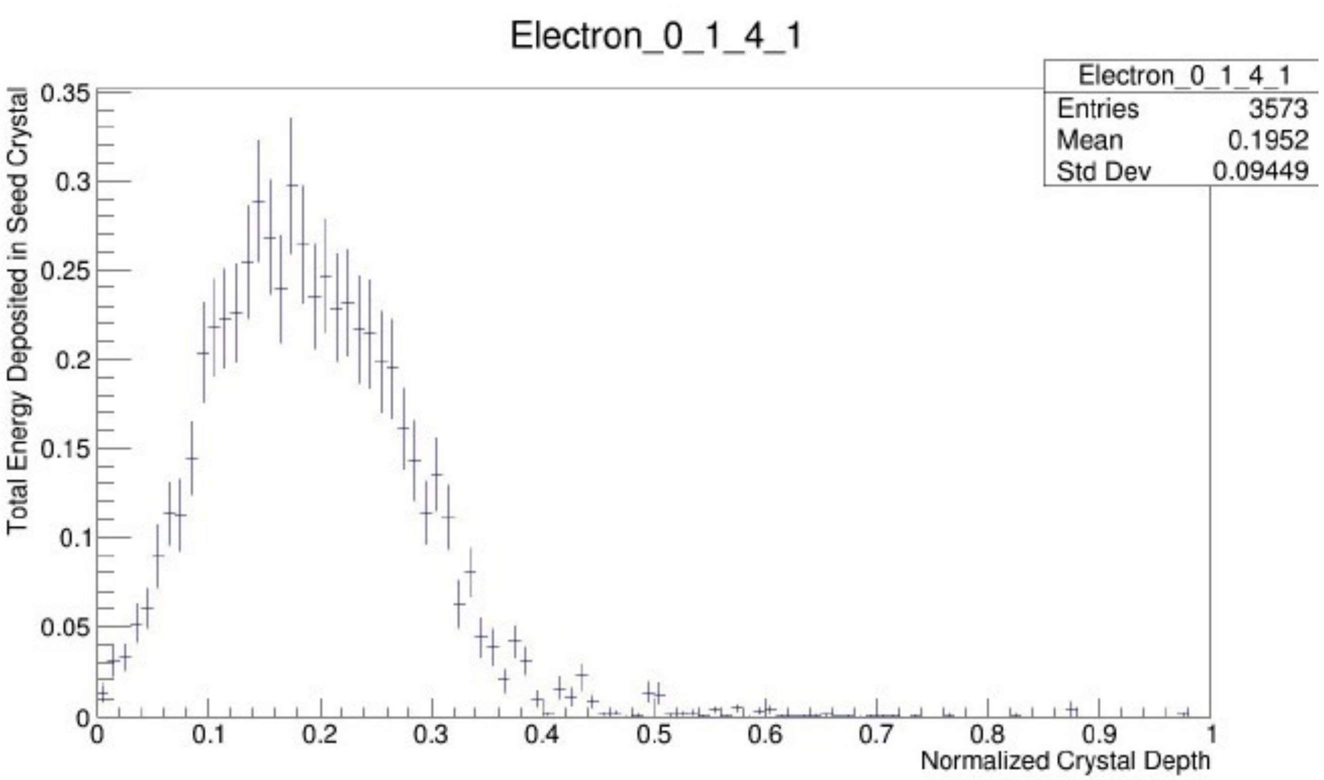
# Robustness Tests

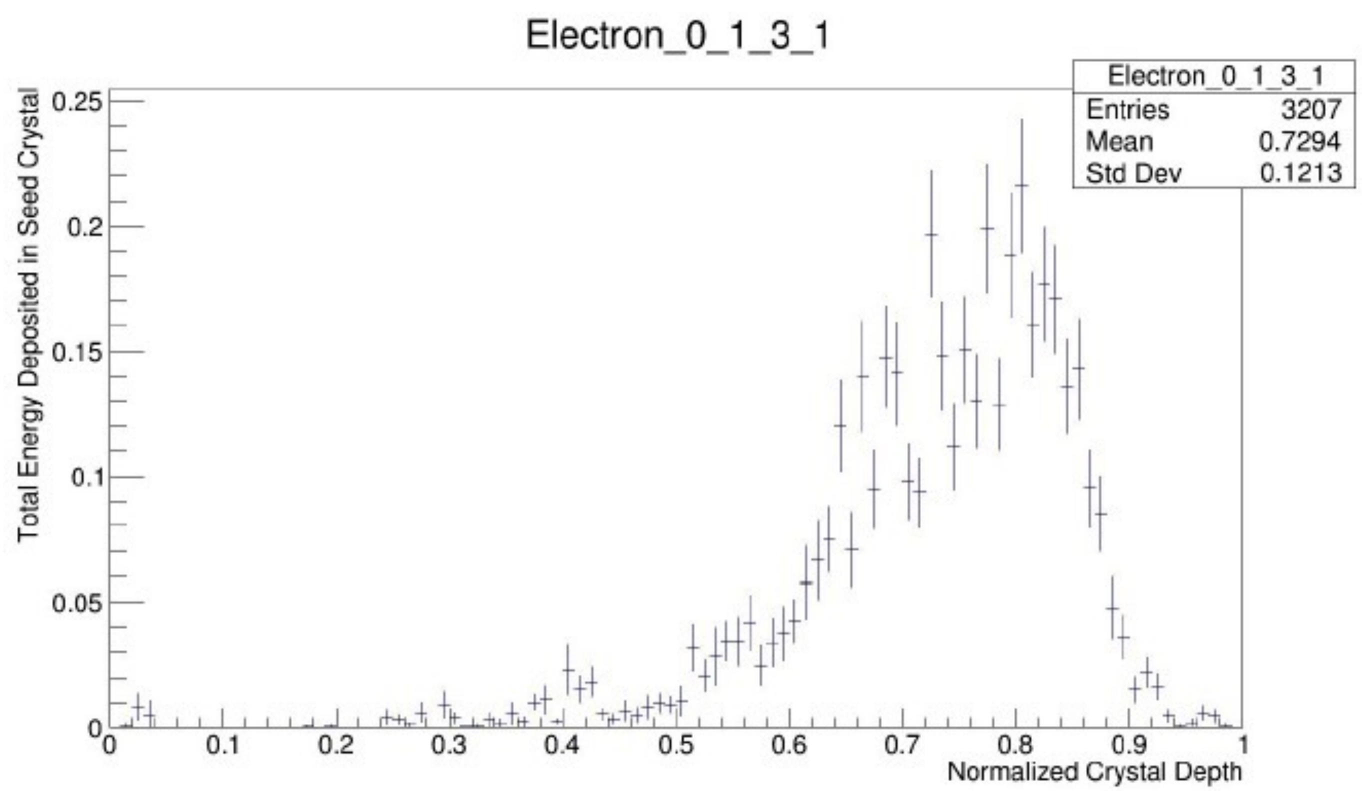
- There are 2 types of warning messages which appear during tracking
  - "Track stuck, not moving for 10 steps". Extra push is provided to recover the tracking failure - could be due to lack of accuracy or overlap in geometry definition
  - "G4MagInt_Driver::AccurateAdvance()" - happens during tracking in B-field

| | Stuck Track Default | Stuck Track Stepper | Stuck Track Smart | Stuck Track Stepper +Smart | FieldMan. Default | FieldMan. Stepper | FieldMan. Smart | FieldMan. Stepper +Smart |
|---|---|---|---|---|---|---|---|---|
| Muon (Barrel) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Muon (Endcap) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Pion (Barrel) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Pion (Endcap) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Elec (Barrel) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Elec (Endcap) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Minimum Bias | 0 | 0 | 0 | 0 | 4 | 0 | 1 | 0 |
| t-tbar | 0 | 0 | 0 | 0 | 8 | 0 | 5 | 0 |

# Some Bugs Live Long

- Study of radiation damage in ECAL required to look into the longitudinal shower development in the crystals
- There is a facility to provide the shower depth in units of radiation length as the depth index of PCaloHit
- The current implementation of shower depth calculation showed peculiar profiles for 2 different sets of crystals for electrons and photons



# How Depth is Computed

- The crystals are defined as trapezoid with the crystal axis along z-axis
- Computation of depth involves
  - Transform the hit point to the local frame of reference
  - Add the crystal length to the z-coordinate

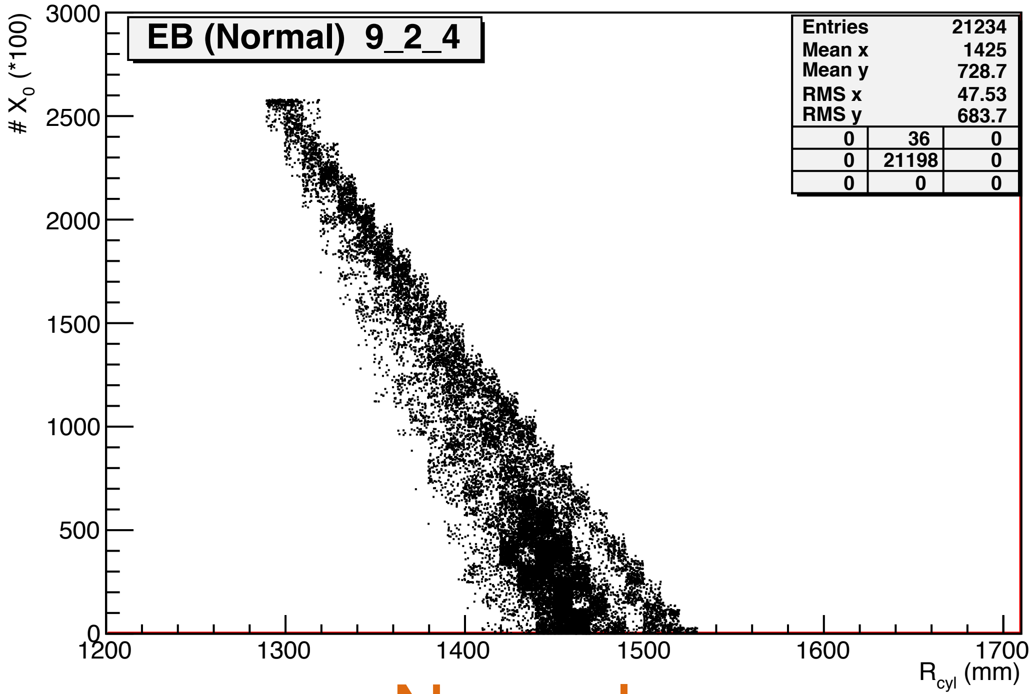
```
G4LogicalVolume* lv = stepPoint->GetTouchable()->GetVolume(0)->GetLogicalVolume();
G4ThreeVector localPoint = setToLocal(stepPoint->GetPosition(),stepPoint->GetTouchable());
double crlength = crystalLength(lv);
double dapd = 0.5 * crlength - localPoint.z();
if (dapd <= 100.) weight = 1.0 + slopeLY - dapd * 0.01 * slopeLY;
else              weight = 1.0;

double depth = 0.5 * crlength + localPoint.z();
```

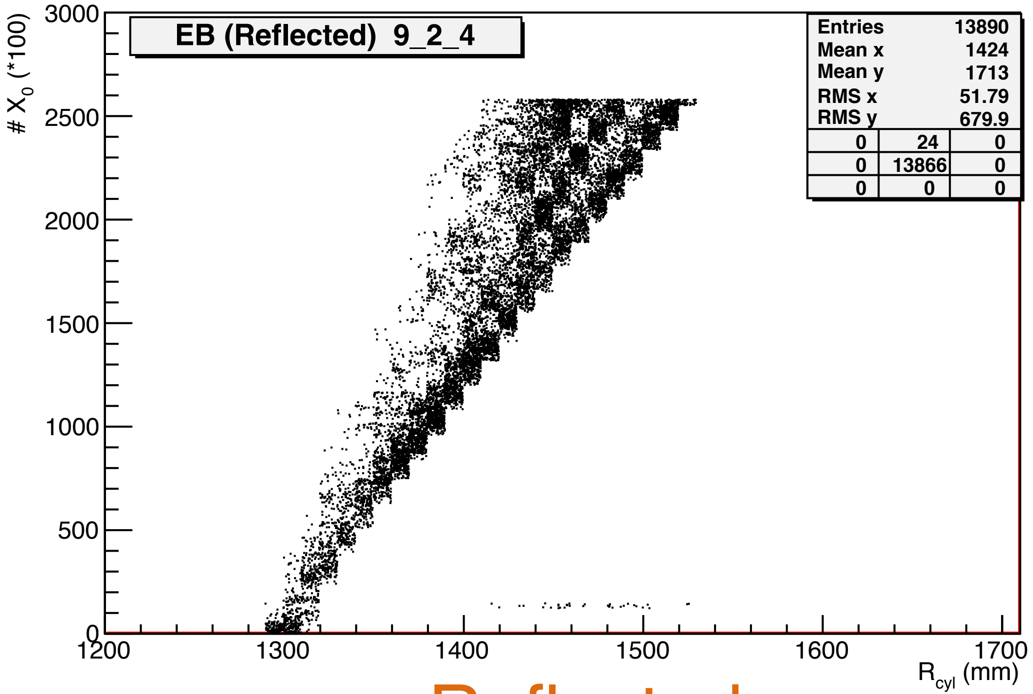
- The same calculation is used to compute distance of the hit point from the optical transducer. This helps in estimating non-uniformity of light collection efficiency
- In the simulation of fast timing option for PhaseII studies, the method getLayerIDForTimeSim(…) is invoked - but there the calculation is done by testing the occurrence of “refl” in the name of the volume

# Depth Computation in the Buggy Version

- Generate di-muon events of fixed $p_T$ in the region $|\eta| < 3.0$
- Compute depth in units of $X_0$ and plot it as a function of
  - $R_{cyl}$ for crystals in the barrel region
  - $|z|$ for crystals in the endcap region

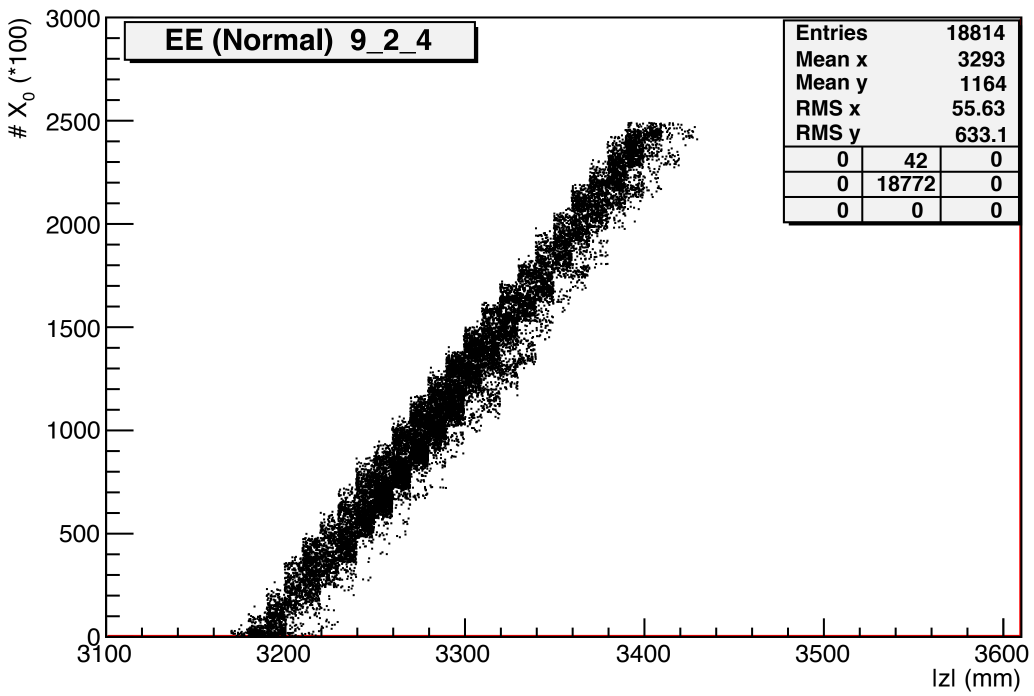



EB

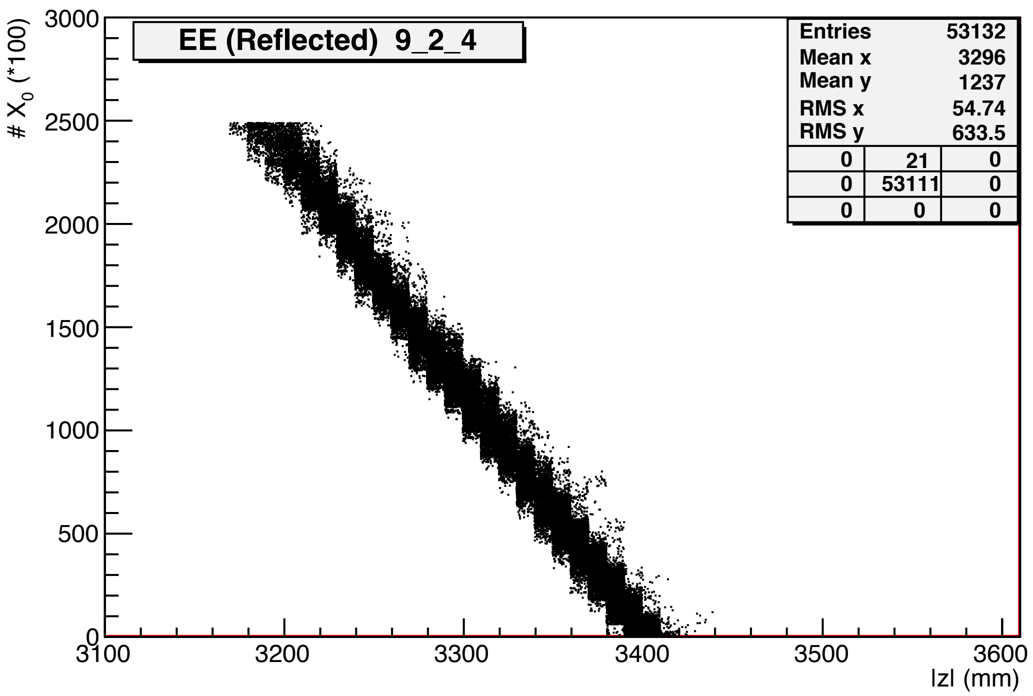



Normal

Reflected

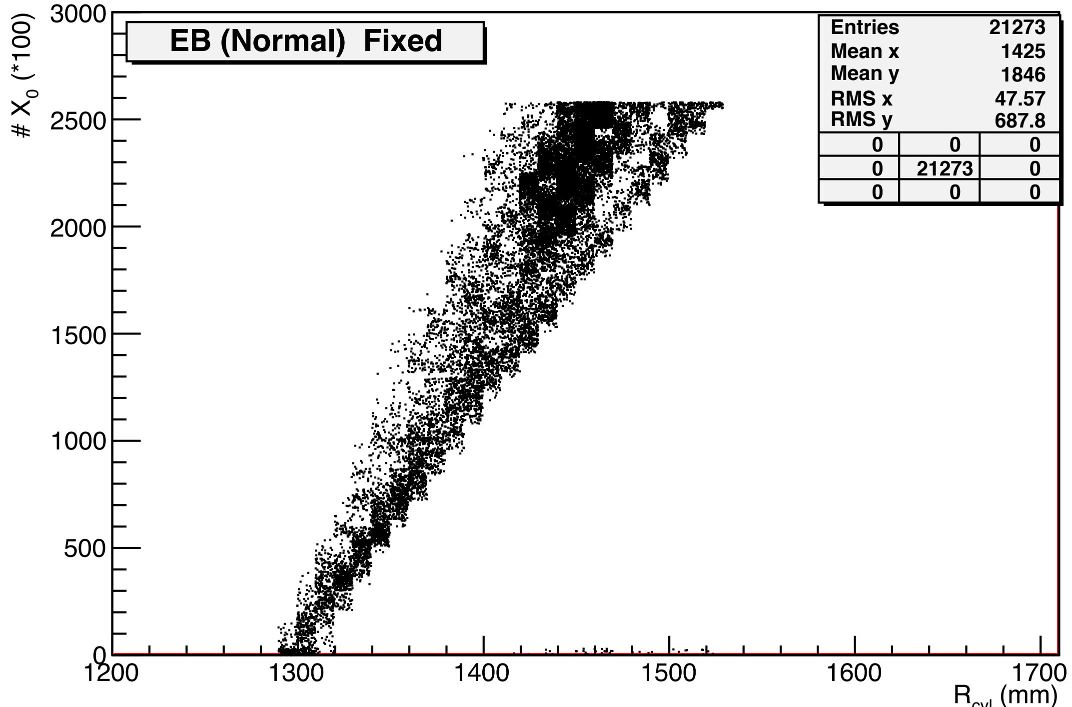



EE

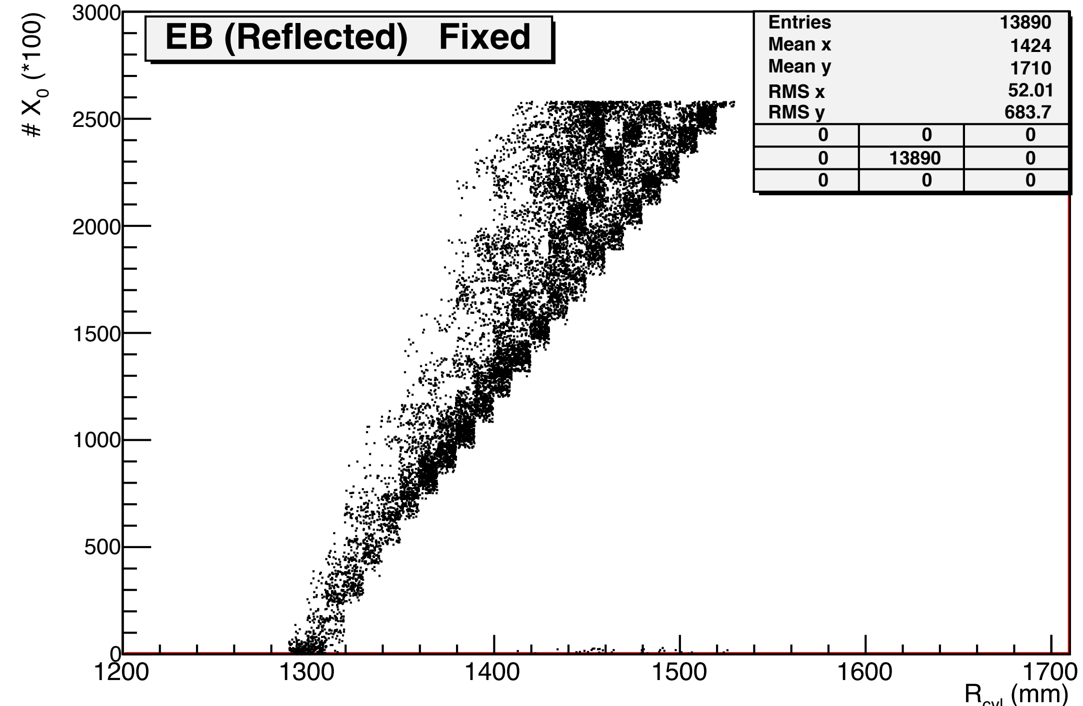


# Fix Cures the Issue

- Treat the normal and “refl” volumes differently both for EB and EE

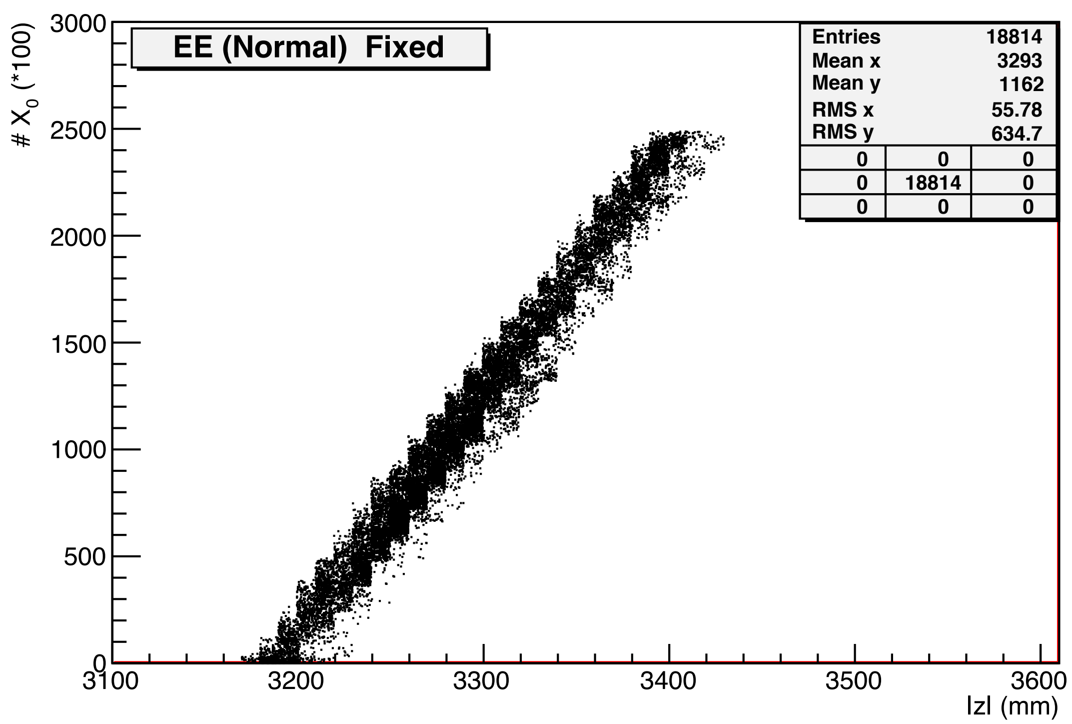



EB

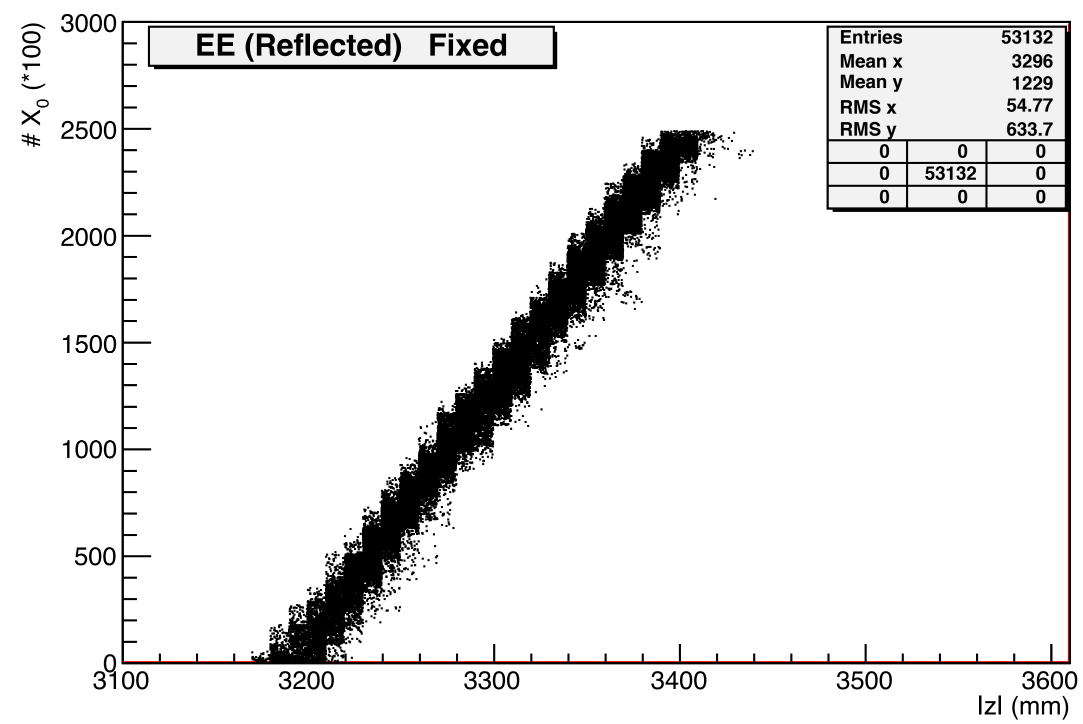



Normal

Reflected



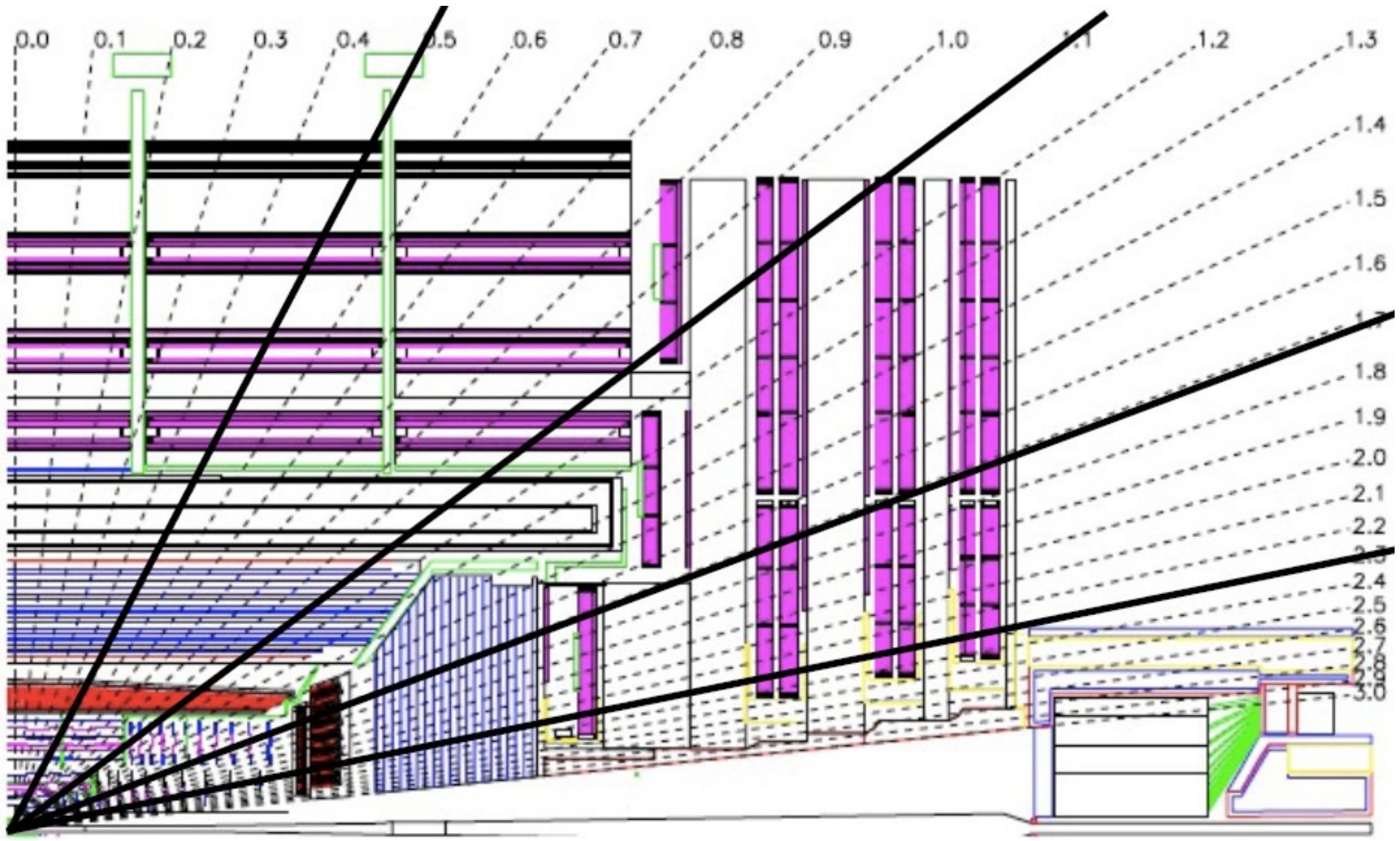

EE

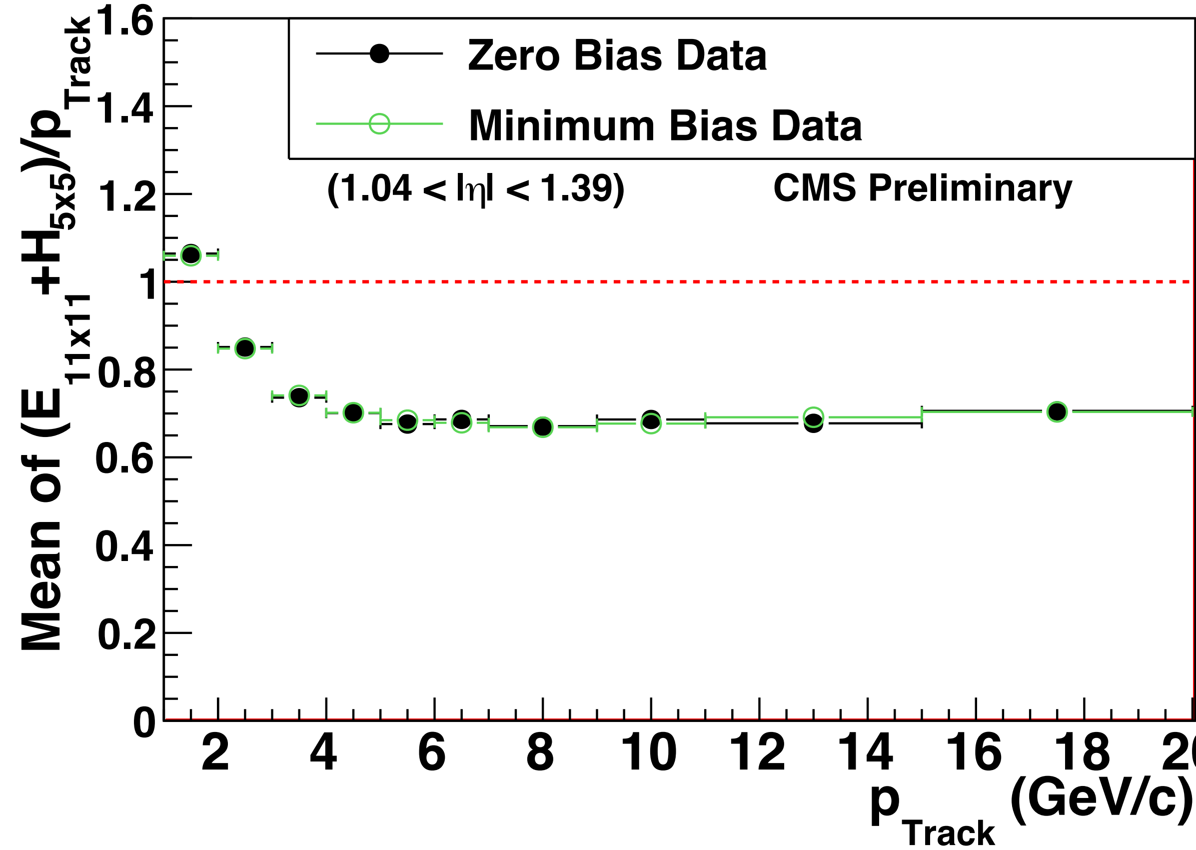


# What is Done for Physics Validation

- Compare ratio of calorimeter energy measurement to beam or track momentum between data and MC. Data come from
  - test beam studies with identified particle type
  - isolated charged hadron sample in collision data
- Source of Data:
  - 2006 test beam set up in the SPS H2 beam line with HB prototype and one EB supermodule
  - Low luminosity runs taken during 2016B run period using Zero Bias and Minimum Bias triggers
- For Monte Carlo, events are generated using FTFP_BERT_EMM Physics List for Geant4 versions 10.2.p02, (10.3.p03, 10.4.beta), and 10.4 with CMSSW_9_2_5_patch1. All the non-standard Geant4 versions are custom made with the same same CLHEP etc as used in 9_2_5_patch1:
  - Generate 50k events at each beam energy for the said type and for calibration generate 50k electron events in setups with and without EB
  - Generate 100k single particle event sample using a flat energy distribution between 1 and 20 GeV with a given admixture of pions, kaons and protons and anti-protons (as expected in minimum bias sample)

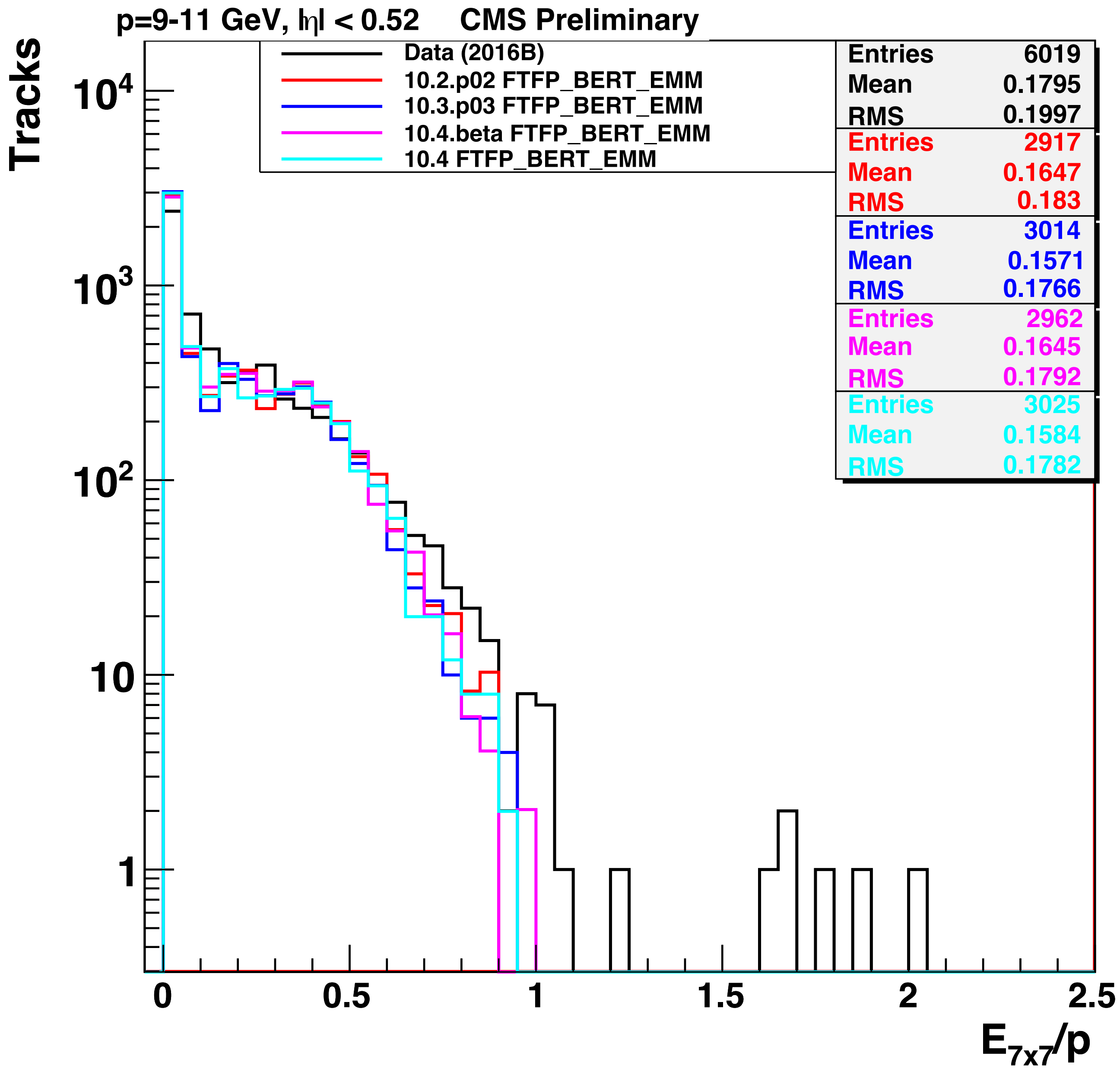
# Energy Measurements

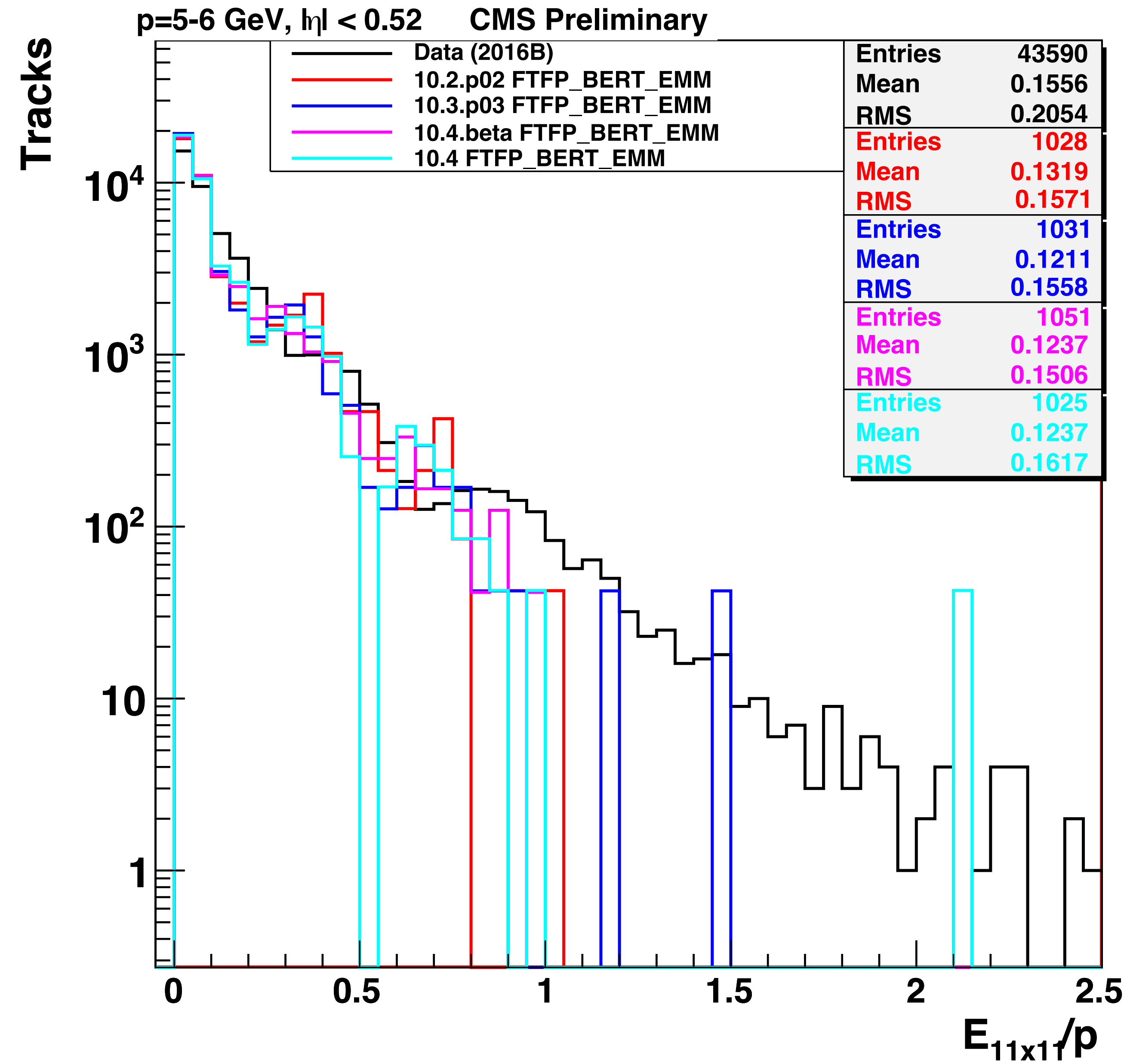





- Select good charged tracks using standard cuts
- Propagate them to calorimeter surface and select those which are well isolated from other charged or neutral particles in the calorimeter surface
- Measure energy by combining energy measurements from a matrix of NxN cells around the cell hit by the extrapolated track to the calorimeter surface
- For the data use the two low luminosity data sets from the 2016B runs
  - The distributions from Zero Bias and minimum bias data agree quite well
  - For comparison with Monte Carlo these two data sets are combined

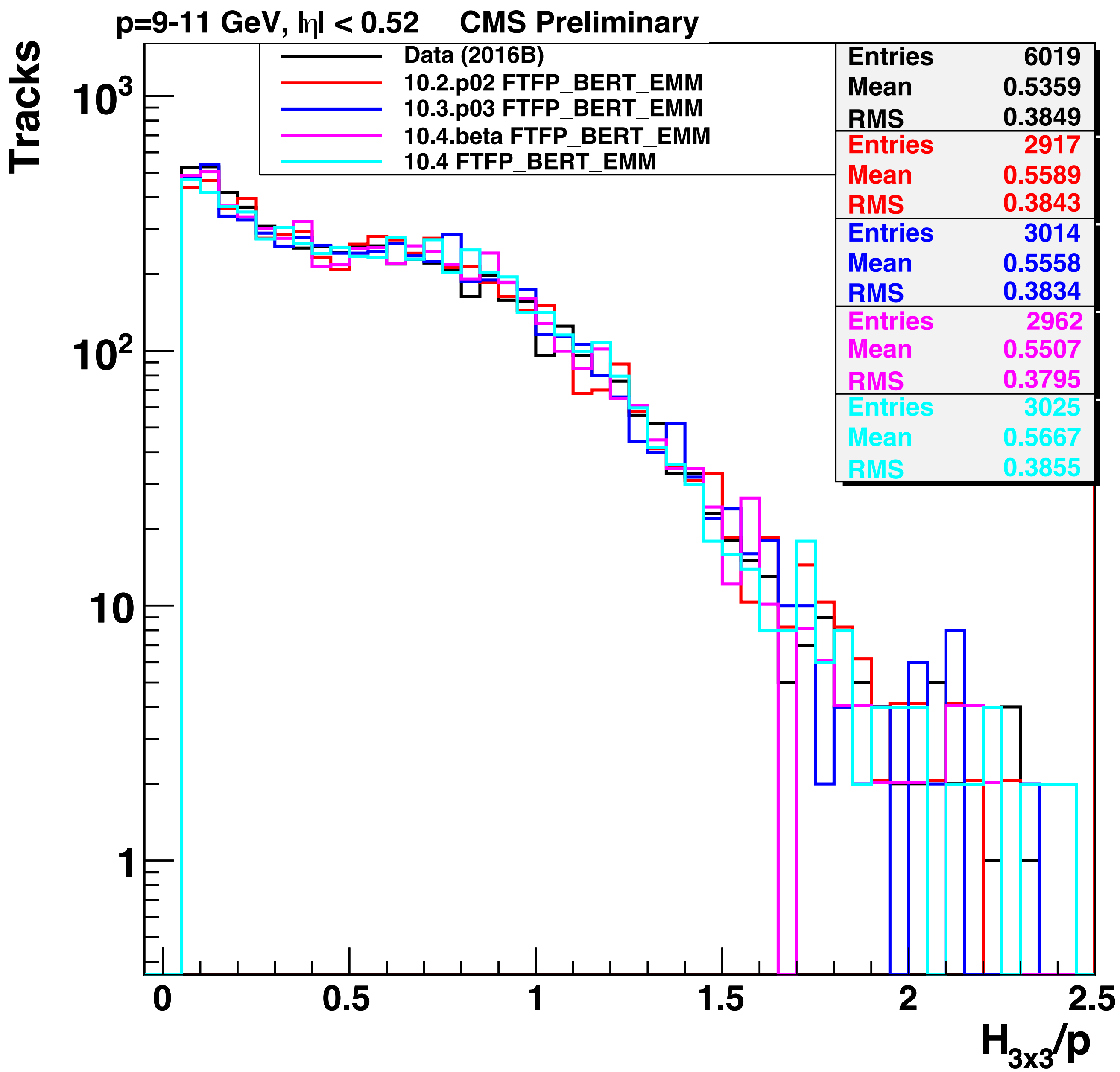
# Energy in ECAL

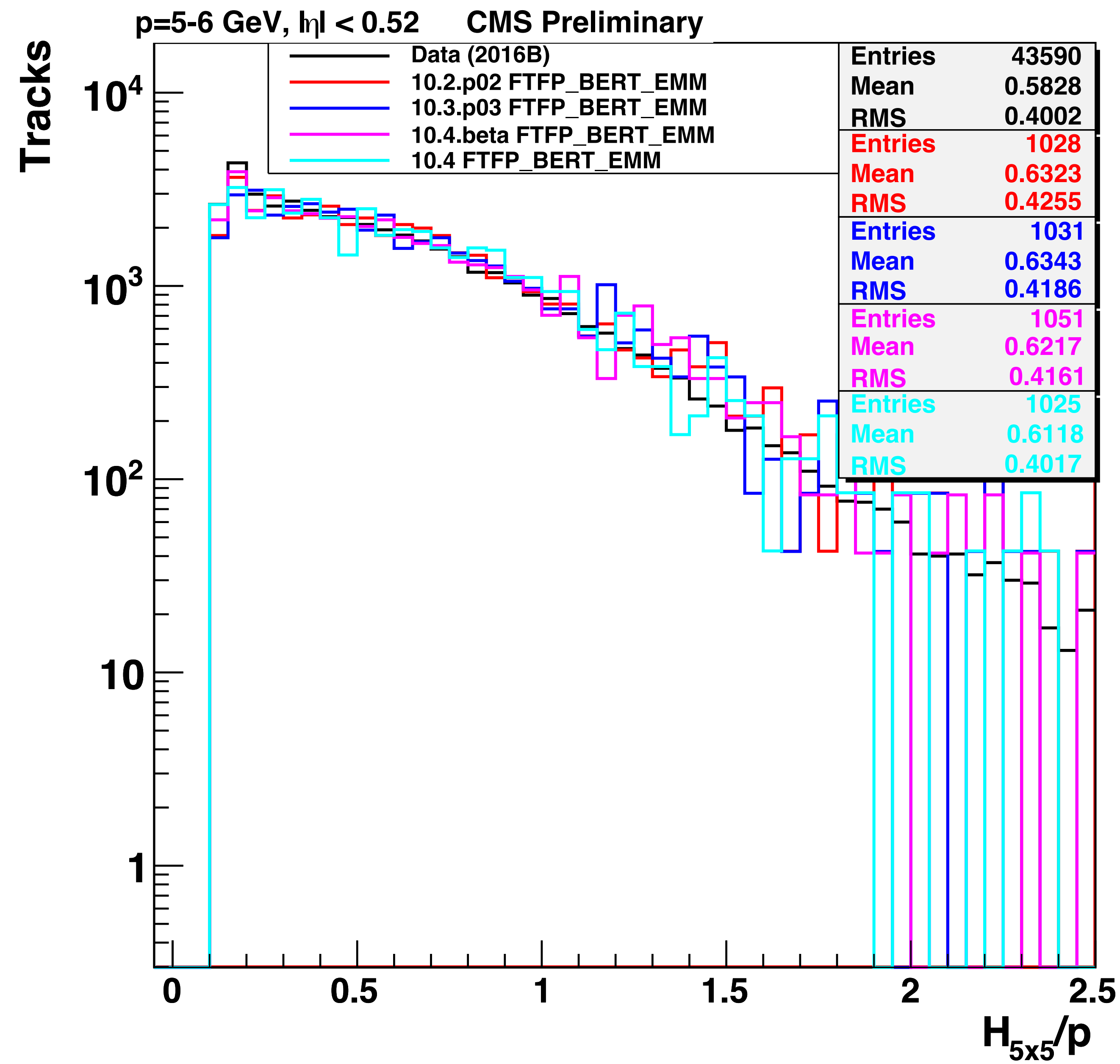



- Energy fraction in ECAL for narrow (7x7) or wide (11x11) matrix agrees reasonably between data and MC. The disagreement in the tail is partially due to limited statistics in the MC sample

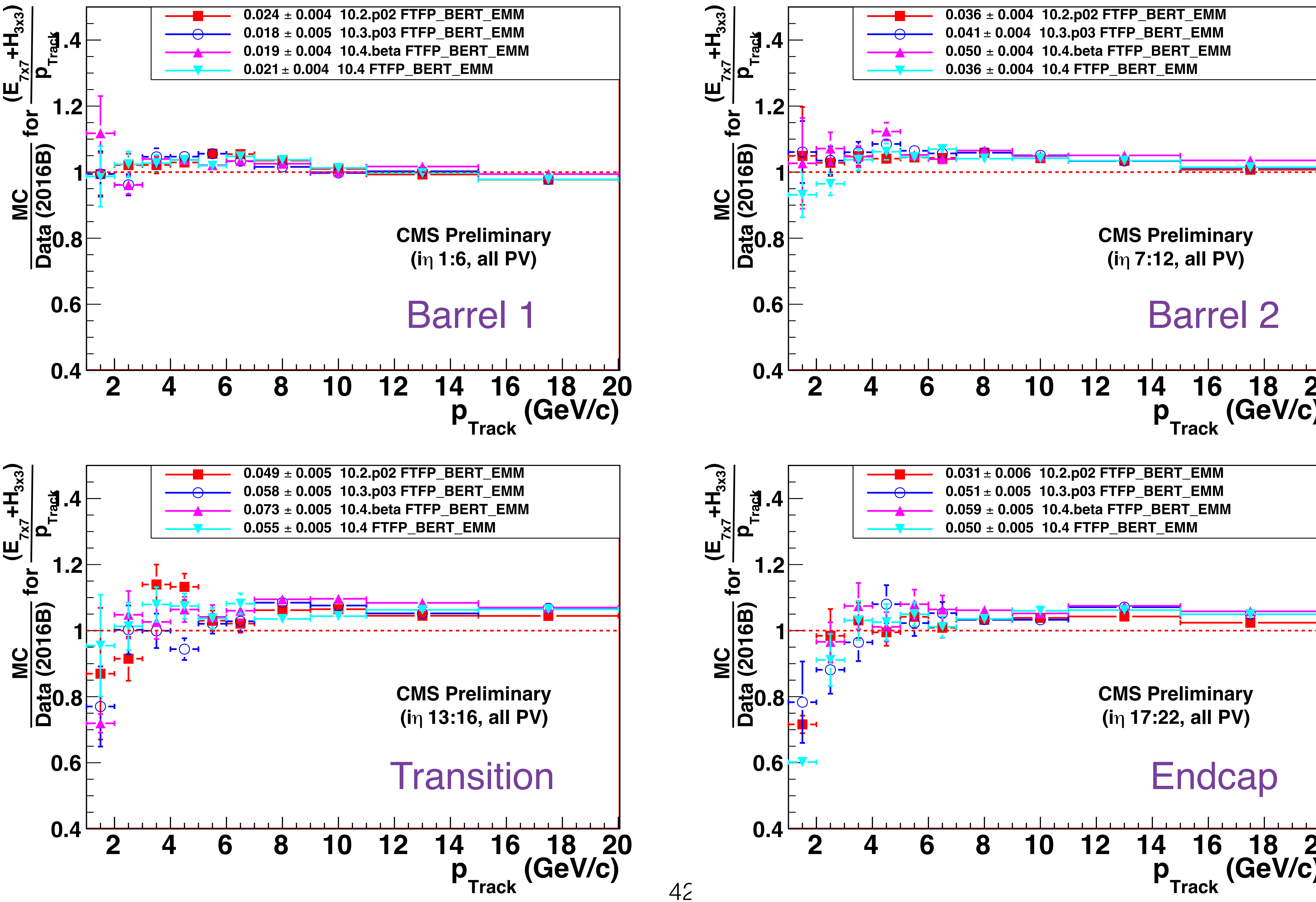
# Energy in HCAL




- Fairly good agreement between data and MC for energy measured in the HCAL for narrow (3x3) as well as wide (5x5) matrix

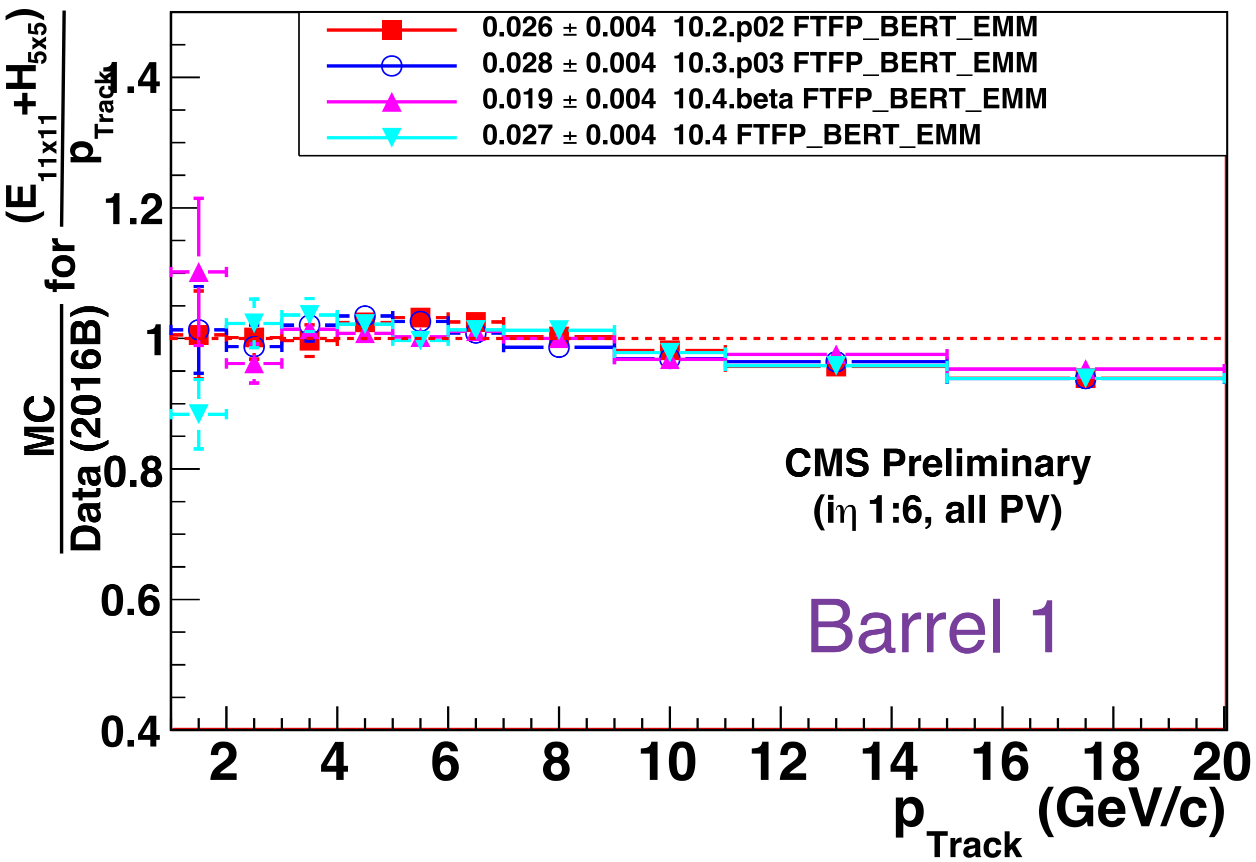
# Combined Calorimeter Energy (7x7+3x3 matrix)

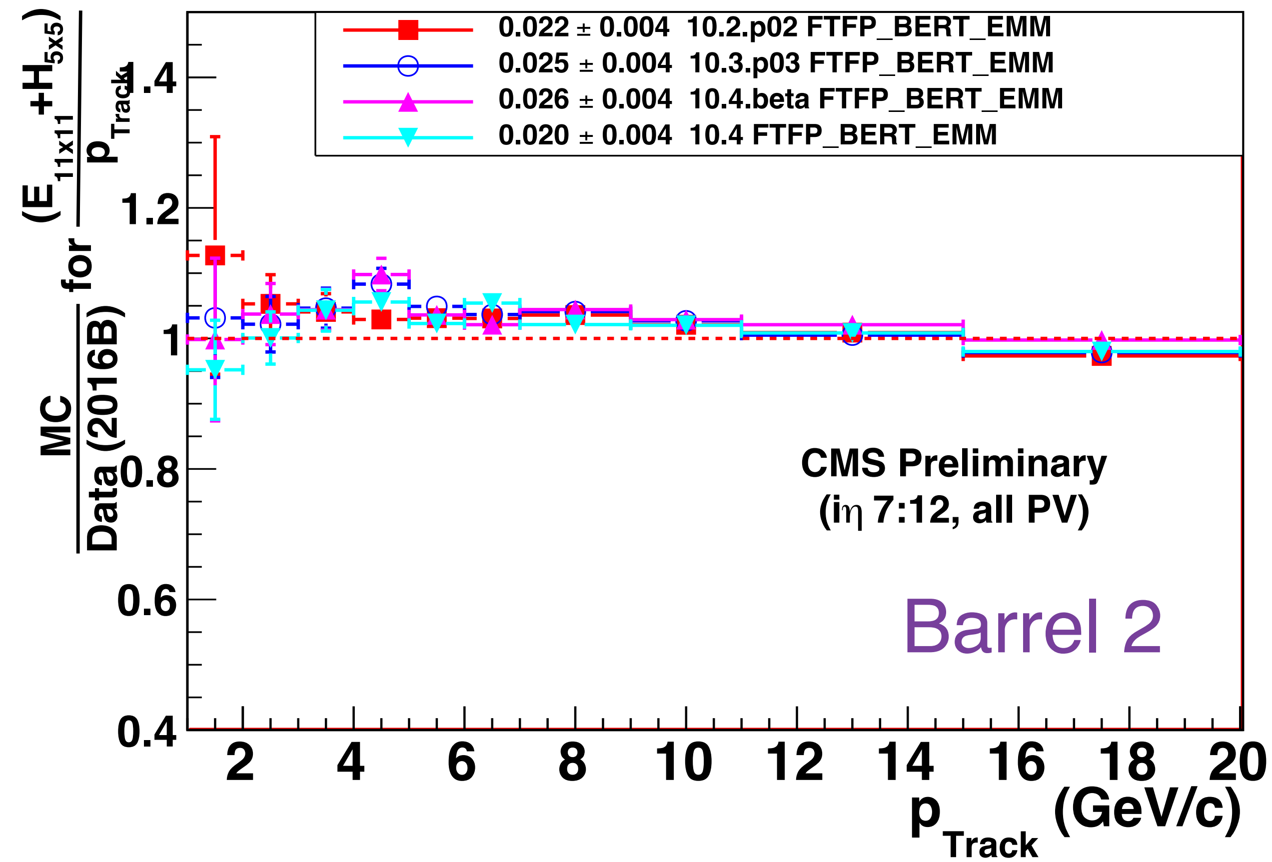


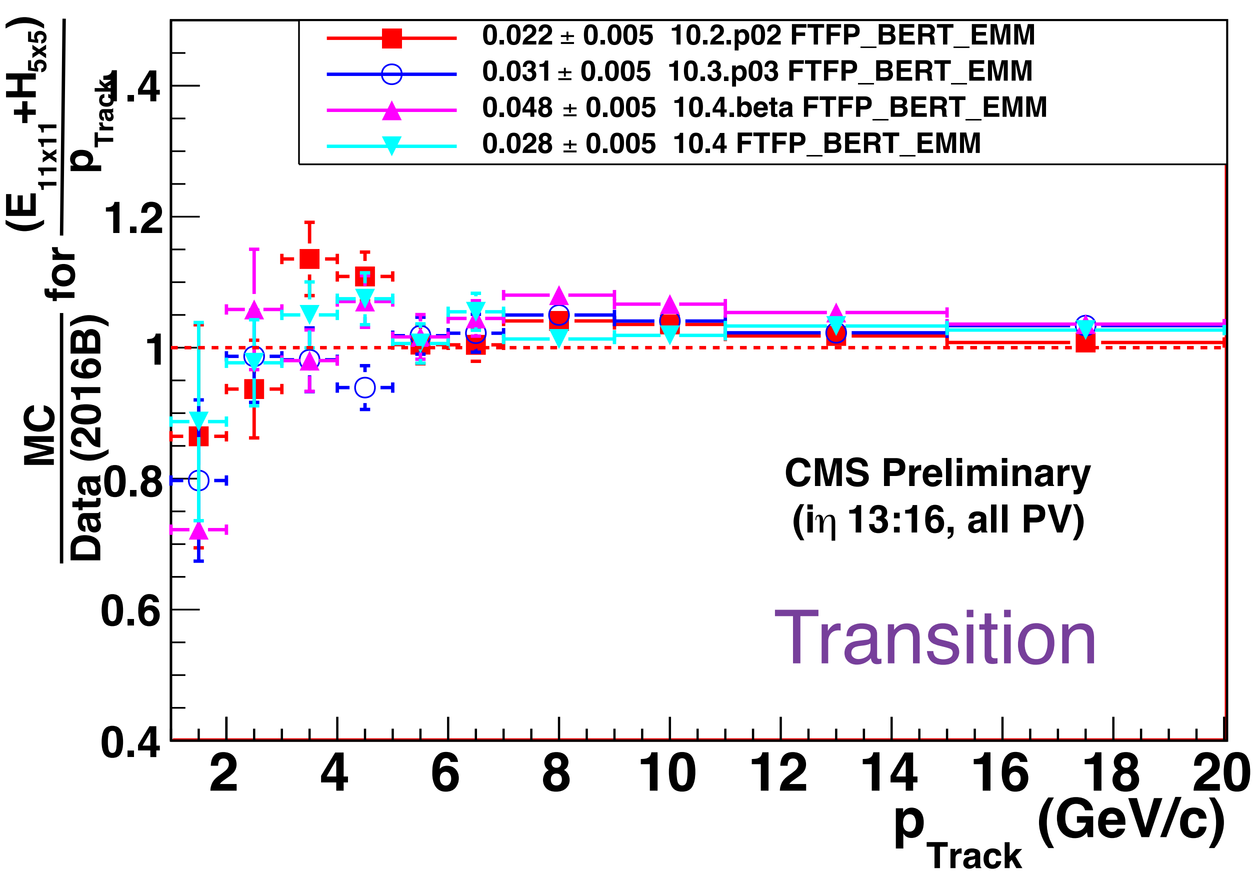


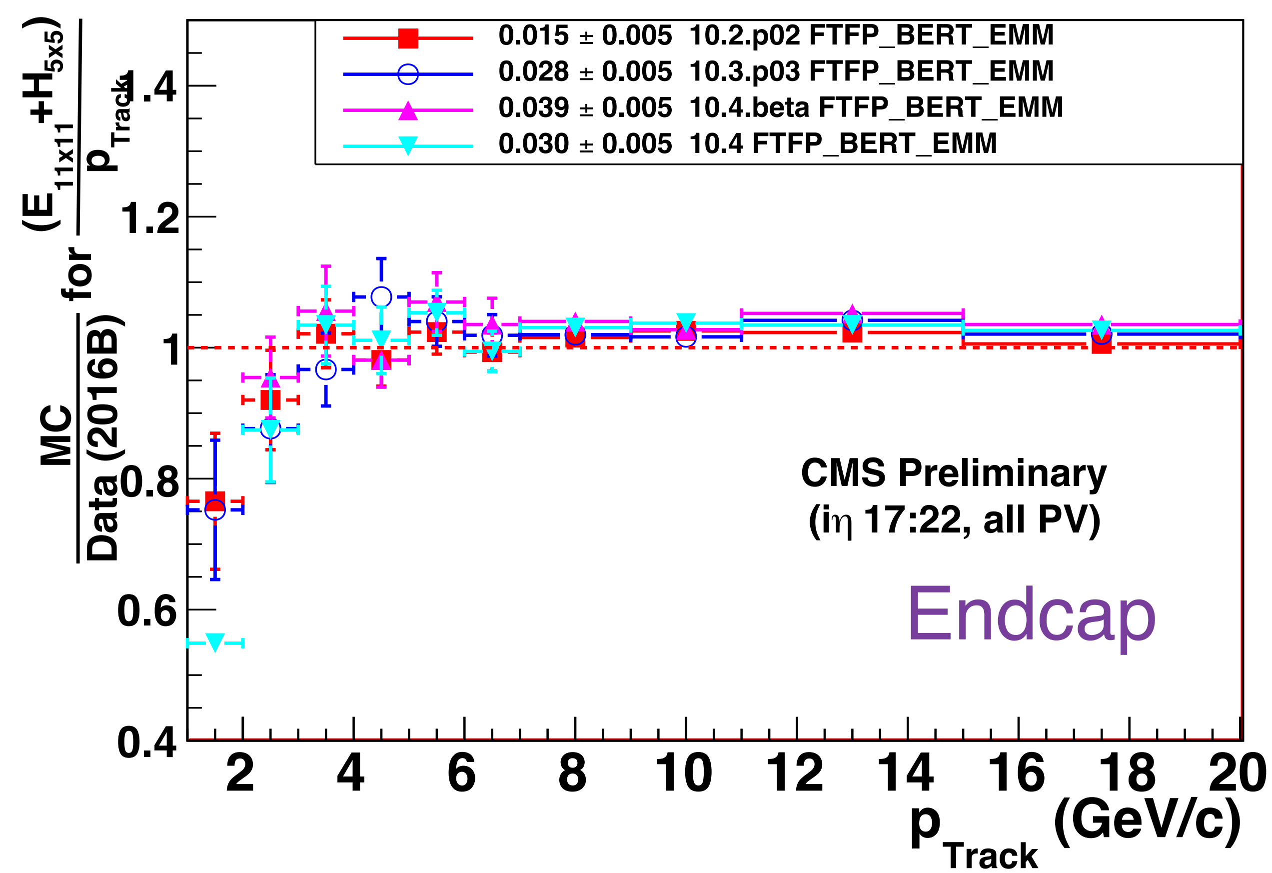

jee

# Combined Calorimeter Energy (11x11+5x5 matrix)



# Collision Data

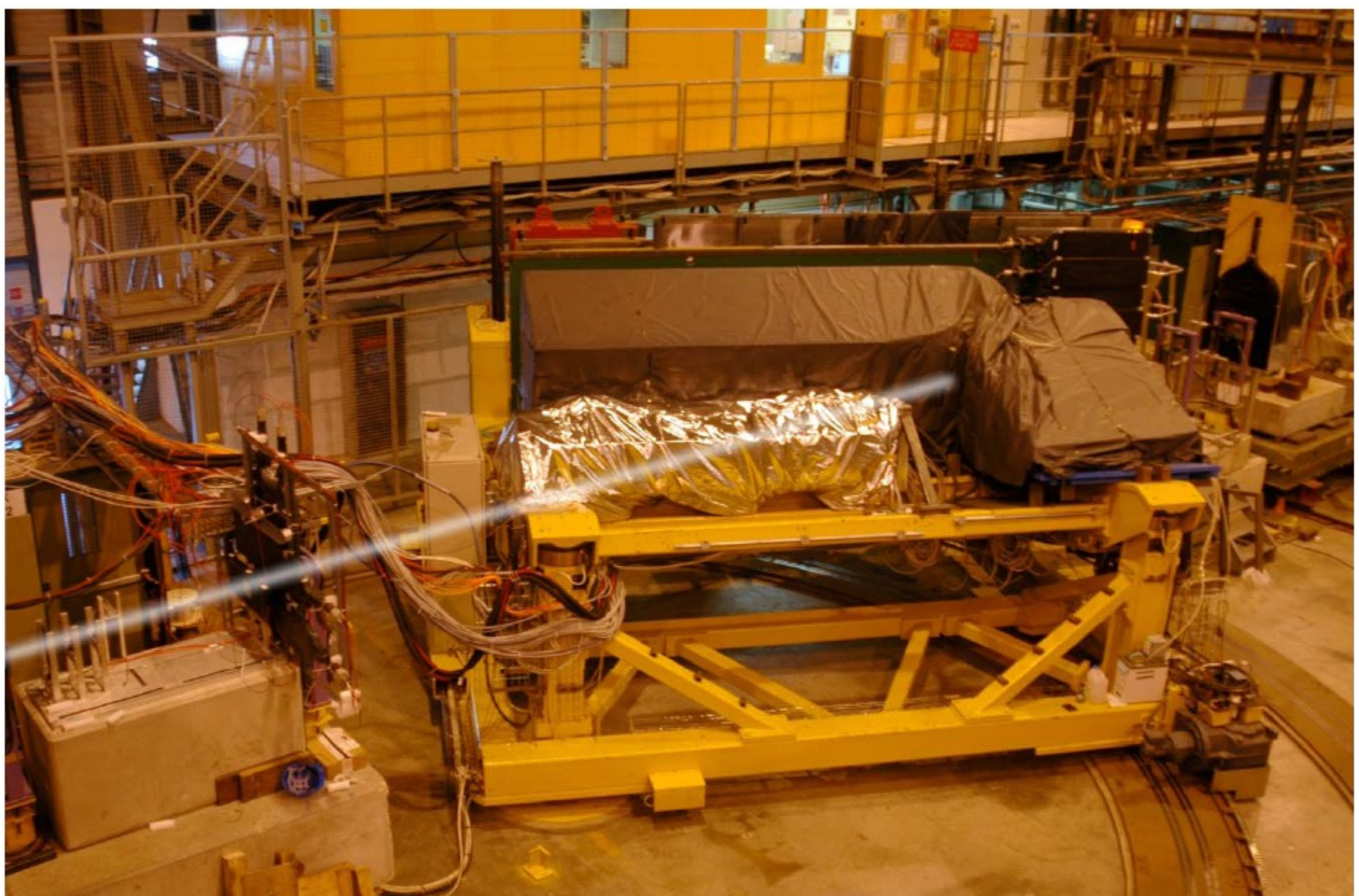
- The level of disagreement between data and MC is between 2.0% to 5.5% for the Geant4 version 10.4 depending on the region of the detector

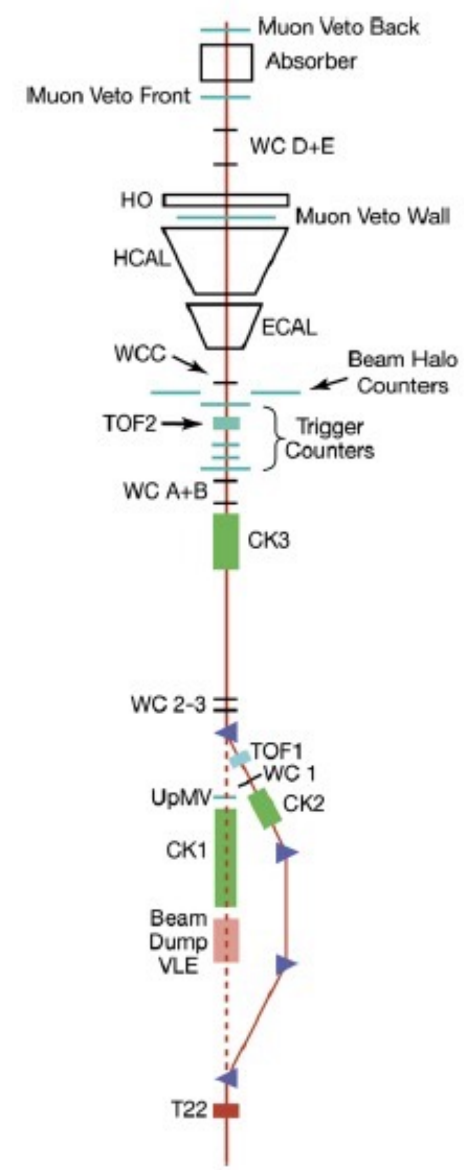
Mean level of disagreement between MC and data

| | $(E_{7x7}+H_{3x3})/p$ 10.2.p02 | $(E_{7x7}+H_{3x3})/p$ 10.4.beta | $(E_{7x7}+H_{3x3})/p$ 10.4 | $(E_{11x11}+H_{5x5})/p$ 10.2.p02 | $(E_{11x11}+H_{5x5})/p$ 10.4.beta | $(E_{11x11}+H_{5x5})/p$ 10.4 |
|---|---|---|---|---|---|---|
| **Barrel 1** | (2.3±0.4)% | (1.9±0.4)% | (2.1±0.4)% | (2.6±0.4)% | (1.9±0.4)% | (2.7±0.4)% |
| **Barrel 2** | (3.6±0.4)% | (5.0±0.4)% | (3.6±0.4)% | (2.2±0.4)% | (2.6±0.4)% | (2.0±0.4)% |
| **Transition** | (4.9±0.5)% | (7.2±0.5)% | (5.5±0.5)% | (2.2±0.5)% | (4.8±0.5)% | (2.8±0.5)% |
| **Endcap** | (3.1±0.3)% | (5.9±0.5)% | (5.0±0.5)% | (1.5±0.5)% | (3.9±0.5)% | (3.0±0.5)% |

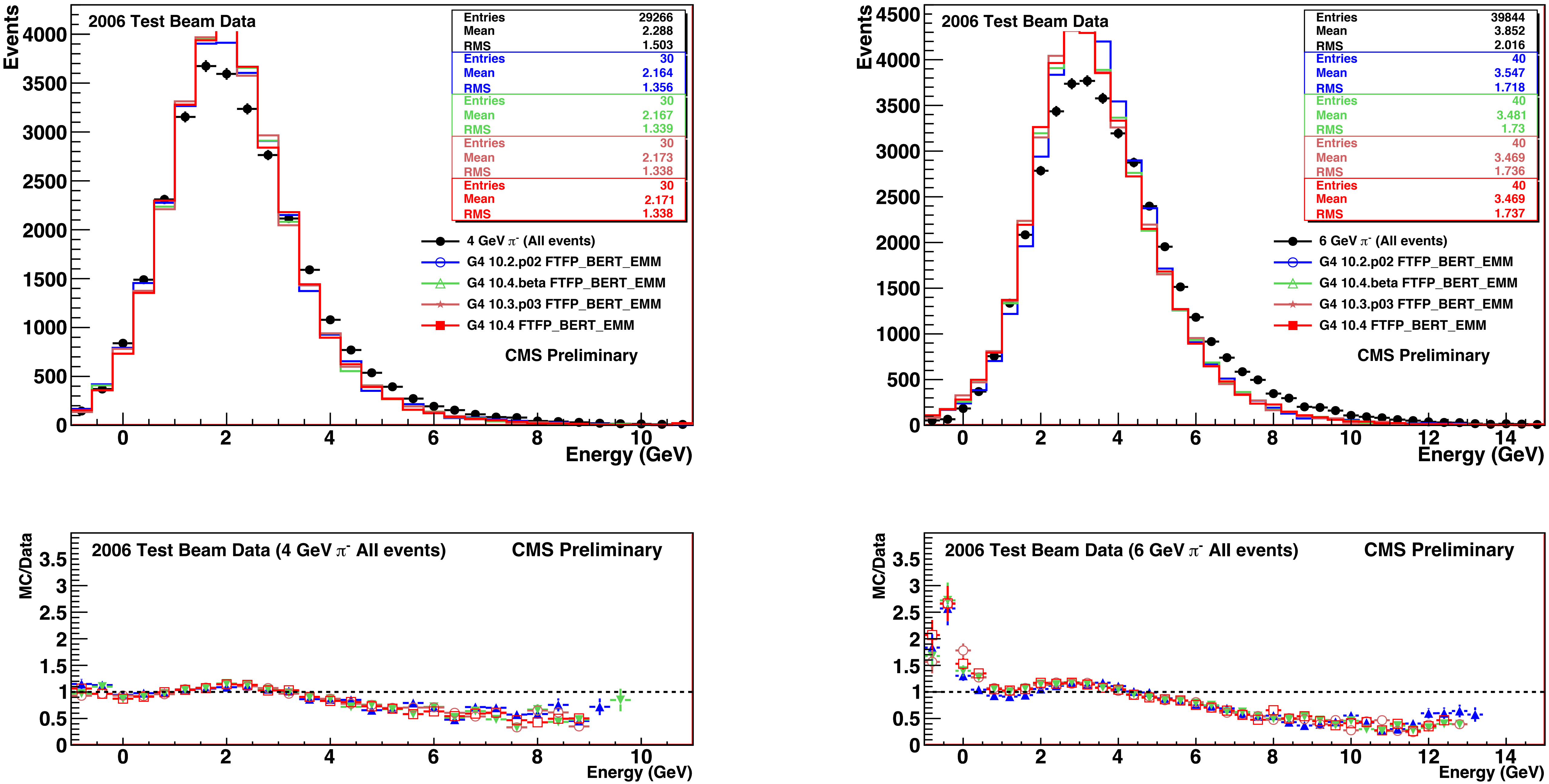
# 2006 TestBeam Data

- The data correspond to energy response due to well identified single particles over a large momentum range (2 to 350 GeV)
- The results consist of the energy distributions for well identified particles at a fixed momentum, mean response and energy resolution as a function of beam momentum and fraction of events not interacting in the ECAL
  - Particle identification is rather good for beam momenta at or below 9 GeV
- Use the setup described within CMSSW to simulate events with single particles.



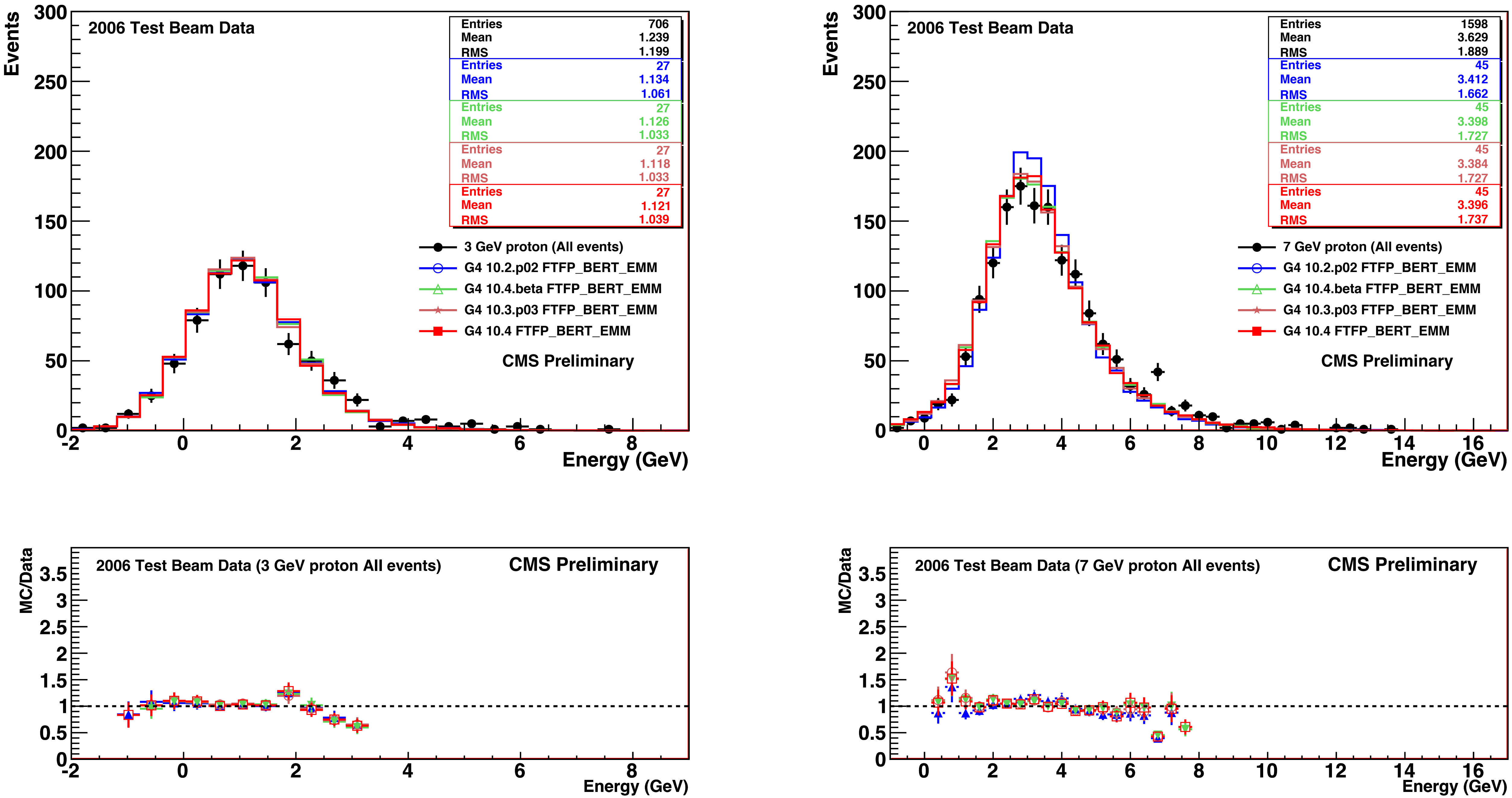


# Energy for negative pions



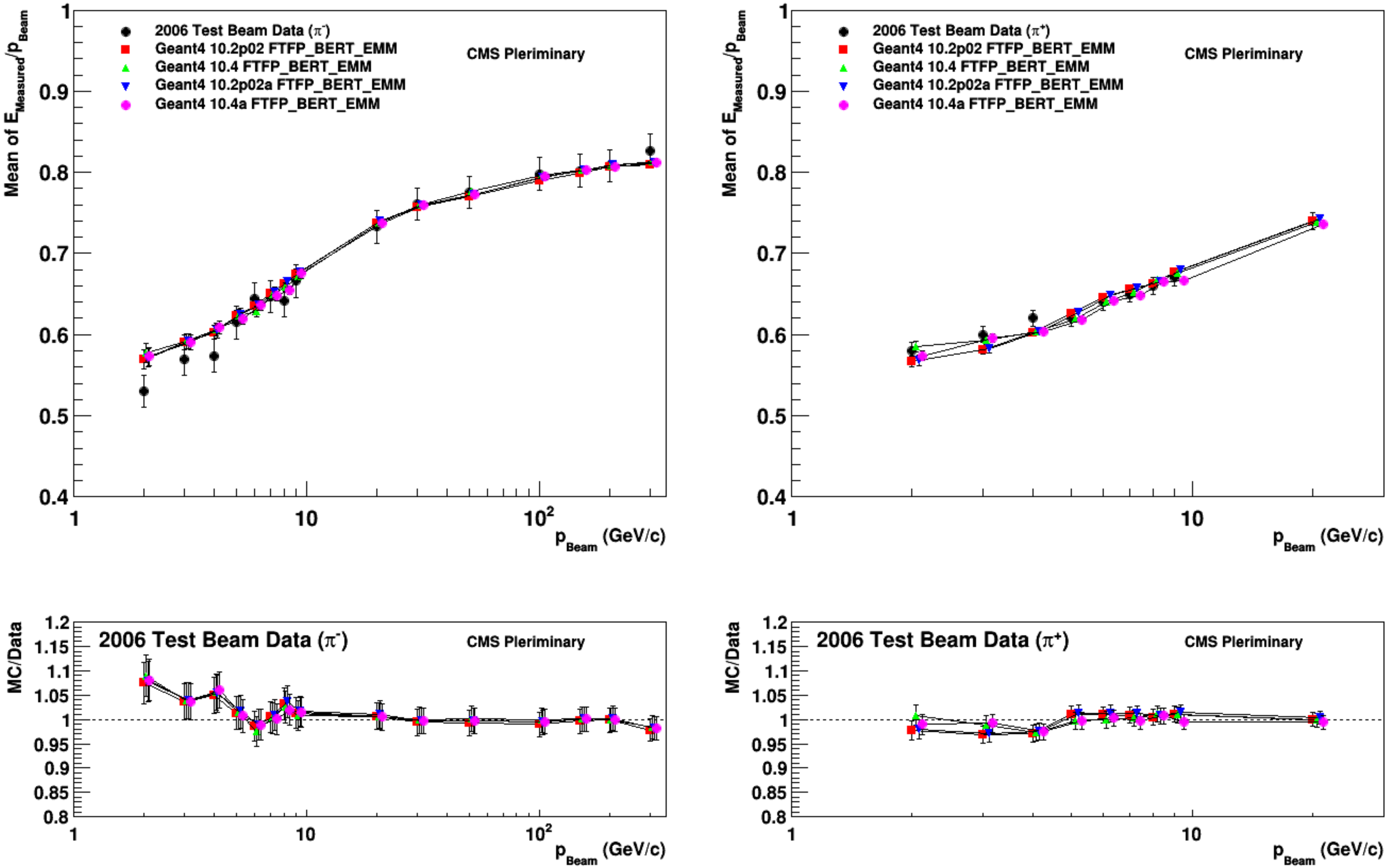

- Total energy measured for negative pion beams of 4 GeV/c and 6 GeV/c
- The data have a longer tail than the MC (mean level of disagreement could be as high as 20%)

# Energy for Protons



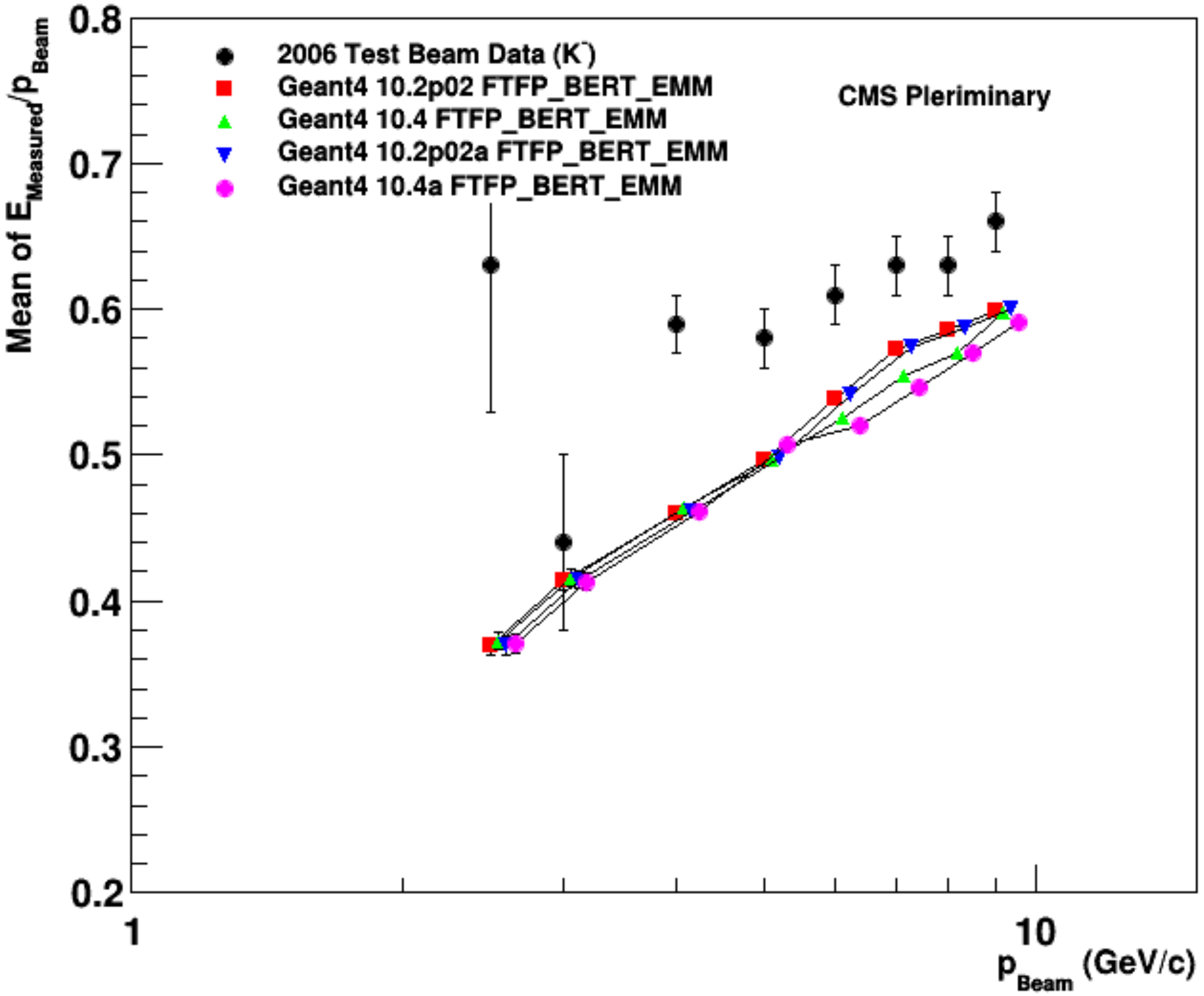

- Energy distribution for protons at 3 and 7 GeV/c
- All four versions of Monte Carlo provide a decent description of the data (the level of agreement is better than in the case of pions)

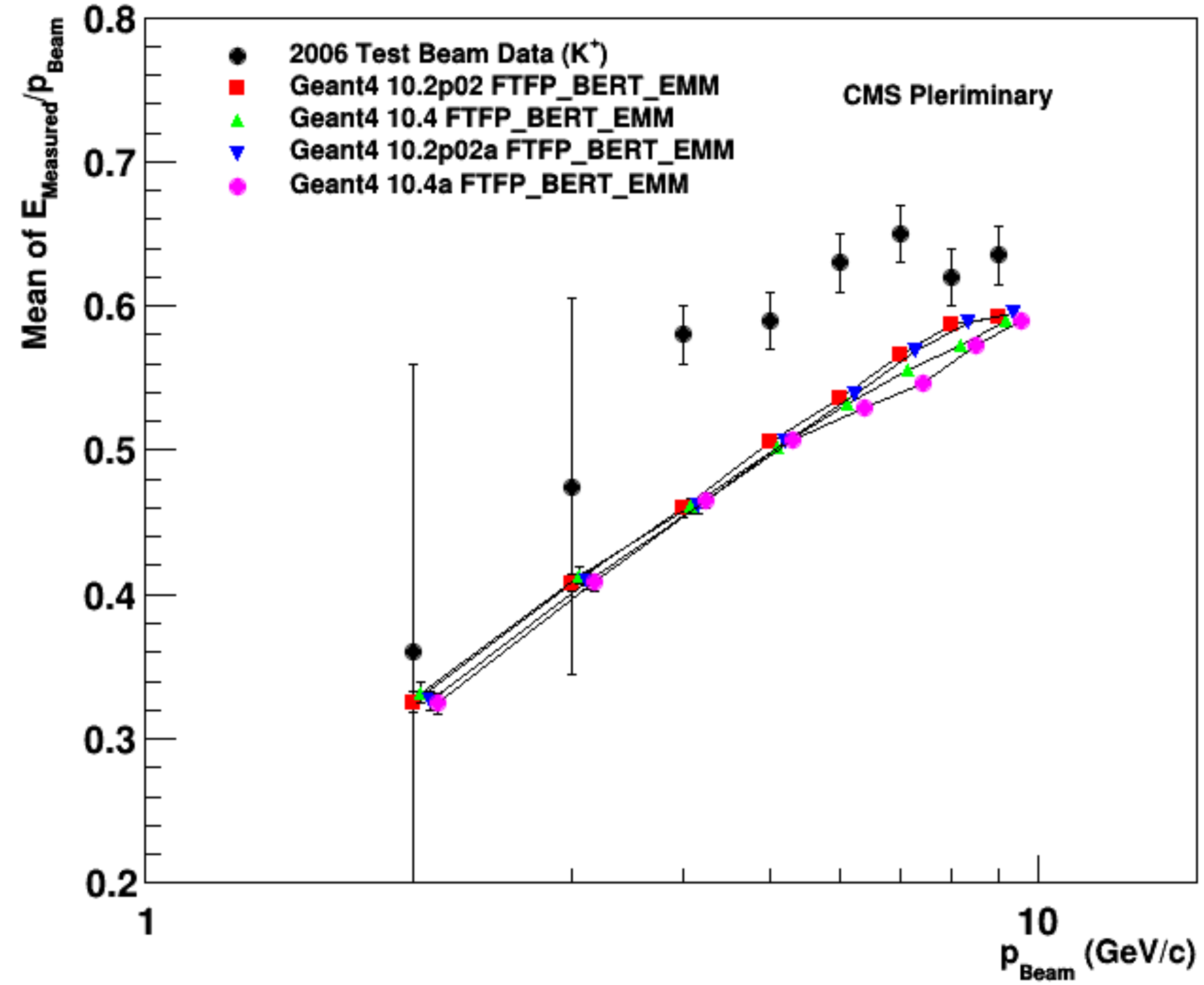
# Mean Response for Pions

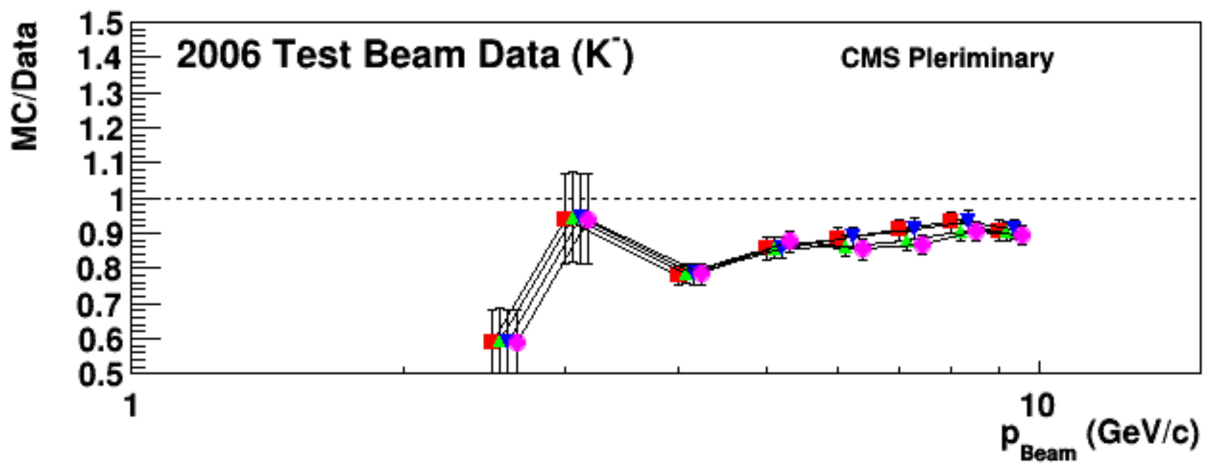


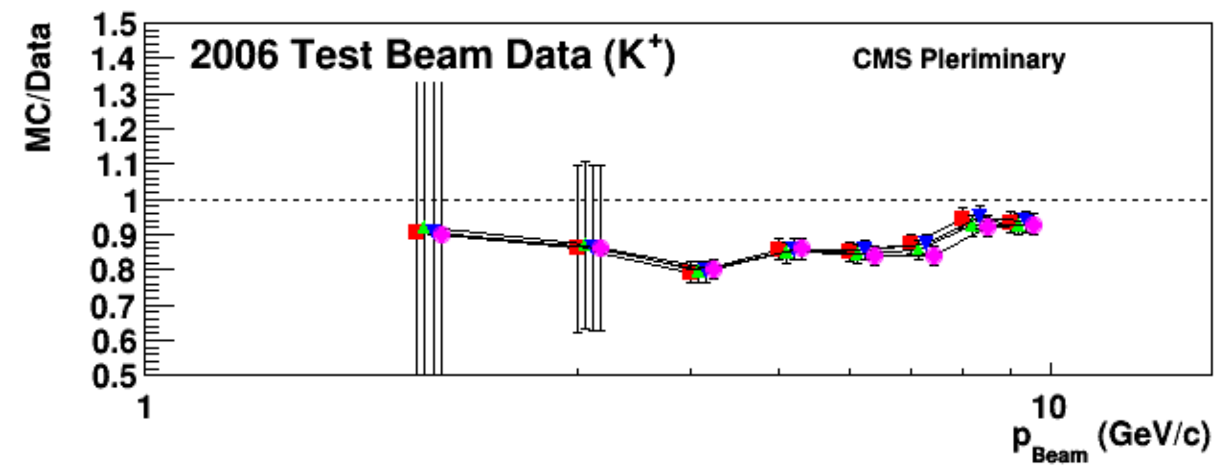

- Use 2 versions of CMSSW builds using Geant4 versions 10.2.p02 and 10.4

# Mean Response for Kaons



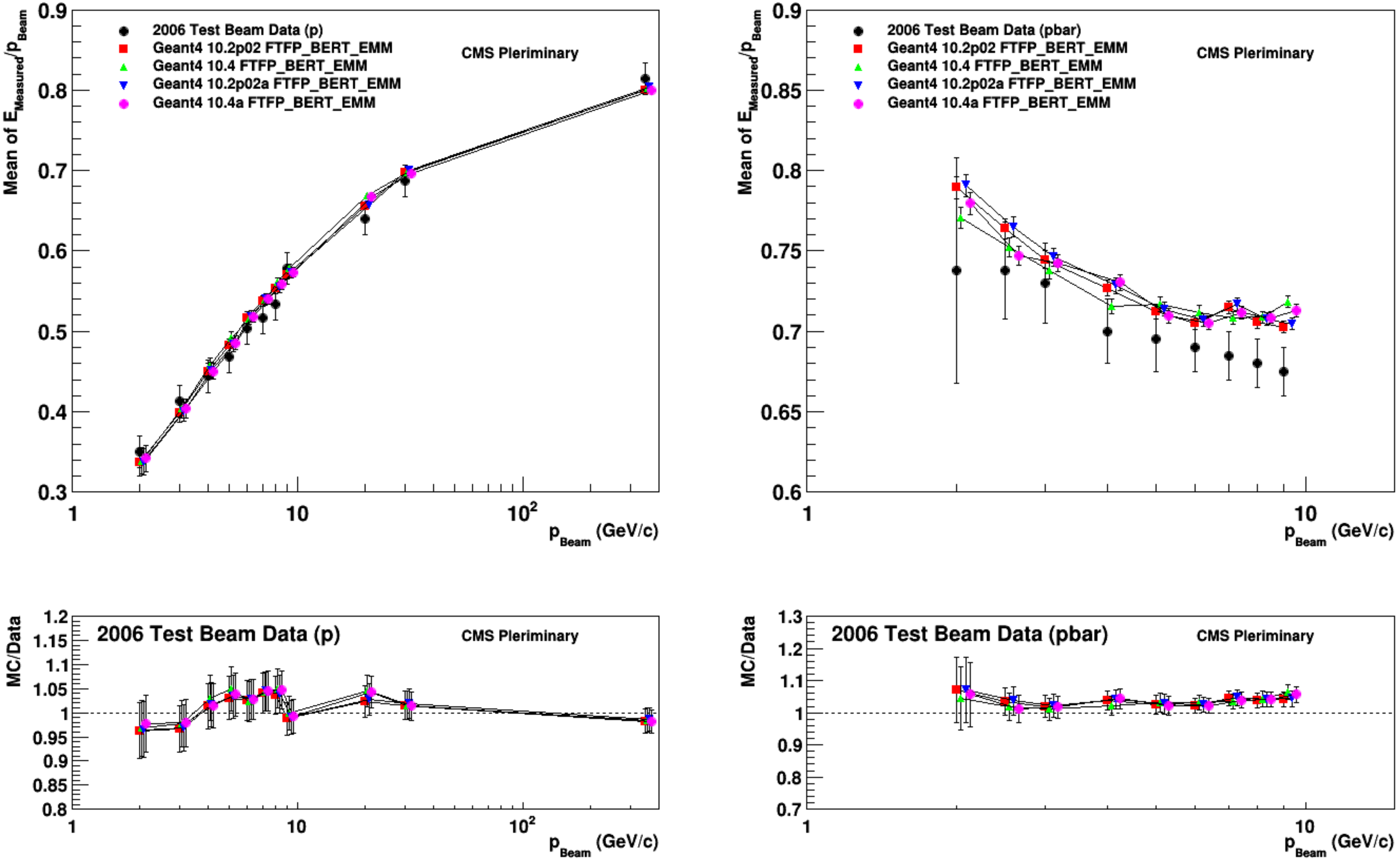



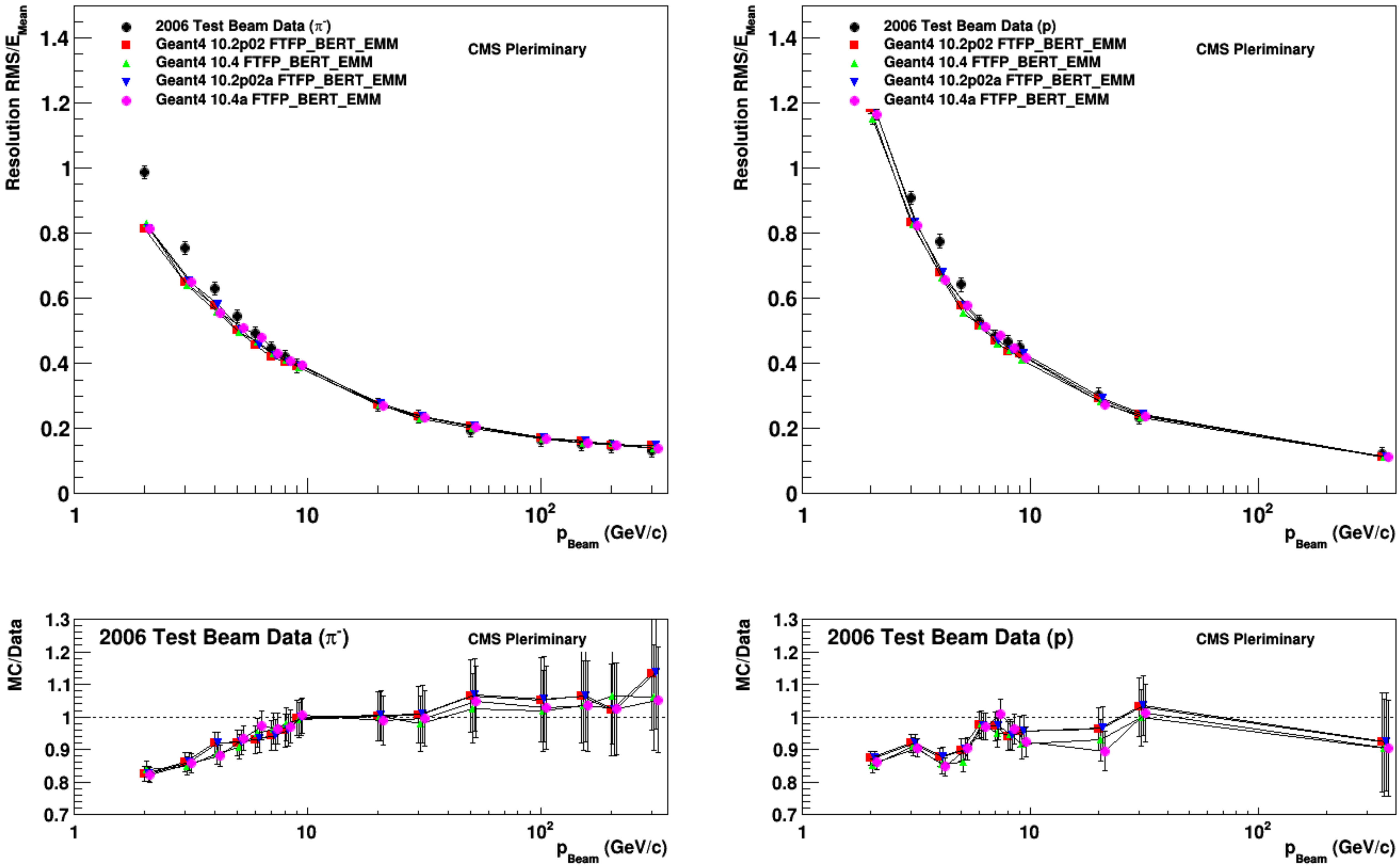

- Both Geant4 versions under predict for kaons (of either charge)

# Mean Response for Protons-AntiProtons



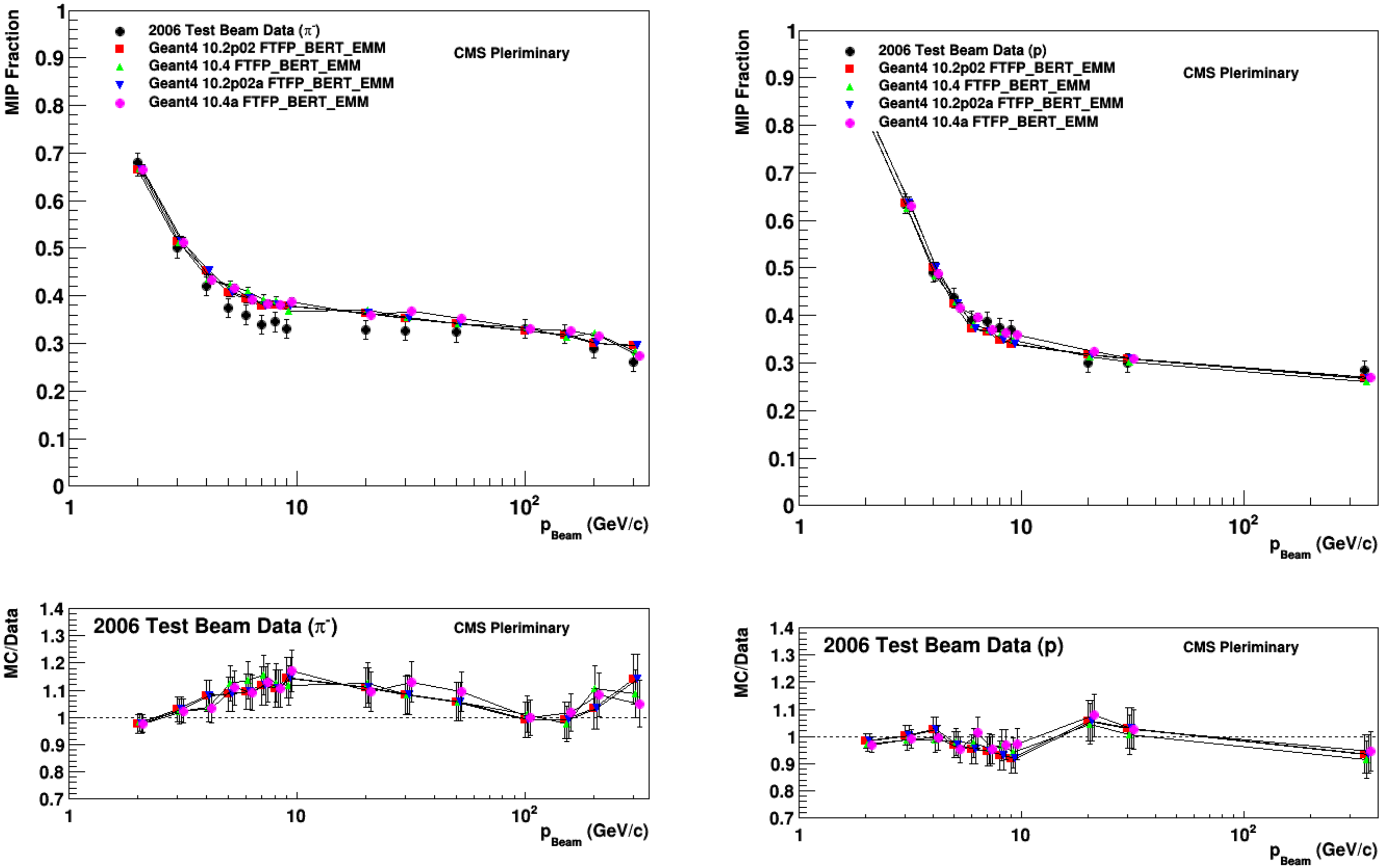

- Both Geant4 versions over-predict mean response for anti-protons

# Energy Resolution for π⁻ and protons




- Energy resolution is under-predicted for π⁻ for beam momenta below 8 GeV/c

# MIP Fraction for π⁻ and protons

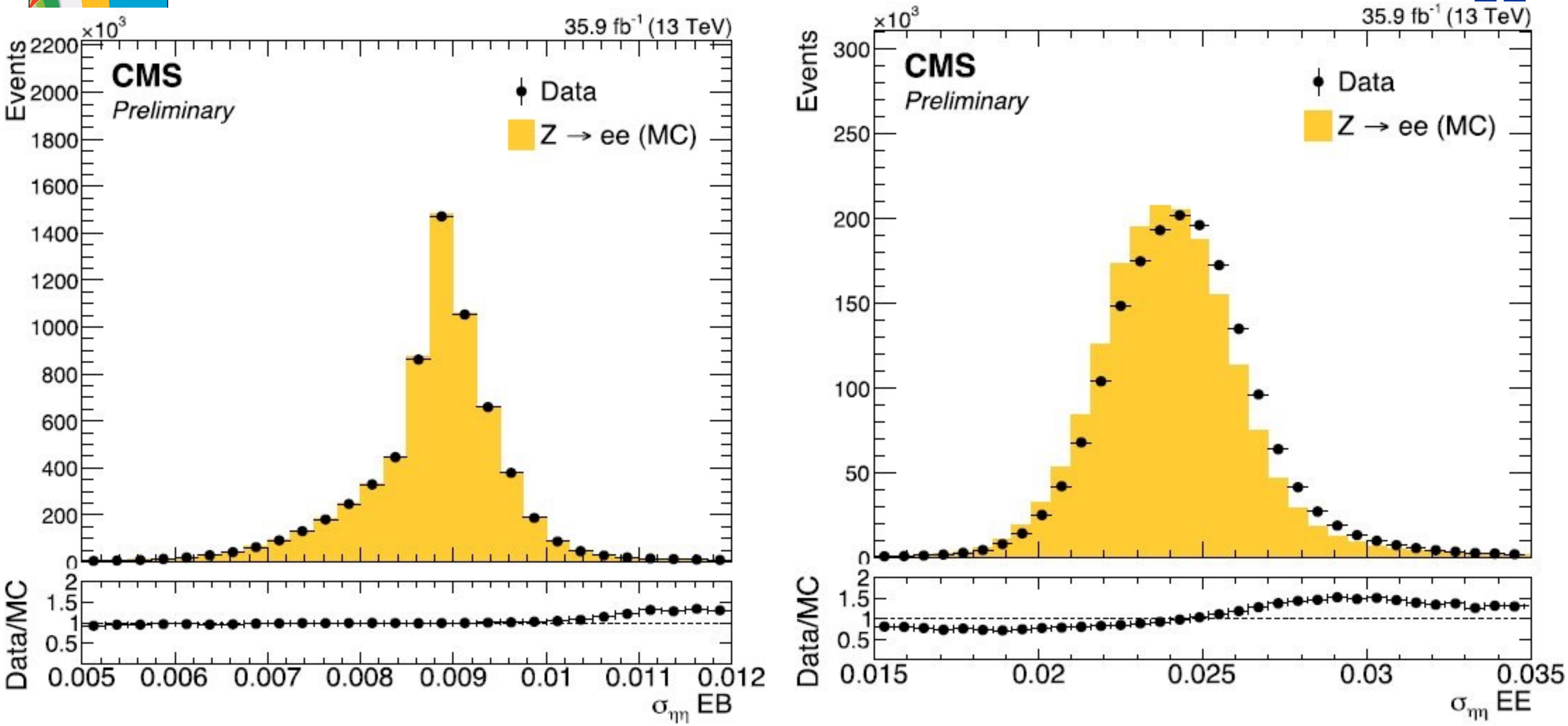



- Fraction of events with energy less than 0.8 GeV in ECAL agrees well for protons

# Electrons in CMS

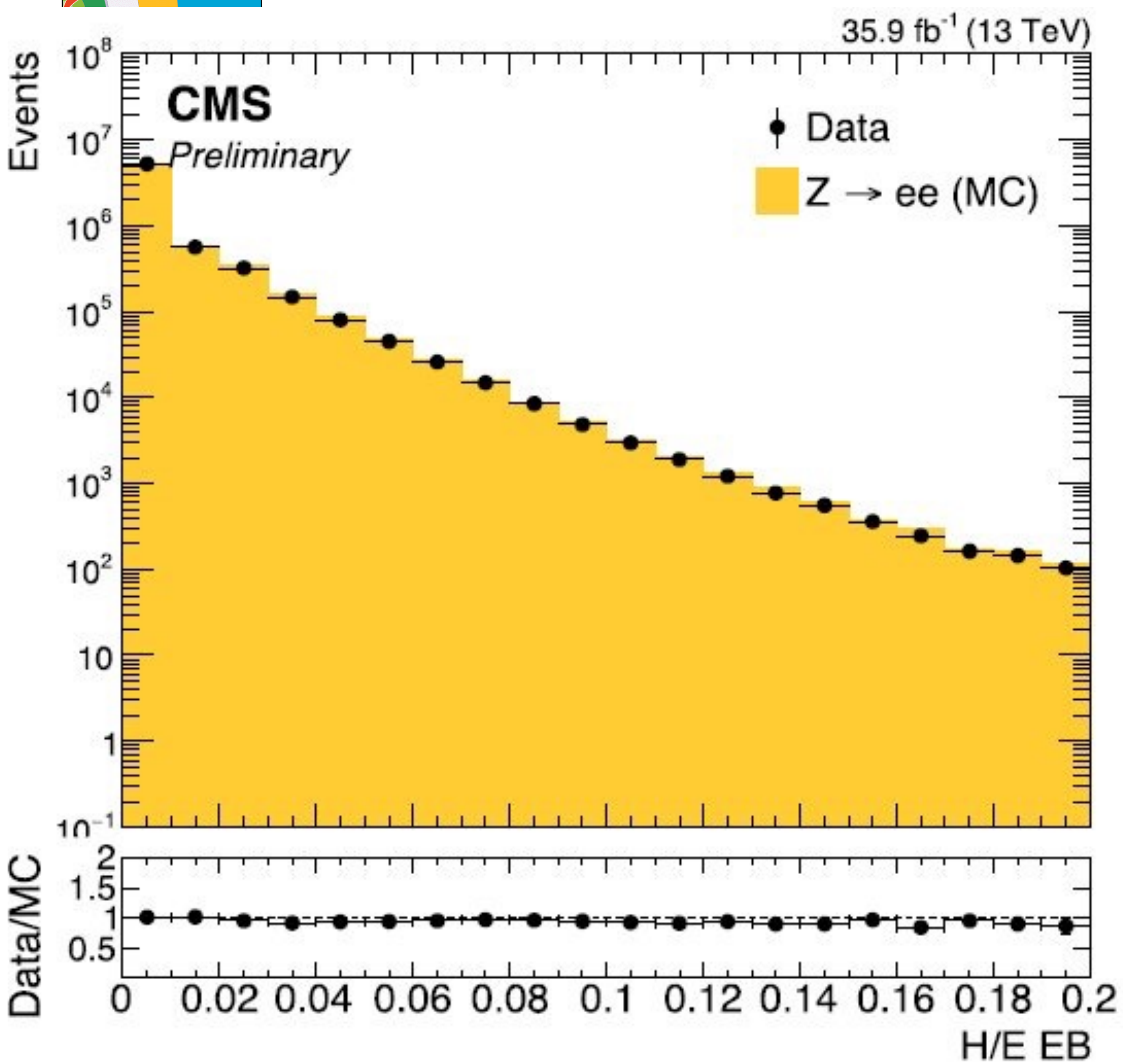
- Use 35.9 $fb^{-1}$ data collected by CMS during 2016
- Utilize events recorded by single electron trigger
- Use tag and probe method and look for electrons from Z decays
- Tag electron selection
  - $p_T > 40$ GeV
  - $|\eta| < 2.1$
  - Tight identification criteria
  - Matching the trigger e/γ candidate
- Probe electron selection
  - $p_T > 25$ GeV
  - $|\eta| < 2.5$
  - Loose identification criteria
  - $80 < M(ee) < 100$ GeV
- Monte Carlo sample corresponds to Drell Yan process using Madgraph generator

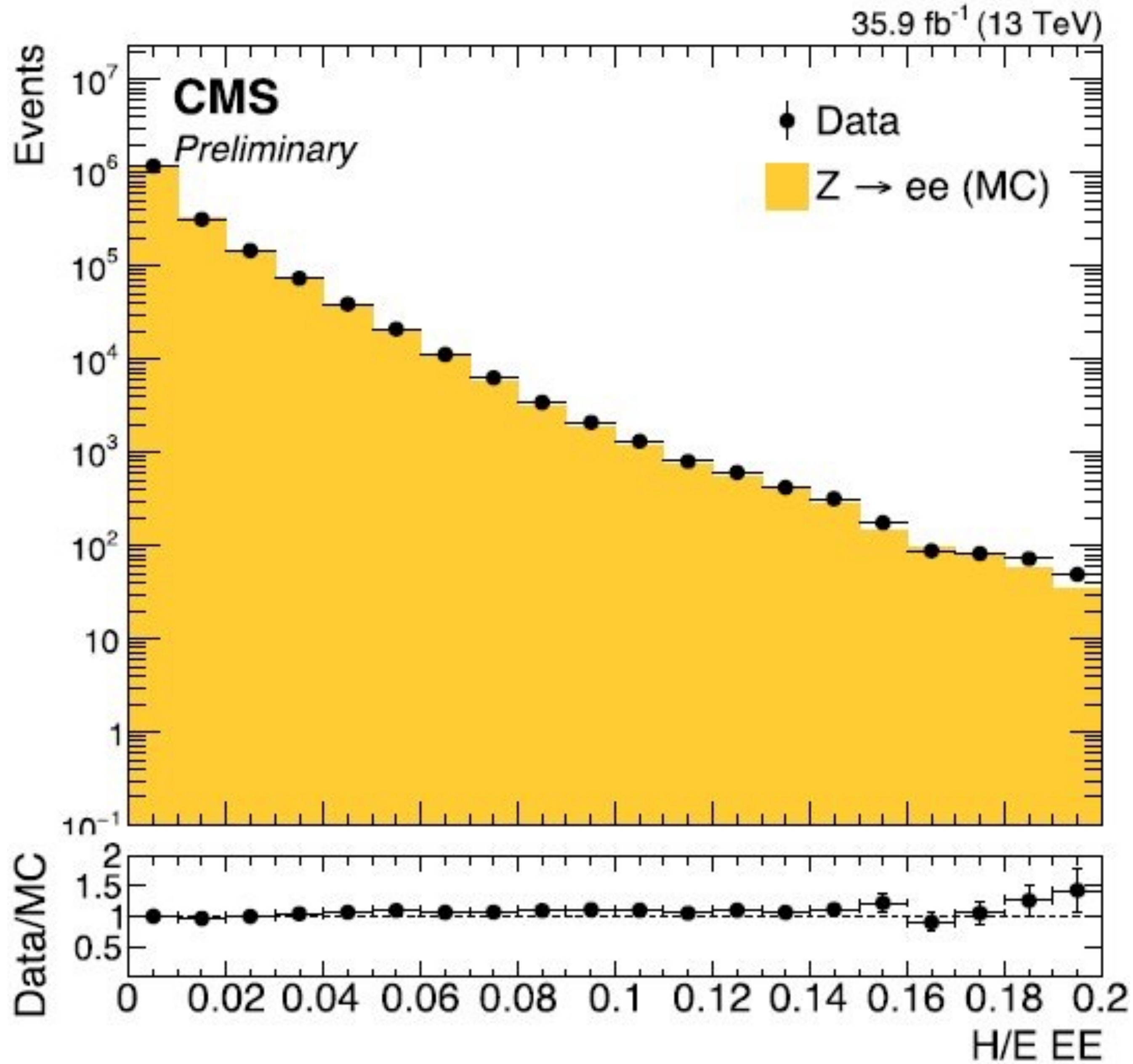
# Crystal-based Shower Width




- Crystal based shower width in the η direction

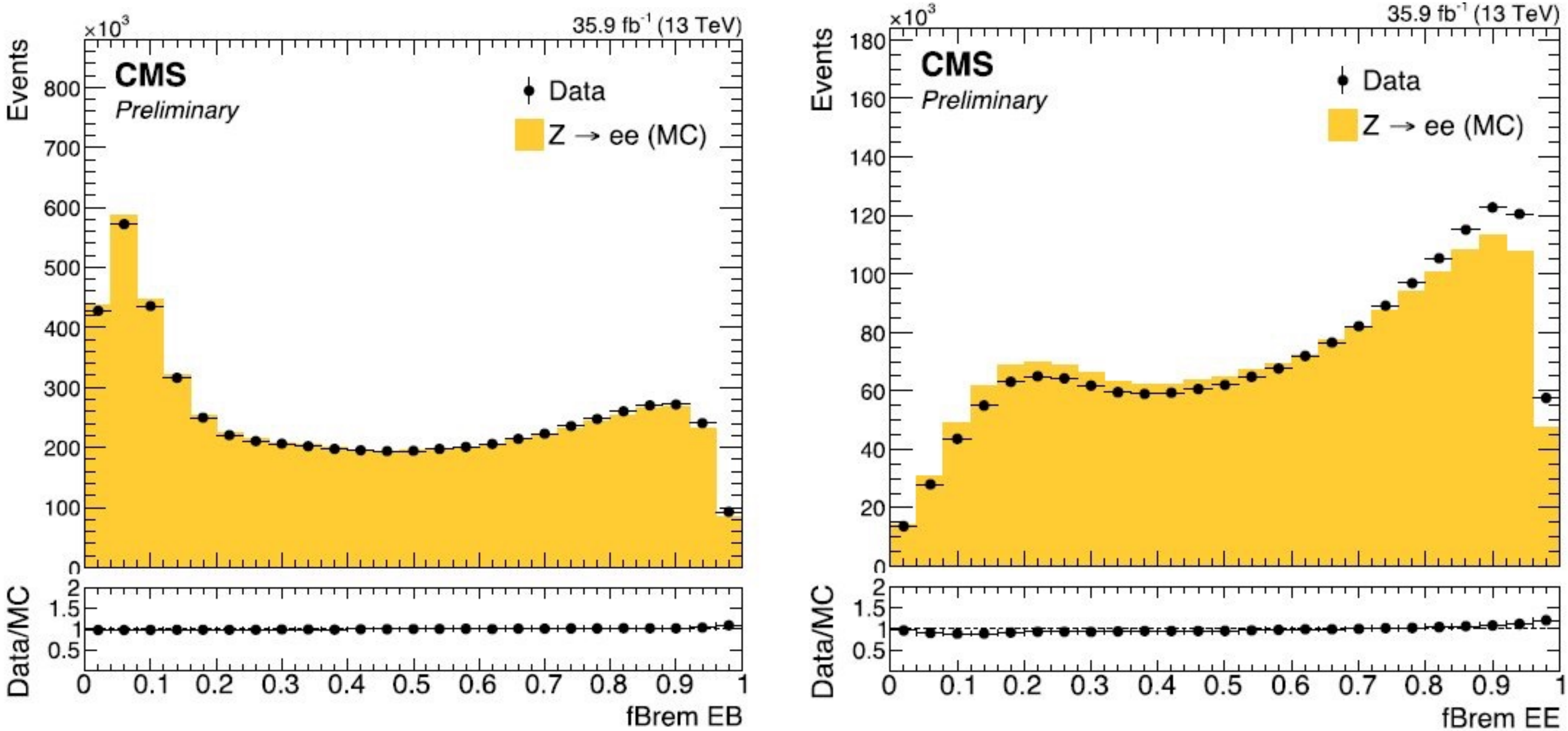
# Hadronic Energy Fraction




- Ratio of energy measured in the hadron calorimeter in a $\Delta R = 0.15$ cone behind the electron seed over the energy measured in the ECAL

# Bremsstrahlung Fraction

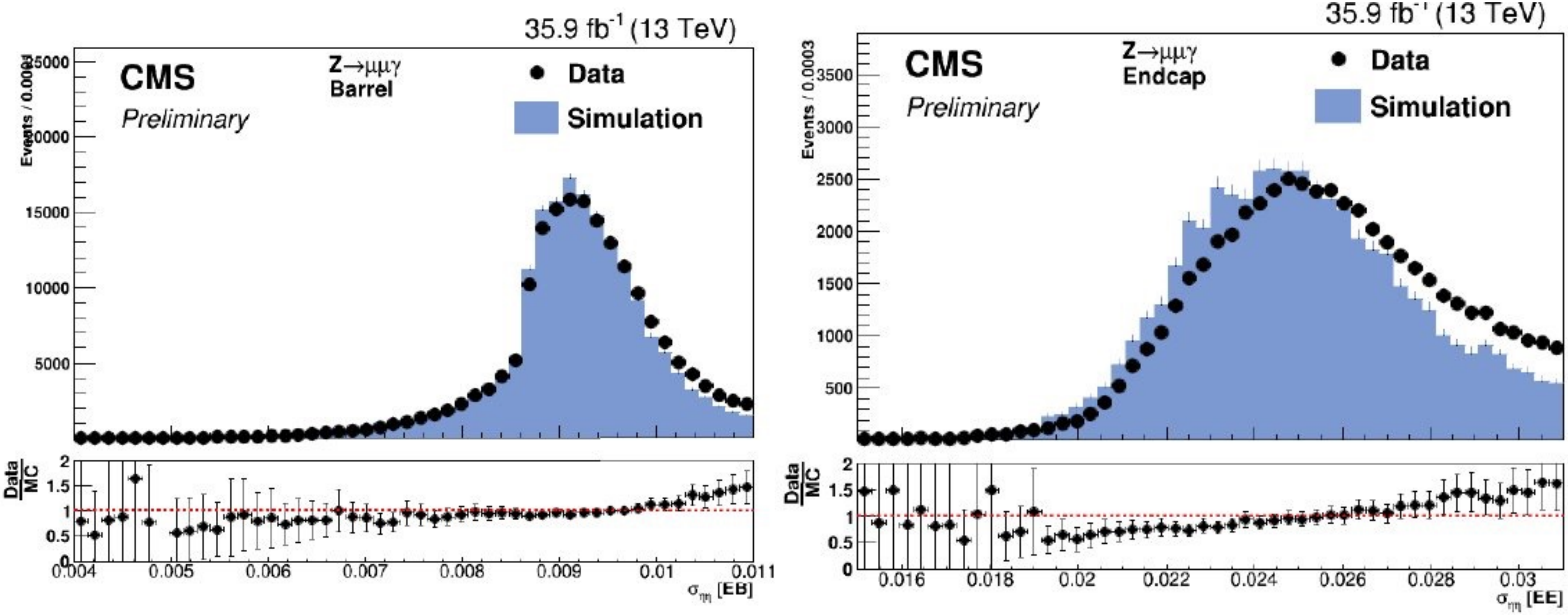



- Fraction of momentum lost to bremsstrahlung measured in the tracker

# Photons in CMS

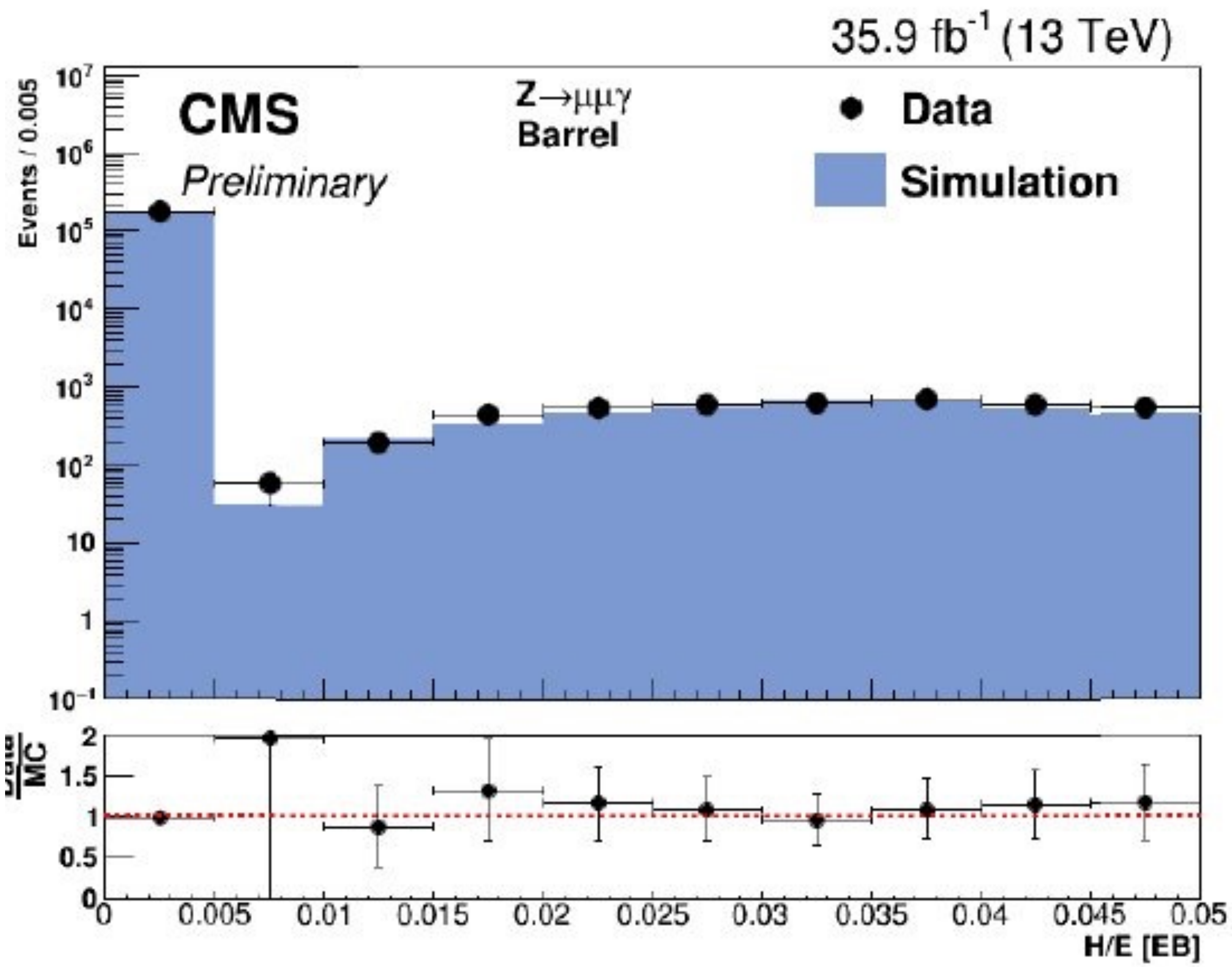
- Use 35.9 $fb^{-1}$ data collected by CMS during 2016
- Utilize events recorded by double muon trigger
- Select events corresponding to Z decays to $\mu\mu\gamma$
- Muon selection
  - $p_T > 20, 10$ GeV
  - Tight identification criteria
  - Matching the trigger candidates
- Photon selection
  - $p_T > 10$ GeV
  - Identification using H/E, $\sigma_{\eta\eta}$ and charge isolation
  - $0.1 < \Delta R(\mu,\gamma)_{min} < 0.8$
  - $70 < M(\mu\mu\gamma) < 110$ GeV
- Monte Carlo sample corresponds to Drell Yan process using Madgraph generator

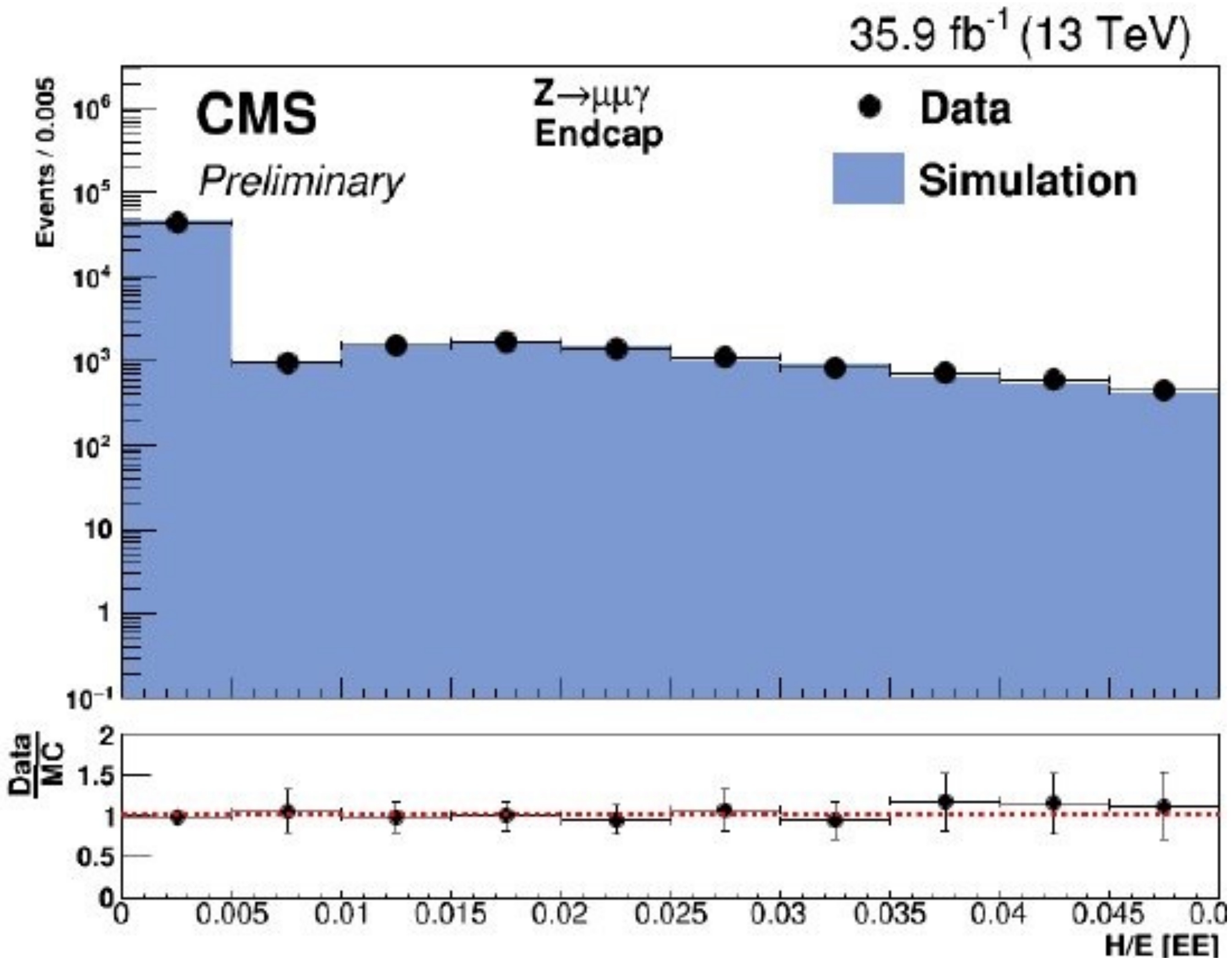
# Crystal-based Shower Width




- Crystal based shower width in the η direction

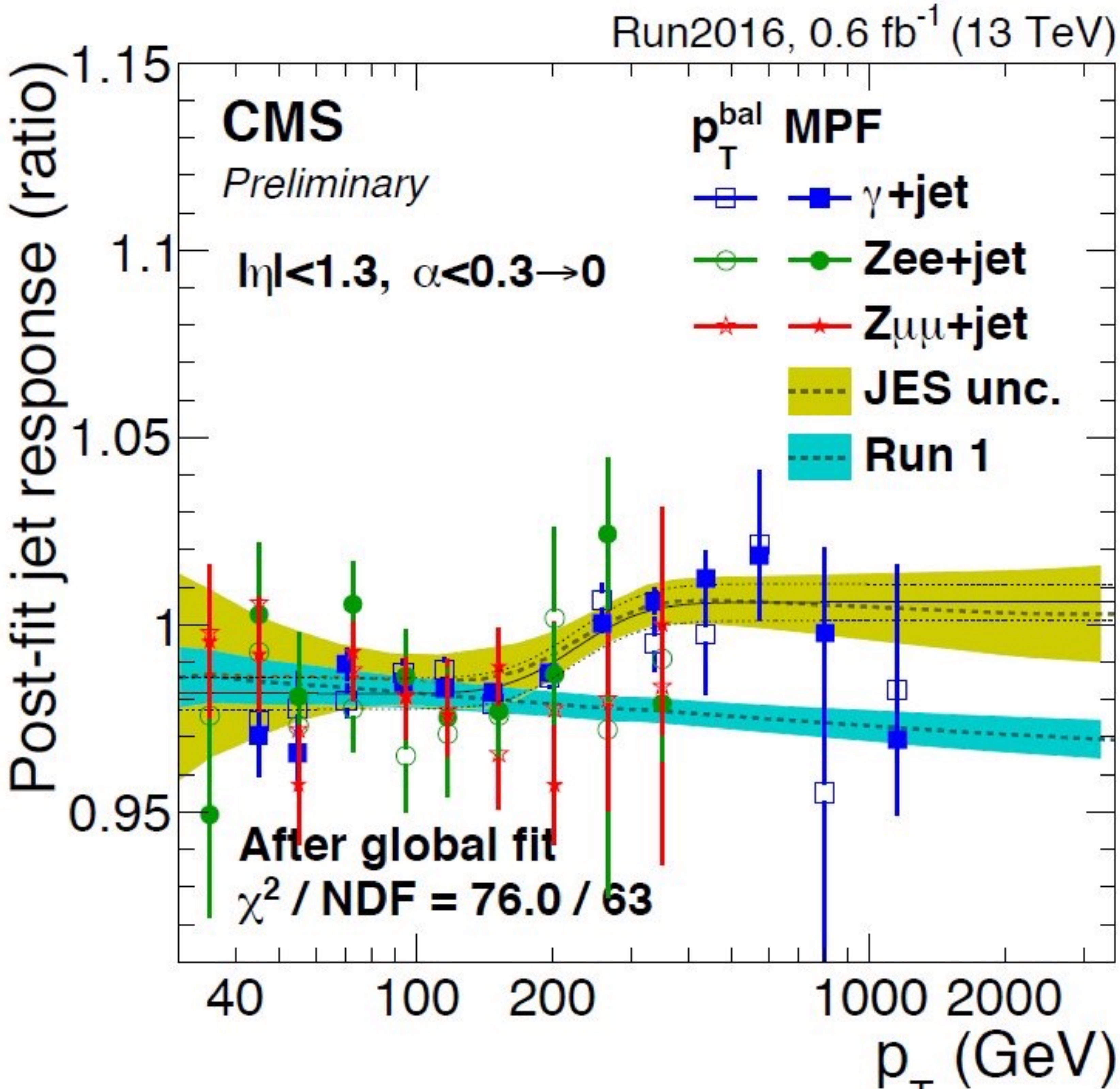
# Hadronic Energy Fraction




- Ratio of energy measured in the hadron calorimeter in a $\Delta R = 0.15$ cone behind the photon seed over the energy measured in the ECAL

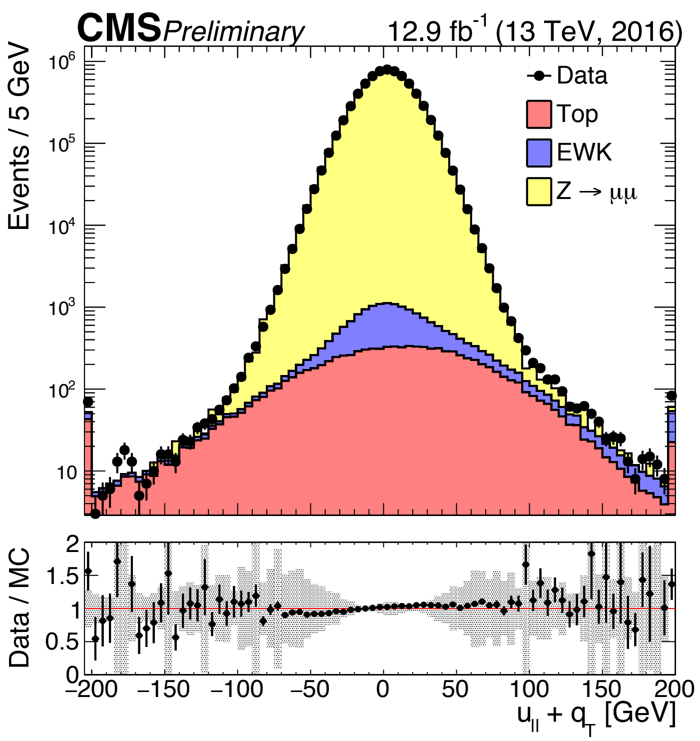
# Jets in CMS

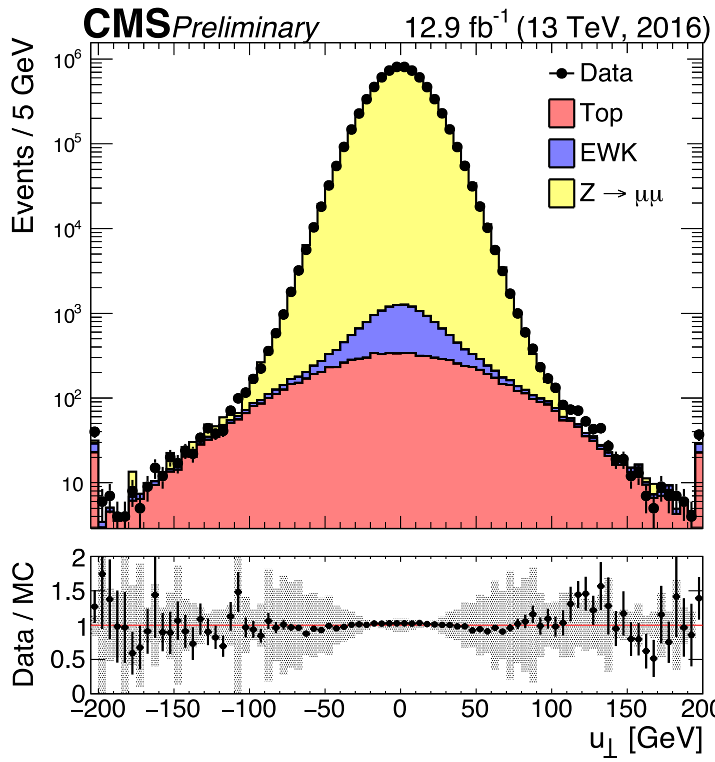


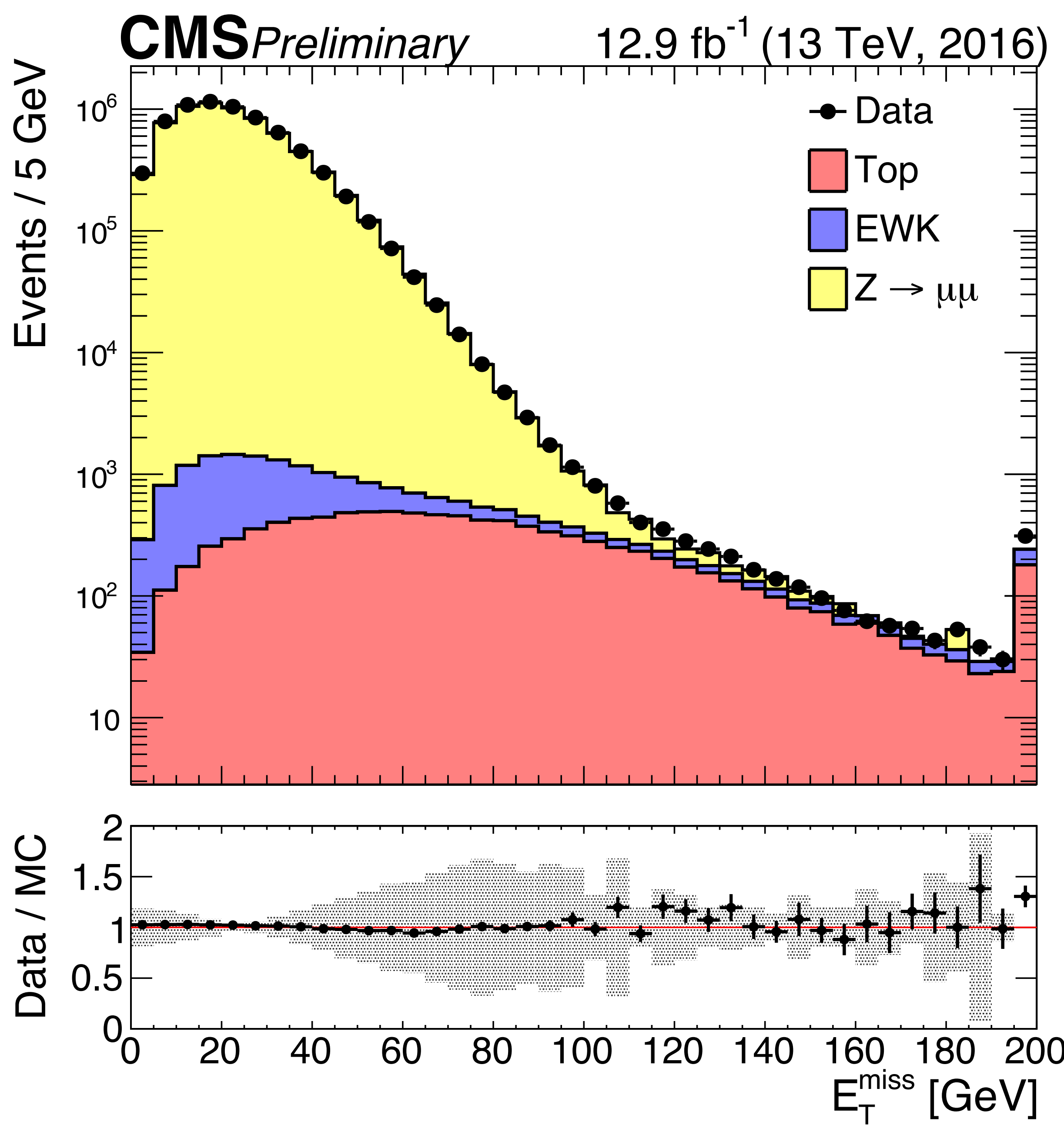

- Jet energy scale evaluated in four samples
  - $\gamma$ + jets
  - Z ($\rightarrow$ $\mu\mu$) + jet
  - Z ($\rightarrow$ ee) + jet
  - Multijet events
- Compare between data and MC as a function of jet $p_T$
  - Kink observed in 13 TeV data at 150 GeV
  - Scale is ~ 0.98 for low $p_T$ jets
  - Scale is ~ 1.0 for high $p_T$ jets

# MET in CMS

- Inclusive Z sample with Z decaying to a pair of μ's are compared with MC samples for Z, W, WZZ, WWZ and t-tbar from MadGraph interfaced with Pythia
- Data selected from di-lepton triggers with threshold of 17 and 8 GeV
- Also compared momentum of recoil hadrons parallel and perpendicular to the Z-direction



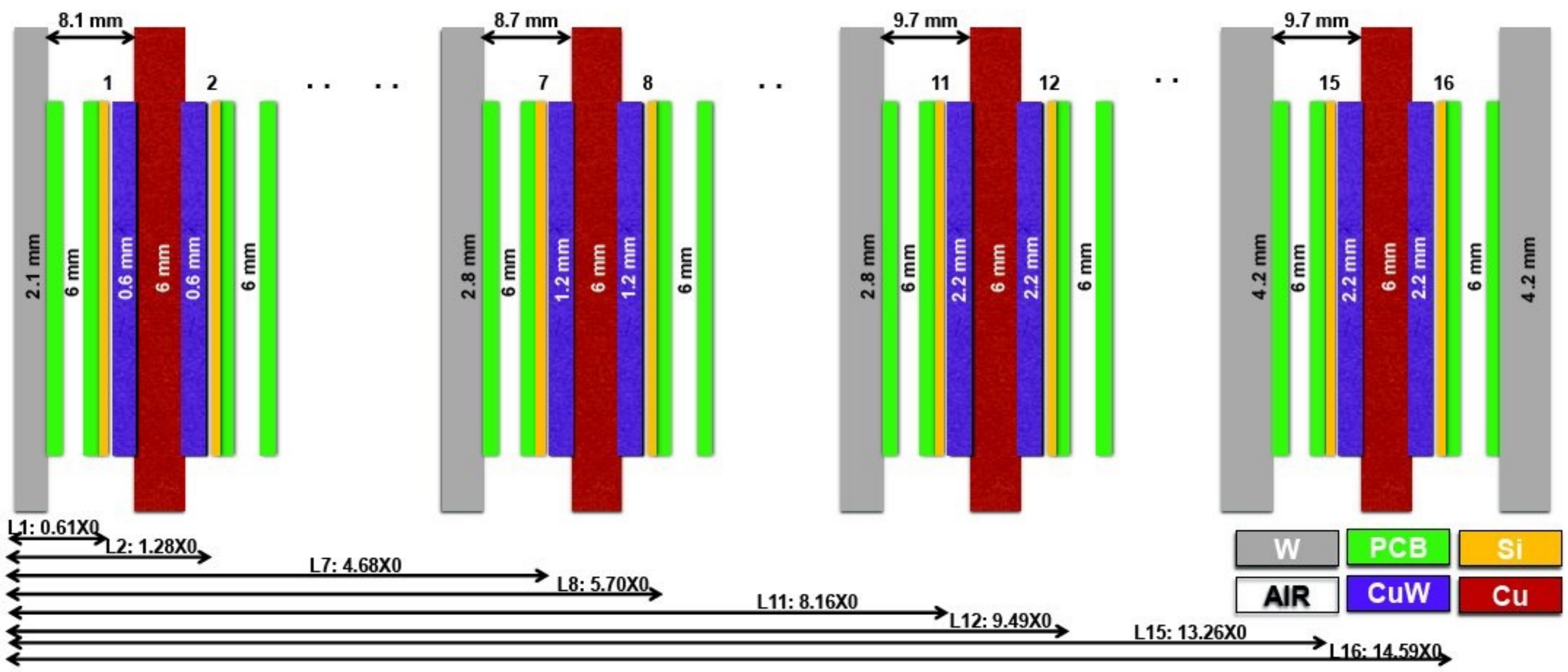



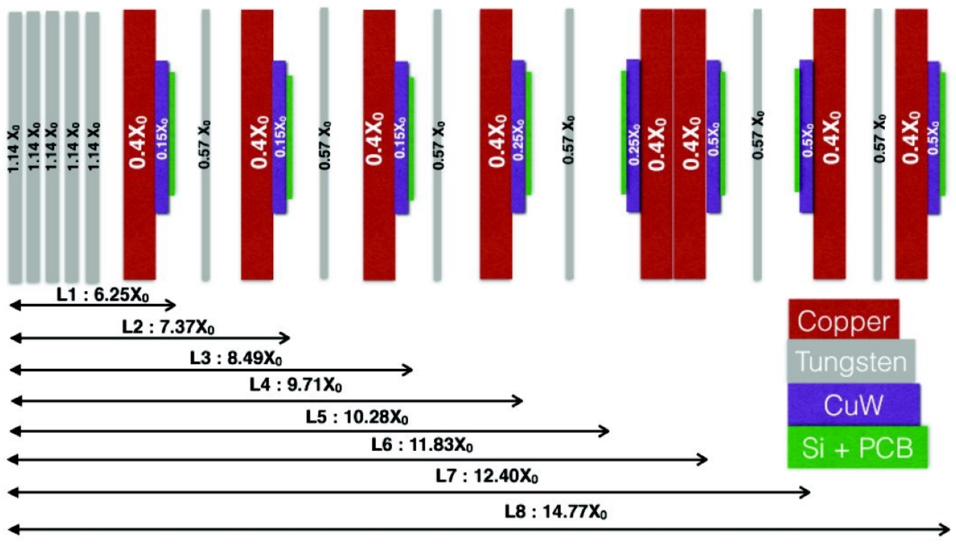



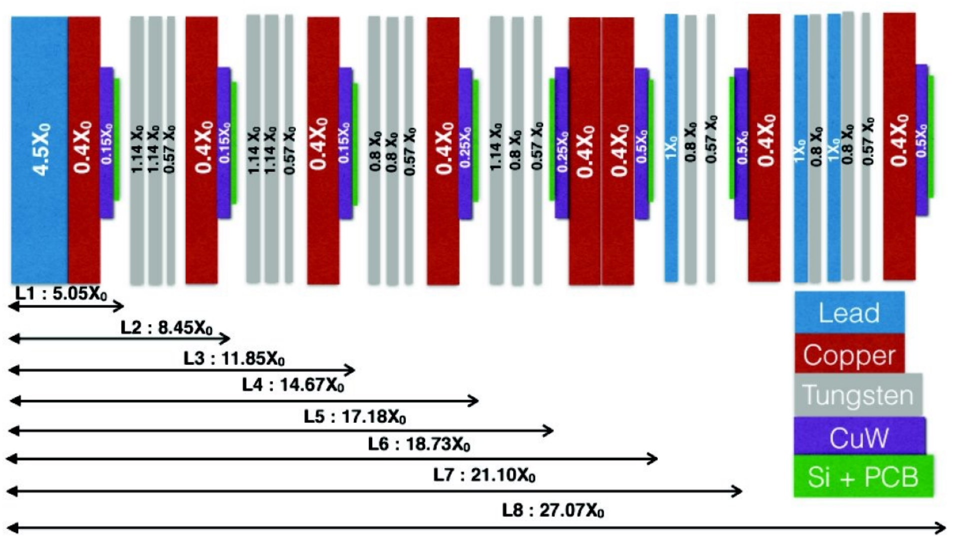

CMS Detector Simulation

# Test Beam with HGCal Prototype

- Prototype of the electromagnetic component of high granularity calorimeter is exposed to electron beams at Fermilab (up to 32 Gev) and at CERN (up to 250 GeV)
  - 16 modules at Fermilab in a setup of ~14.6 $X_0$
  - 8 modules at CERN in two setups of 14.8 $X_0$ and 27.1 $X_0$
- All three setups are simulated using Geant4 version 10.2.p02 and physics list FTFP_BERT_EMM

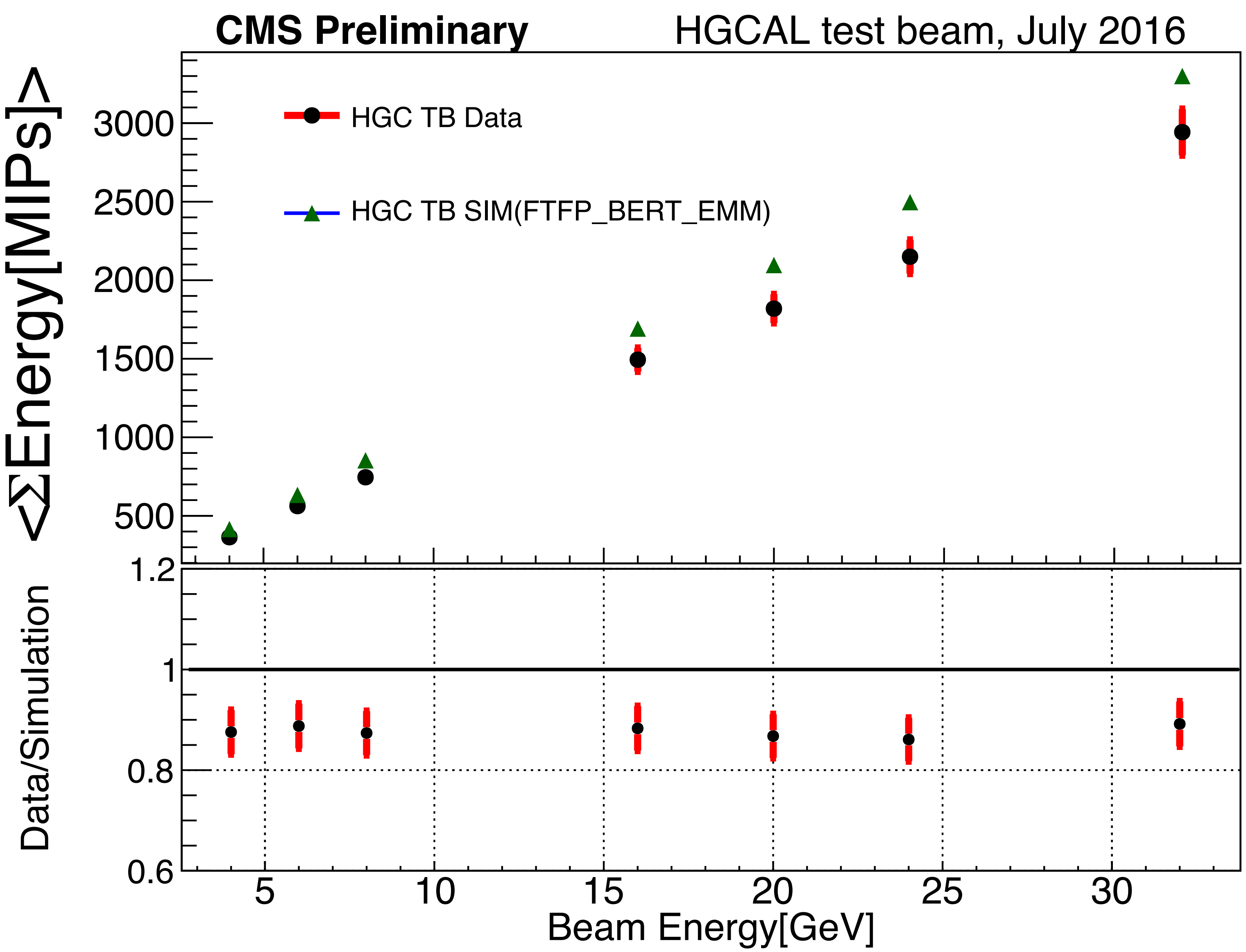


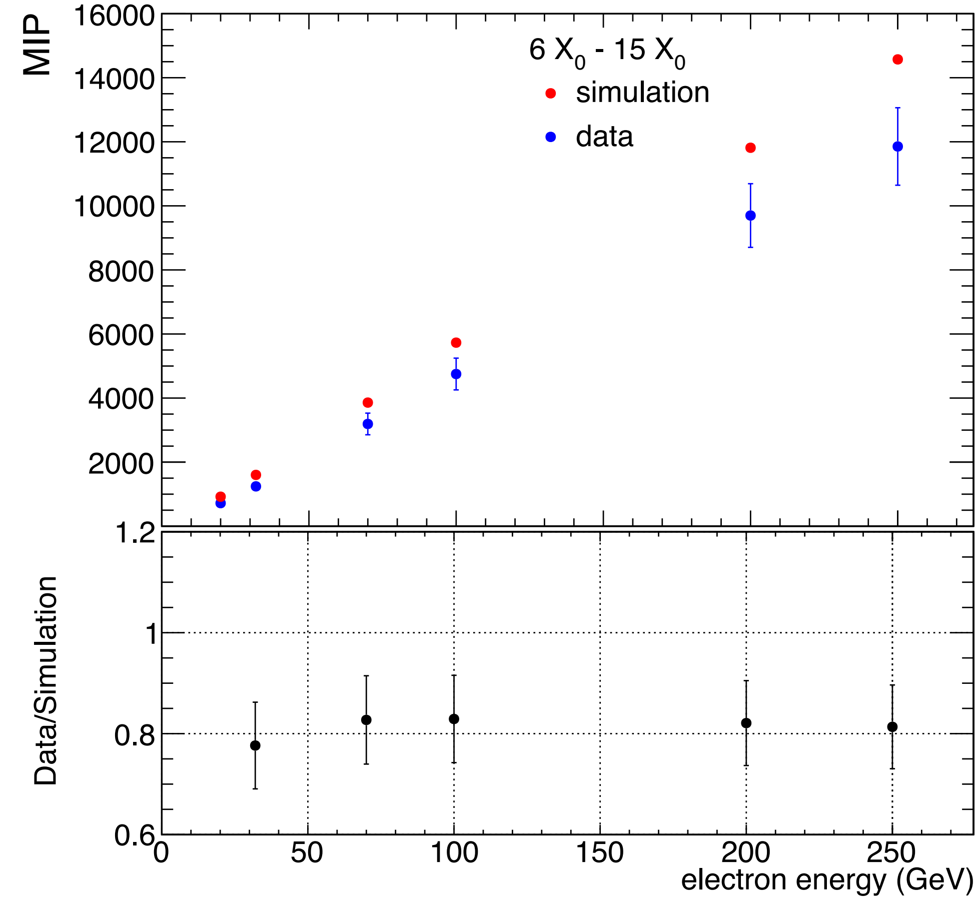



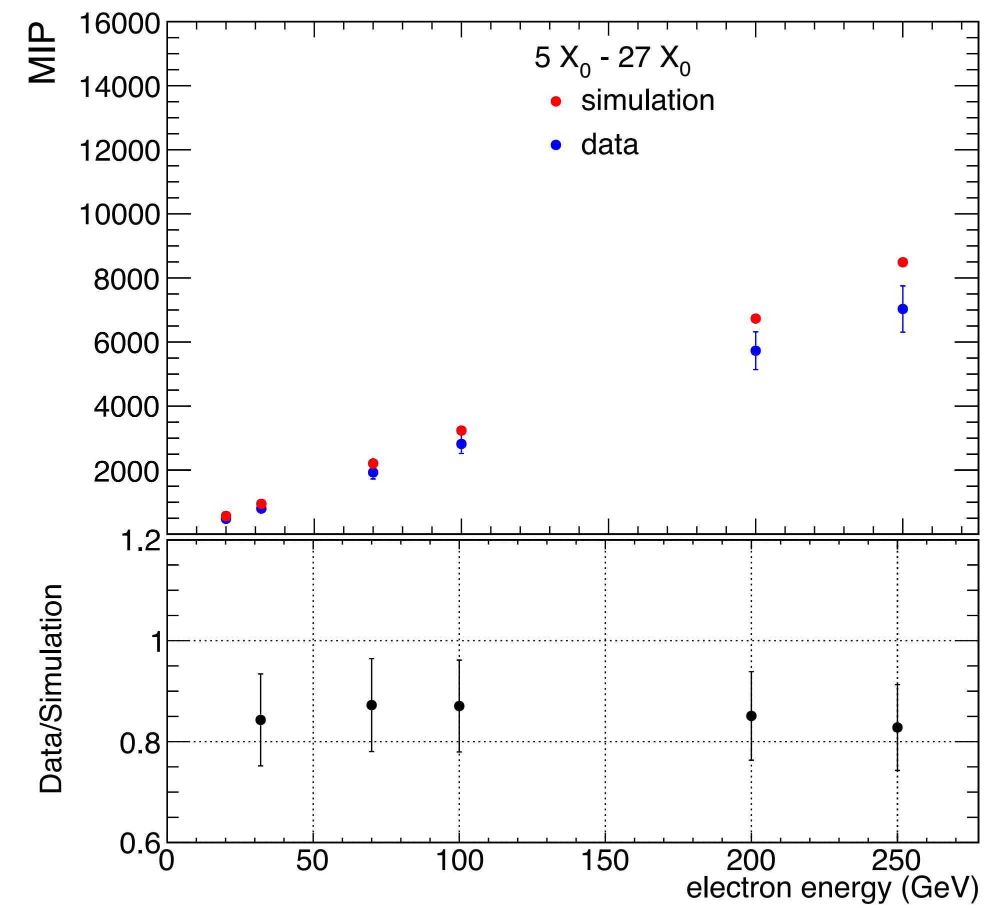



# Response to Electrons

- Both data and MC samples are calibrated with MIP
- Energy deposit in each layer is weighted by a factor depending on material in front
- Linearity is observed in data and MC
- Energy scales are ~10% different in the two cases

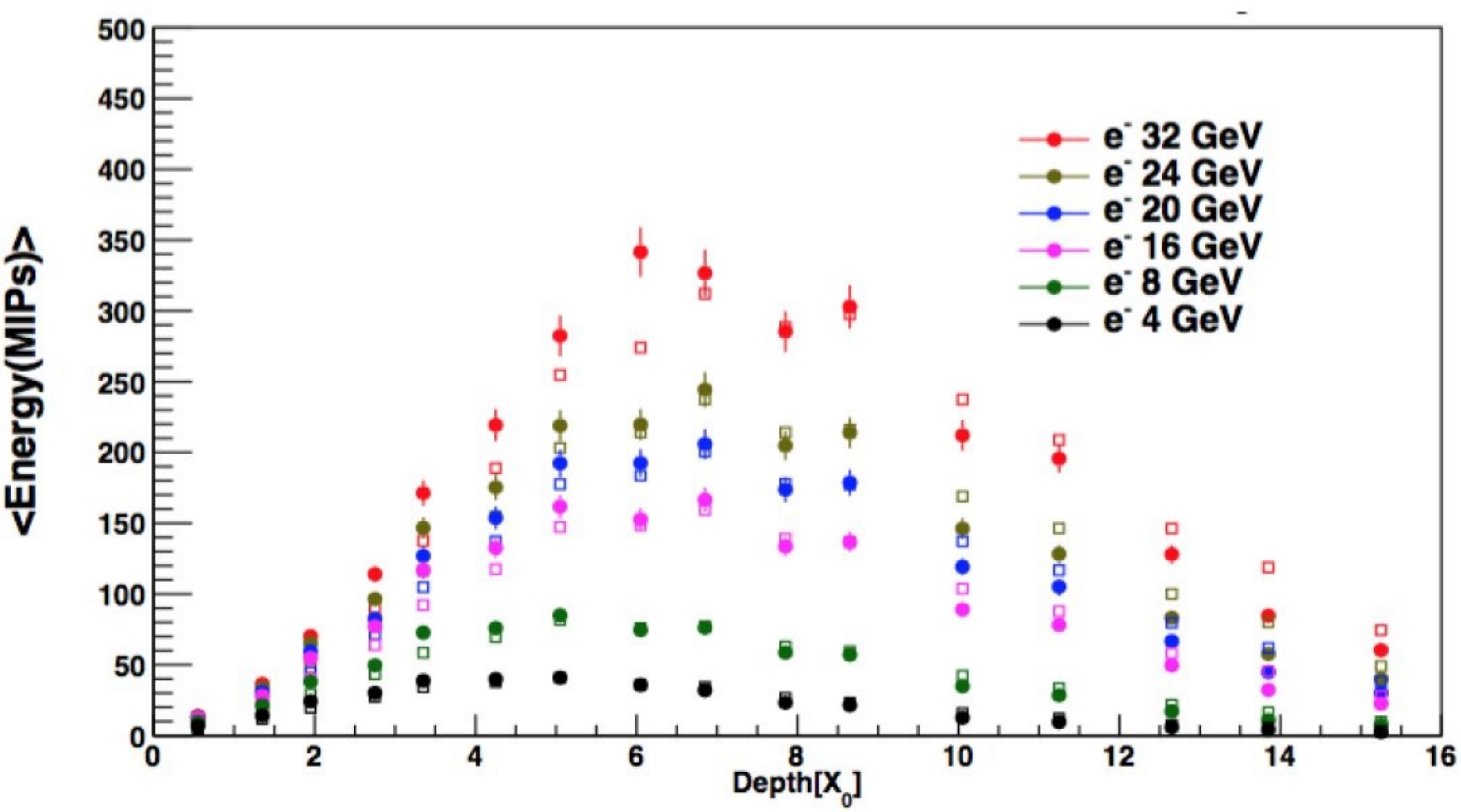


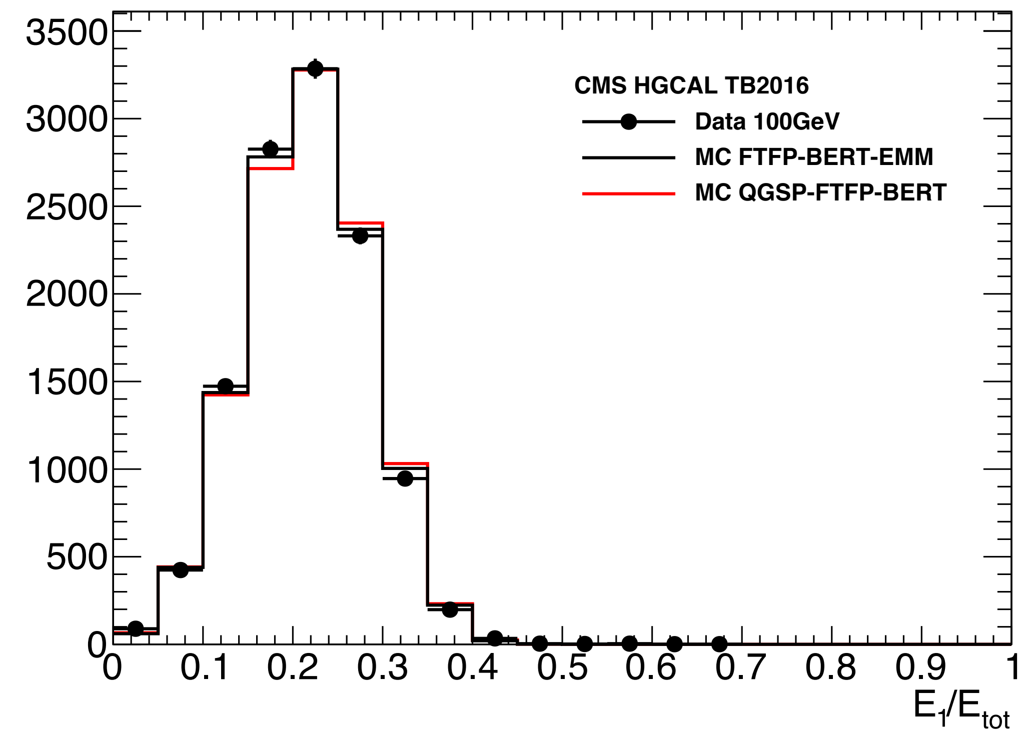



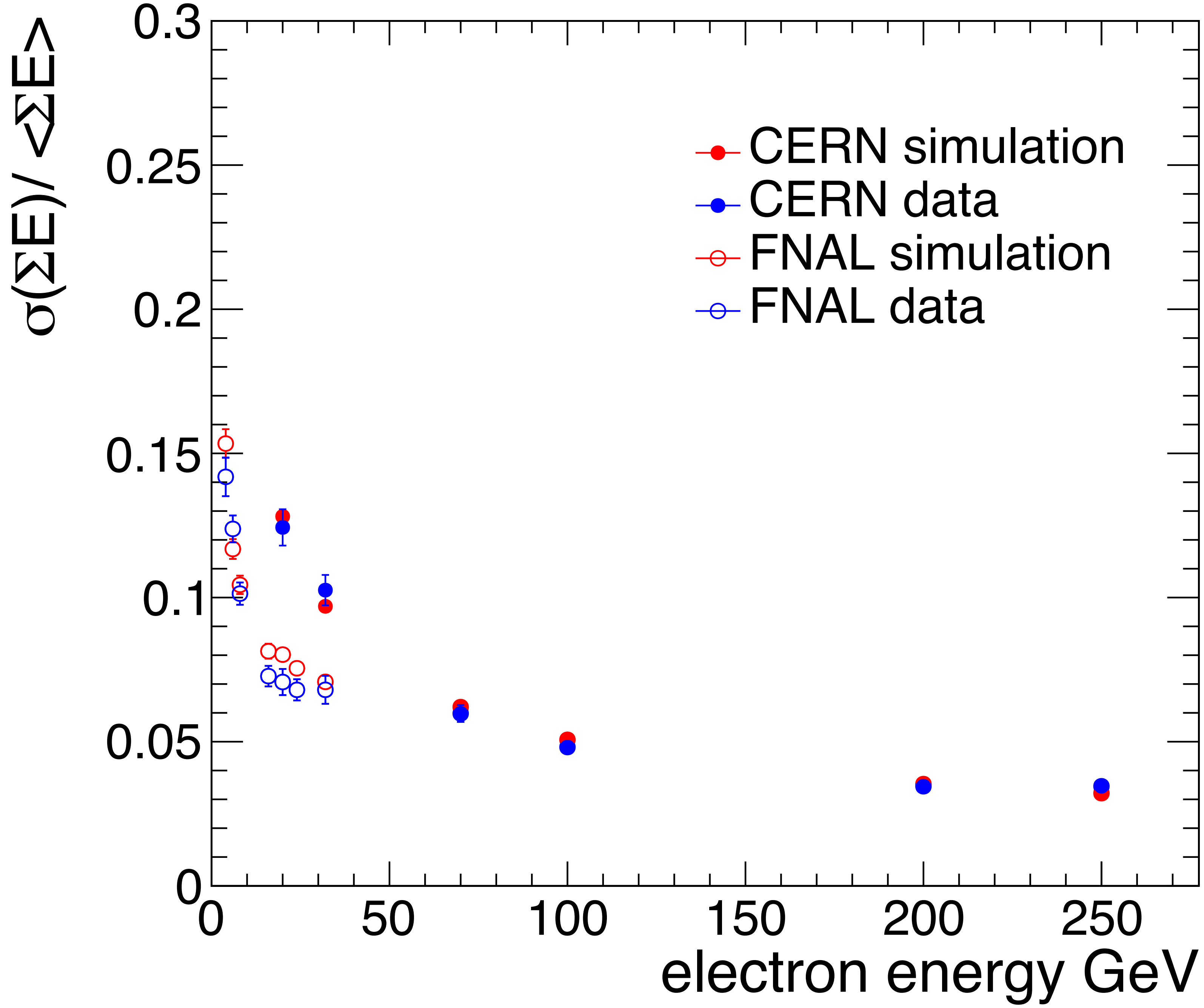
# Shower Shapes

- Shower shapes are compared
  - Longitudinal shower profile studied from mean energy as a function of depth measured in units of $X_0$
  - Lateral profile measured from energy ratio of the central crystal to the total
- Fairly good agreement observed between data and MC



# Energy Resolution



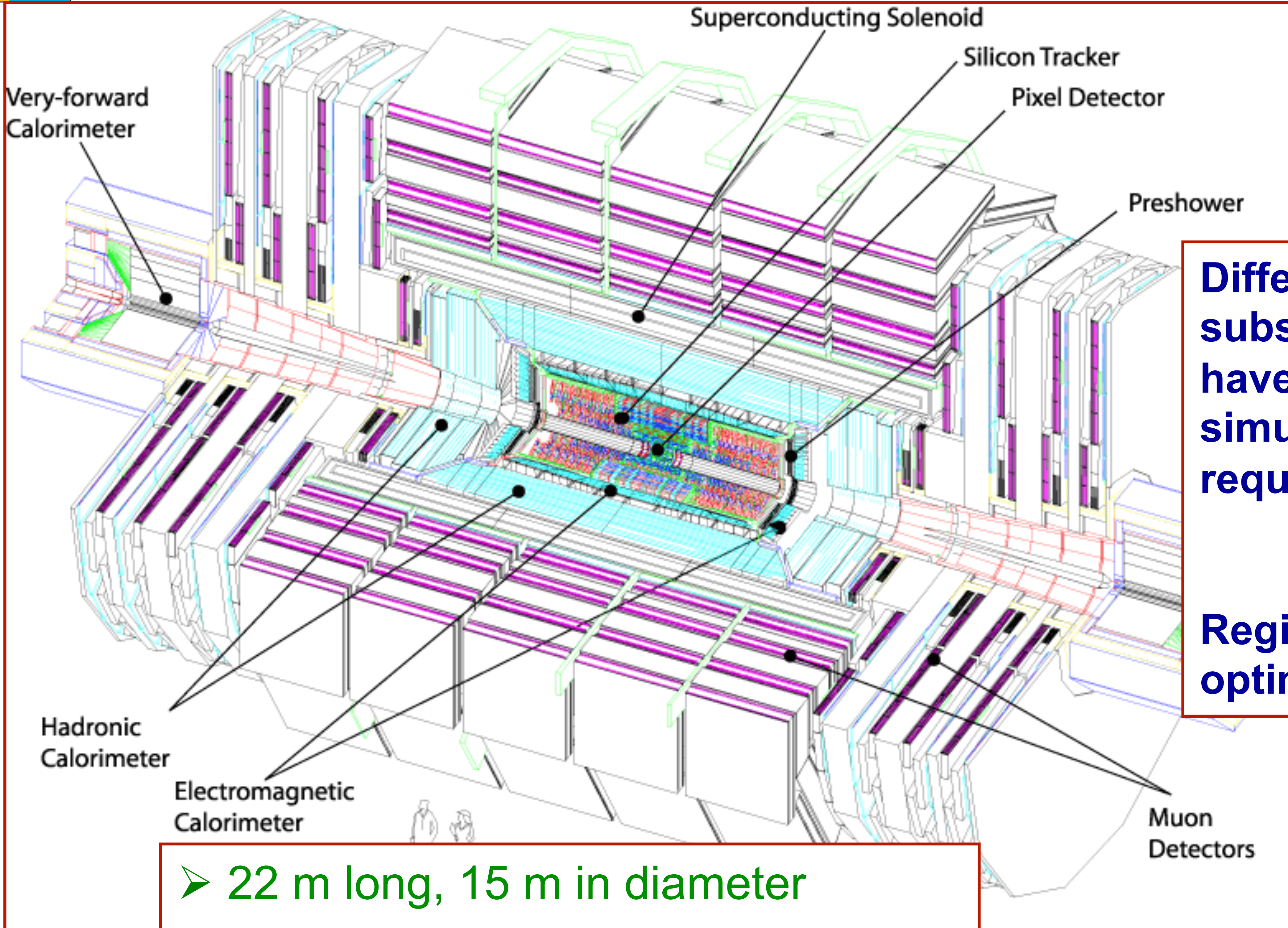

- Energy resolution compares well between data and MC

# Summary

- Detector simulation is an essential tool for modern day nuclear and high energy physics experiments
- CMS has gone through a series of developments to meet the challenges:
  - Accuracy in the predictions
  - Performance in speed and in memory
  - Robust against unusual circumstance
  - Extension for the changing detector
- The ultimate test of detector simulation is how it performs against the data
  - Test beam data with identified particle types are used as one source of validation while isolated charged particles from collision data are used as a second source
  - There is a good agreement between data and Monte Carlo for the new version of physics list (FTFP_BERT_EMM) to be used by CMS for its future event production using Geant4 version 10.4
  - Validation of physics within Geant4 are continued using CMS data from new test beam data
- CMS tries to achieve a well tested Monte Carlo program to get to better results from the CMS experiment

# Additional Slides

# CMS Detector




**Different subsystems have different simulation requirements**

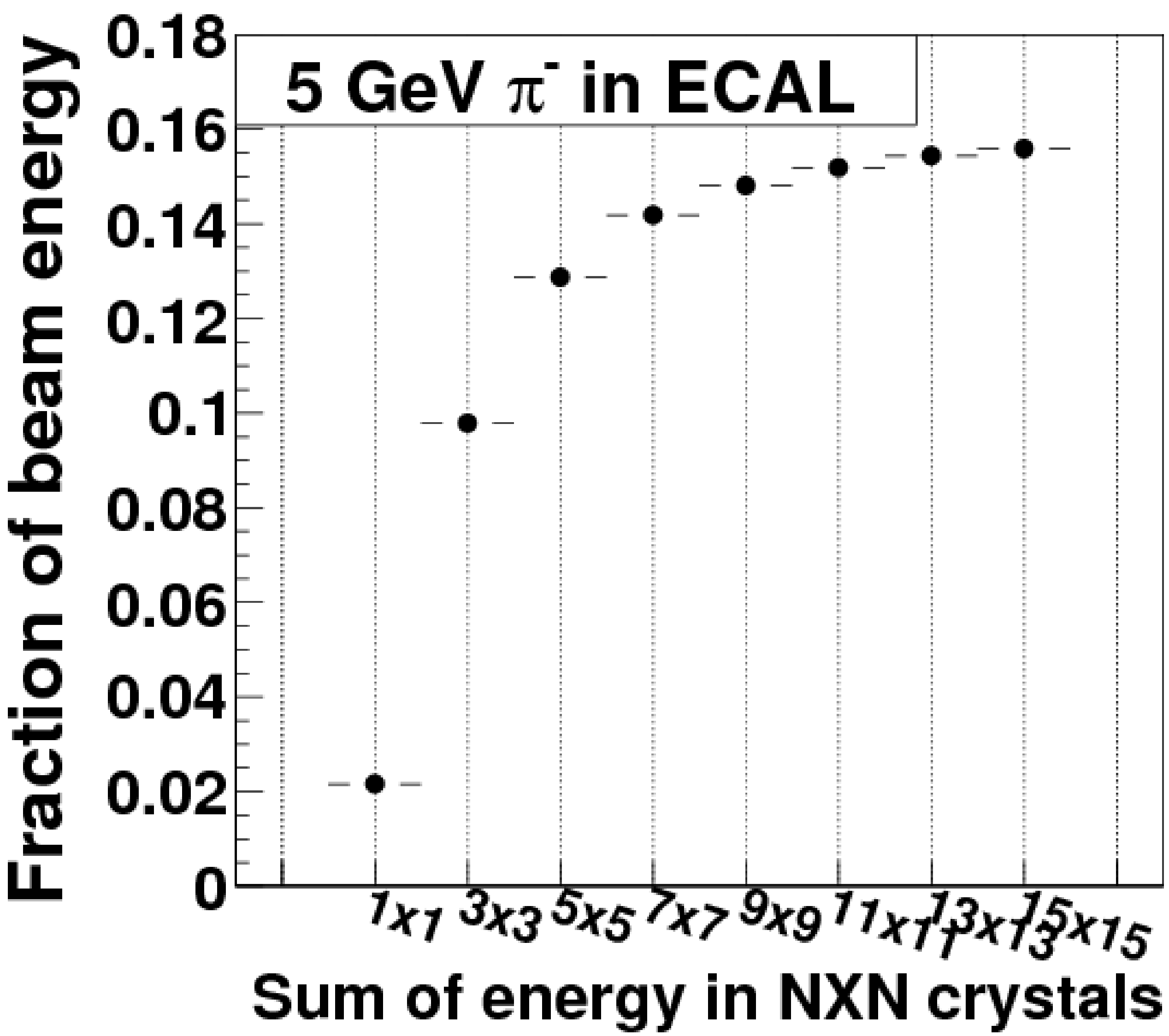
⇓

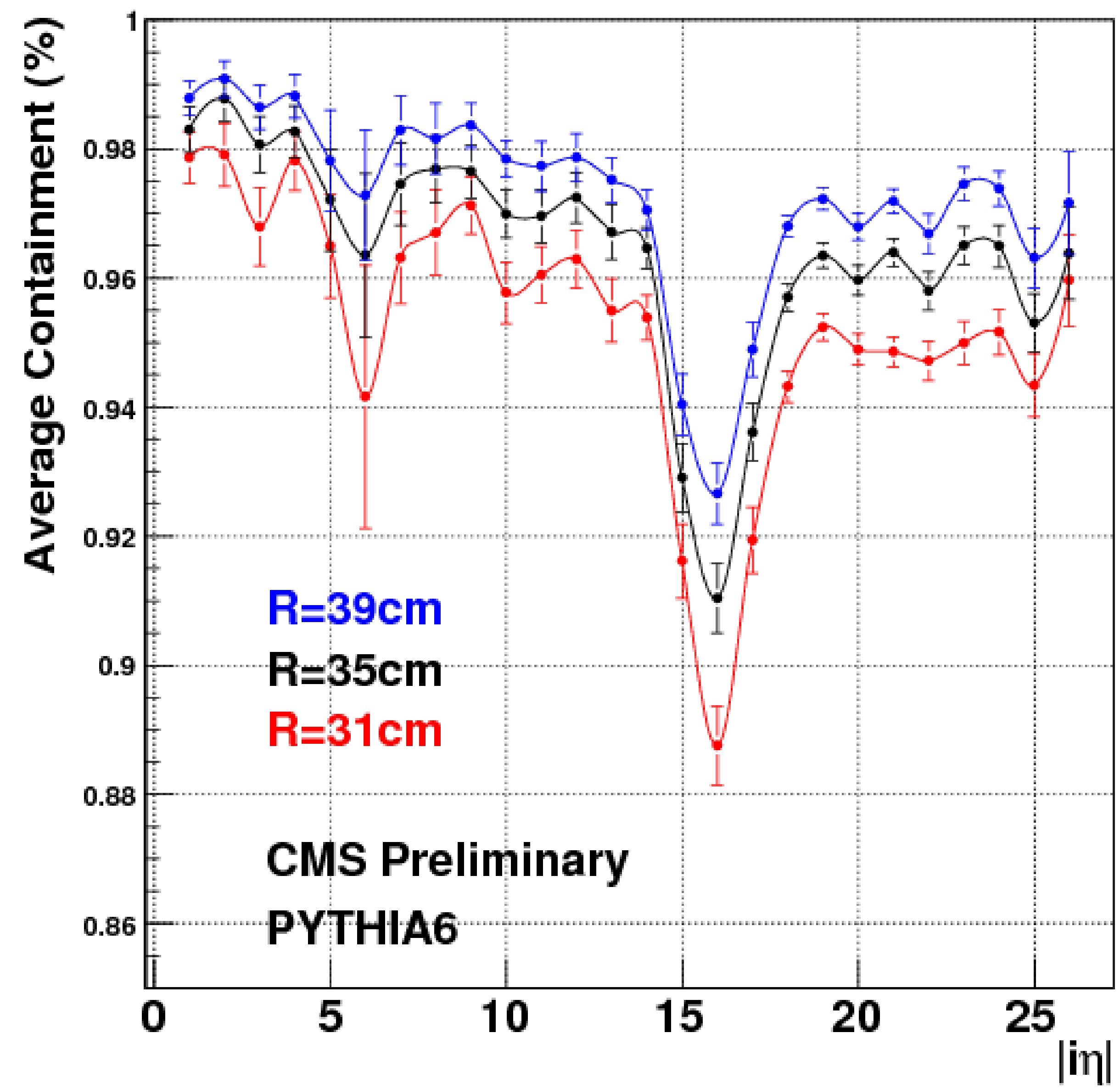
**Region based optimization**

- 22 m long, 15 m in diameter
- Over a million geometrical volumes
- Many complex shapes

# Isolation Studies

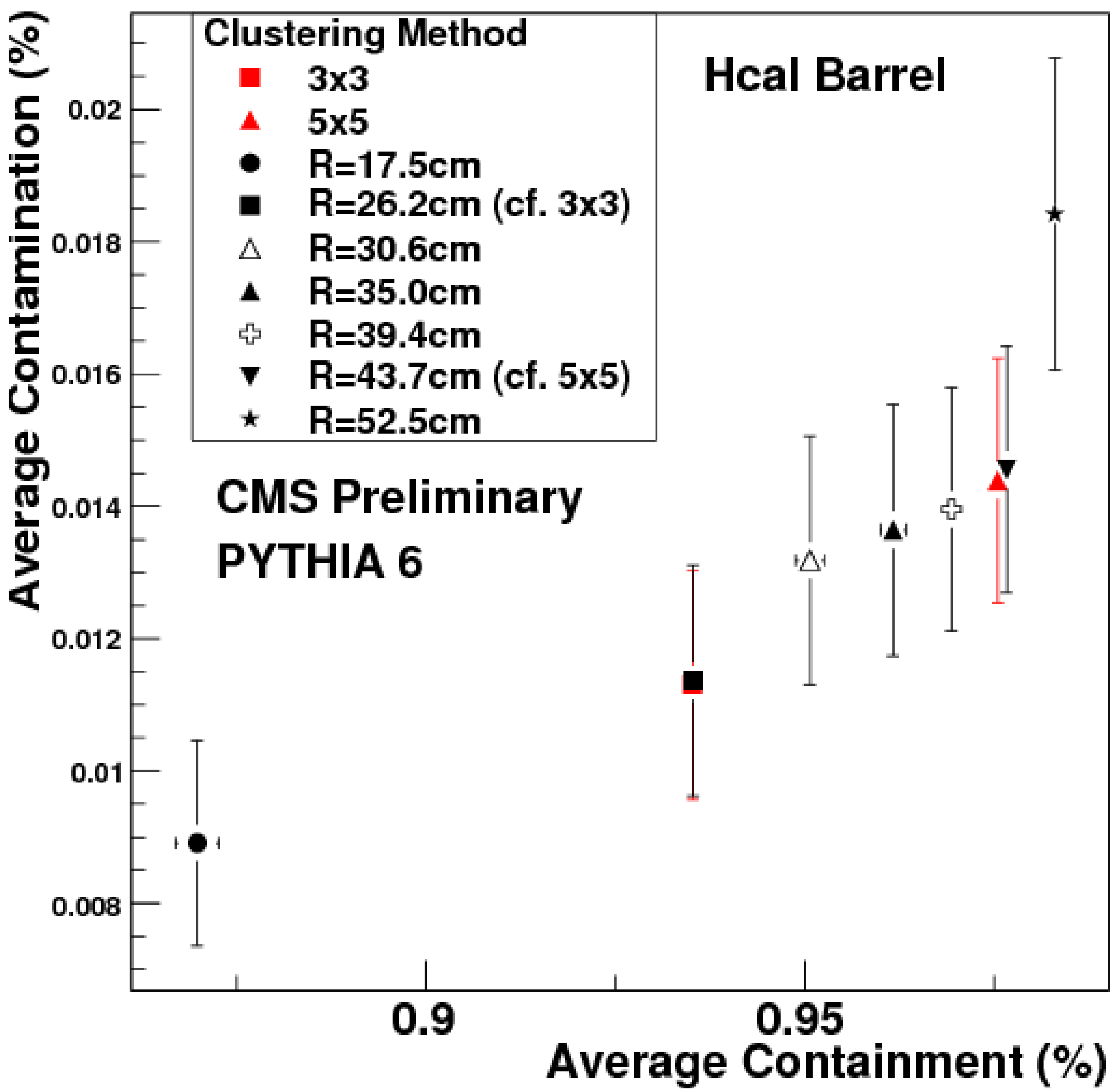
- Isolation was evaluated from MC studies (CMS AN-2010/179)
  - Charge particle isolation was studied from the lateral shower size
  - Neutral isolation came from a comparison of single particle MC and minimum bias MC

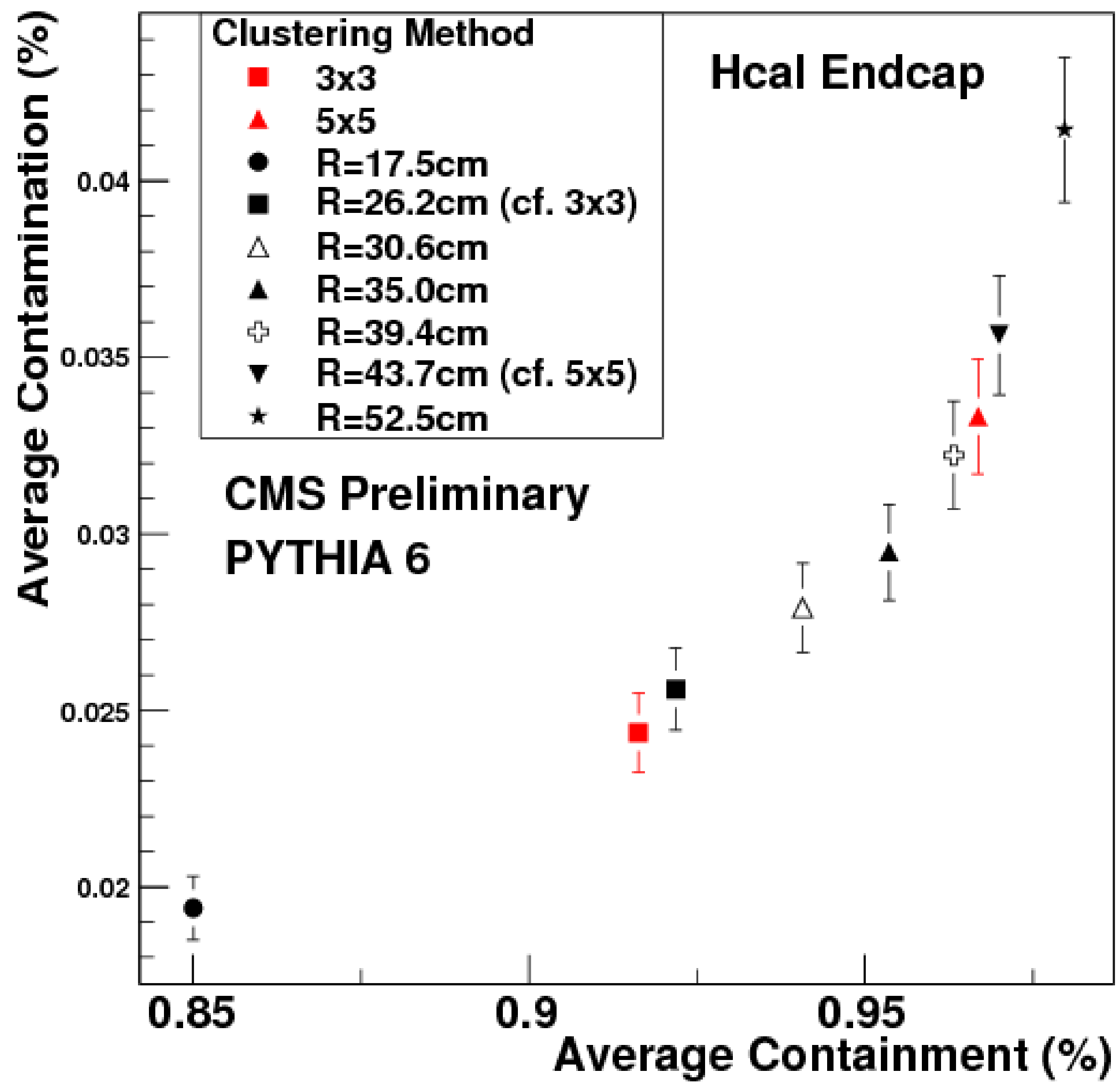


# Containment vs Contamination

- Comparison of containment vs contamination decides the size of signal and isolation zones (CMS DN-2012/002)

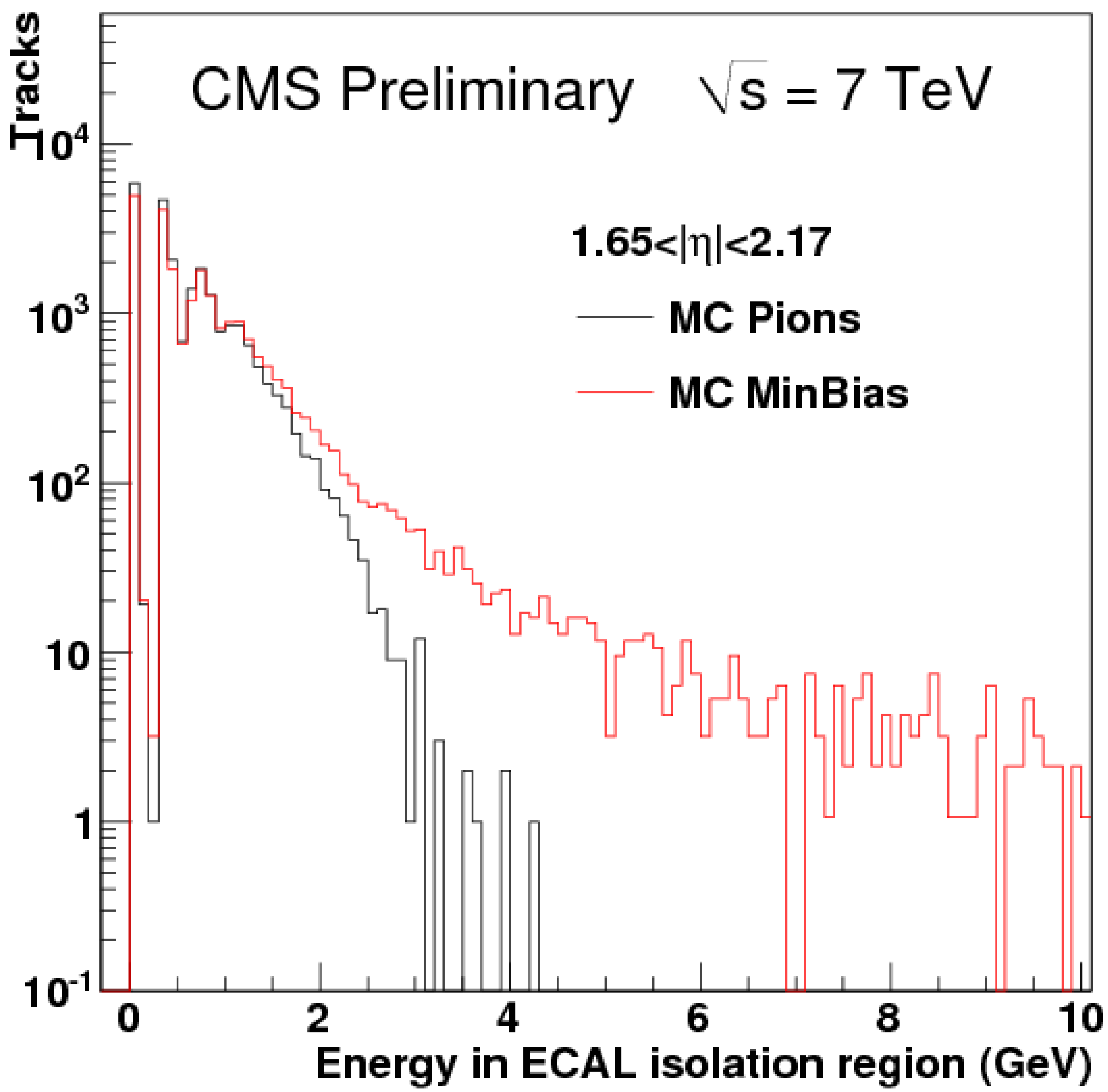


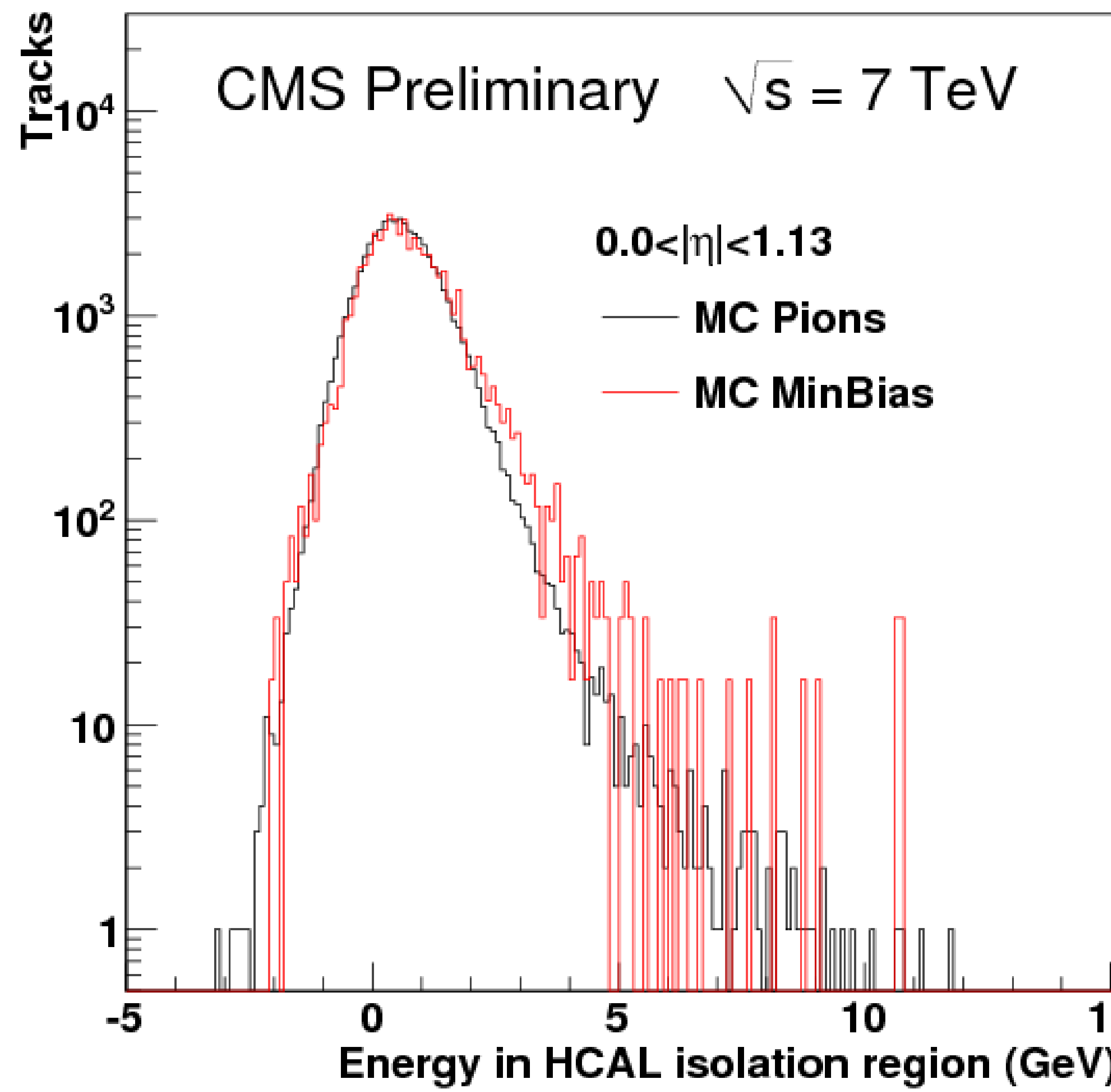



# Isolation Criteria

- For charge isolation, look at the momentum of highest energy charged particle reaching the isolation region around the extrapolated point on calorimeter surface
- The isolation region for neutrals is chosen using energy deposit in the annular region
  - 15x15 - 11x11 for ECAL
  - 7x7 - 5x5 for HCAL

CMS PAS JME-10-008

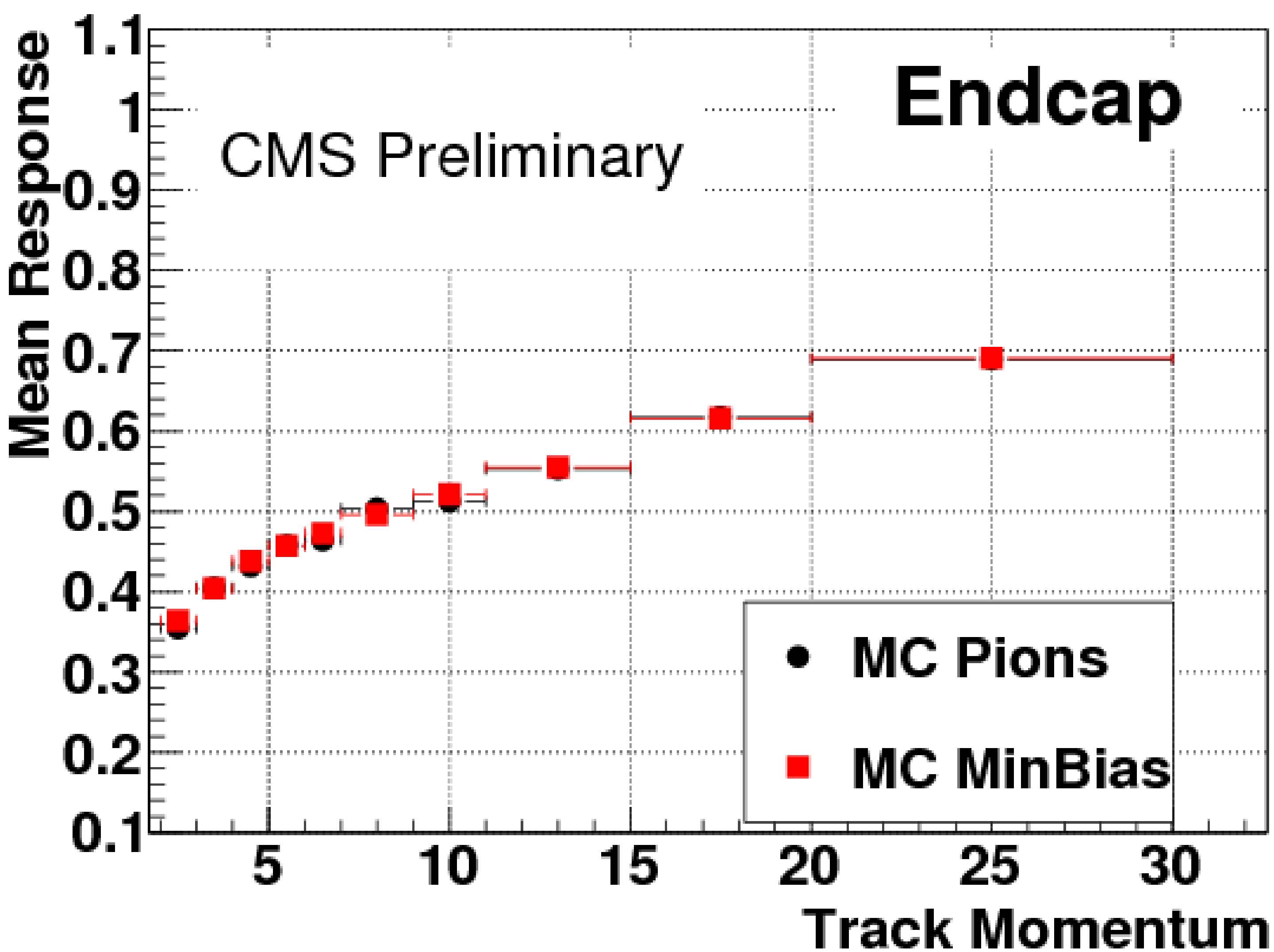


# Closure Test

- Final choice is tested by comparing single particle with minimum bias MC
- Not very critical with 100% pions or a mixture of π/K/p (+anti-particles)

(CMS AN-2010/179)



# Particle Types

- Studied MC sample as a function of track momentum (in the context of calibration of hadron calorimeter)

(CMS DN-2014/033)