Validation and Tuning of the CMS Full Simulation

Mike Hildreth

Université de Notre Dame du Lac & Fermilab

Representing the CMS Collaboration
Overview of CMS Simulation

Physics Generators

Particle 4-vectors

Geometry/Material Description

Geant 4

Simulated Hits

Simulated Hits from Pileup Interactions

Noise Model

Electronics Simulation

Simulated Raw Data

---

(1) 20 October, 2010

Mike Hildreth - CHEP 2010, Taipei, Taiwan
Overview of CMS Simulation

Physics Generators

Particle 4-vectors

Geant 4

Geometry/Material Description

Simulated Hits

Simulated Hits from Pileup Interactions

Noise Model

Electronics Simulation

Simulated Raw Data

Raw Data is processed in a manner identical to real data. Data/MC agreement is an aggregate result.
Overview of CMS

Barrel Muon System
Endcap Muon System

Solenoid
Hadronic Calorimeter
Electromagnetic Calorimeter
Inner Central Tracker
Outer Central Tracker

20 October, 2010
Mike Hildreth - CHEP 2010, Taipei, Taiwan
Overview of CMS

Example of inner detector:

The tracker geometry is quite complicated…

[Root graphical version of G4 geometry shown]
Overview

• Status of Tracker Geometry & Material Description
  – Material studies
    • photon conversions/nuclear interactions
  – Tracker dE/dx results
  – Track distributions

• Calorimeter Modeling
  – Electron bremsstrahlung
  – Jet and Missing Energy studies

• Muon System
  – Hit patterns and isolation variables

• Future Prospects

For other results, see:
Validation of Geant4 Physics Models with LHC Collision Data (PS08-1-170), Sunanda Banerjee
Material distribution in current CMS Tracker (estimated):

- Very large photon conversion probability
- Large effects of multiple-scattering
- Must test with data to validate simulation

20 October, 2010
Mike Hildreth - CHEP 2010, Taipei, Taiwan
Detector Material Studies

Reconstruction of Photon Conversions and Nuclear Interactions allow a mapping of the material distribution in the detector

Reminder - Photon conversion probability in a thin cylindrical shell:

\[ dN_{\text{conv}} = N_\gamma(R, \theta, \phi) \cdot R^2 \sin \theta \, d\theta \, d\phi \frac{P}{X_0} \, dR \]

\[ N_\gamma(R, \theta, \phi) \propto \frac{1}{R^2 \sin \theta} \]

For Nuclear Interactions:
– swap \( P(\text{photons}) \sim 7/9 \) to \( P = 1 \), \( X_0 \rightarrow \lambda_0 \)

(But, \( X_0 \) and \( \lambda_0 \) are sensitive to different physics)

• Different reconstruction characteristics:

Photons:
\[ m = 0: \text{tracks parallel at point of production} \Rightarrow \text{larger uncertainty in radial position of production point} \]

Nuclear interactions:

Good vertex resolution, many soft tracks with large impact parameters
Some examples:

Note the superior position resolution of the nuclear interaction data.

The beampipe isn’t centered!
Extracting the material budget

photon conversions

nuclear interactions

can unfold this distribution using estimates of the photon position resolution
Extracting the material budget

photon conversions

astonishingly good agreement between data and simulation

can unfold this distribution using estimates of the photon position resolution
Extracting the material budget

- Other methods also employed:
  - track multiple scattering, momentum scale, etc.
- Agreement between photon conversions and nuclear interactions on the location and composition of materials gives us good confidence that the simulation geometry is an accurate representation of the real detector
- Uncertainties in the amount of material and its distribution are estimated to of order 5%  

(CMS PAS: TRK-10-003)
Tracker dE/dx Simulation

- Signal simulation in tracker includes charge propagation, charge collection efficiencies, saturation effects, and tracker noise modeling
  - tuned on cosmic data and early collisions
- Detailed test of Geant4 descriptions of energy loss mechanisms in tracker material

CMS Preliminary 2010 $\sqrt{s} = 7$ TeV

production of deuterons not simulated in Pythia

CMS PAS: EXO-10-004

(12) 20 October, 2010  Mike Hildreth - CHEP 2010, Taipei, Taiwan
Charged particle multiplicity

- Minimum Bias events
- Original Pythia 6.4 tunes largely divergent from Data distributions (tune D6T)
  - charged particle multiplicity very different
  - Surprising, given previous Tevatron studies
Charged particle multiplicity

- Minimum Bias events
- Original Pythia 6.4 tunes largely divergent from Data distributions (tune D6T)
  - charged particle multiplicity very different
  - Surprising, given previous Tevatron studies
Charged particle multiplicity

- Minimum Bias events
- Original Pythia 6.4 tunes largely divergent from Data distributions (tune D6T)
  - charged particle multiplicity very different
  - Surprising, given previous Tevatron studies
- New Pythia 8, Tune 1 gives much better agreement
  - new: hard scattering in diffractive interactions
    - relative increase in population of high-$p_T$, “forward” regions
Based on expected $\phi$ symmetry of energy deposition in minbias events

- Non-uniformity of response correction caused by inter-module gaps and different distributions of material in front of the calorimeter

Here: response correction for each calorimeter module:

\[ \int L \, dt = 250 \, \text{nb}^{-1} \]
Electromagnetic Interactions

- Fraction of energy loss in the Ecal for different ranges in $\eta$: $f_{\text{Brem}} = (p_{\text{in}} - p_{\text{out}})/p_{\text{in}}$
  - inclusive distributions based on high-purity track selection
- Gaussian Sum Filter track fit to account for energy loss

- MC Minbias events; again remarkable MC/Data agreement
- depends on accurate modeling of:
  - material distributions
  - showers
  - correct distribution of particle types in Data and MC

CMS PAS: EGM-10-001
Jet-Finding at CMS

- **Calorimeter Jets**
  Jets clustered from ECAL and HCAL deposits (Calorimeter Towers)
  Correspondingly: Calo MET

- **Particle Flow Jets (PF)**
  Cluster derived Particle Flow objects: unique list of calibrated “particles” representing “generator level” information
  Correspondingly: PFMET

- **Jet-Plus-Track Jets (JPT)**
  Subtract average calorimeter response from CaloJet and replace it with the track measurement
  Correspondingly: Tc MET

Default jet clustering algorithm: Anti-$k_T$ with $R = 0.5$
Jet Variables

• Inclusive $p_T$ spectrum

• Jet $\eta$

(19) 20 October, 2010  Mike Hildreth - CHEP 2010, Taipei, Taiwan
Jet Resolutions

- Resolution as a function of average $p_T$
  - extrapolated to zero additional activity in each bin by measuring $\sigma_A$ at decreasing values of the third jet $p_T$
  - same treatment applied to QCD MC
  - within 10% agreement for all three jet algorithms
  - validates combination of generators, material description, shower models

Di-jet Asymmetry Method:

$$A = \frac{p_T^{jet1} - p_T^{jet2}}{p_T^{jet1} + p_T^{jet2}}$$

$$\frac{\sigma(p_T)}{p_T} = \sqrt{2}\sigma_A$$
MET provides a stringent test of noise simulation, showering, and resolution modeling ⇒ all elements have to be correct

Here, $E_T$ resolution is measured using the width of the $E_x$, $E_y$ distributions

- overall $E_T$ calibration from transverse energy balance in $\gamma$+jet events
- at least two jets of $p_T > 25$ GeV required
- identical MC/Data treatment

(Jet Energy Scale corrections are applied to all jets with $p_T > 30$ GeV; remaining unclustered energy is corrected using a scale derived from the hadronic recoil opposite $Z \rightarrow ee$ events)
Comparisons of Calorimeter Resolutions

- Resolution of MET and $H_T$ (total jet transverse energy) for Calorimeter Jets and Particle Flow Jets
  - $H_T$ potentially more robust
  - multi-jet events
    - leading jet $p_T > 40$ GeV
    - width of central gaussian in $H_x, H_y$ and $E_x, E_y$

- Characterization of the non-Gaussian nature of the tails:
  - important for searches
  - plot shows width of data distributions (in sigma) containing $n\sigma$ of a gaussian
  - deviation from gaussian form outside of $2\sigma$

(22) 20 October, 2010

Mike Hildreth - CHEP 2010, Taipei, Taiwan
Muon System

- Energy deposition characterized by proper modeling of the absorber interaction lengths
  - punch-through, decays in flight account for much of the fakes
  - isolation variable critical to differentiate signal from QCD
  - future: significant backgrounds from neutron interactions

![Comparison of isolation-inverted data sample with expectations from QCD MC](image1)

![Sum of track, Ecal, Hcal energy in cone of ΔR < 0.3 around muon](image2)
Electroweak Distributions

- MET also well-modeled in $W \rightarrow \mu \nu$ events

- Here: combination of intrinsic resolution and generator models

(24) 20 October, 2010  Mike Hildreth  - CHEP 2010, Taipei, Taiwan
Conclusions

• “Tuning” of CMS Simulation has been a multi-year process
  – based on Test Beam data, extensive Cosmic running
  – improved by comparisons using collision data
    • no substitute for the real thing…
• “Validation” ongoing
  – continual refinement as the dataset grows
    • higher statistics comparisons possible for a growing number of studies

• Current (excellent) level of Data/MC agreement is a product of a huge amount of work over many years by many people
  – not an accident!