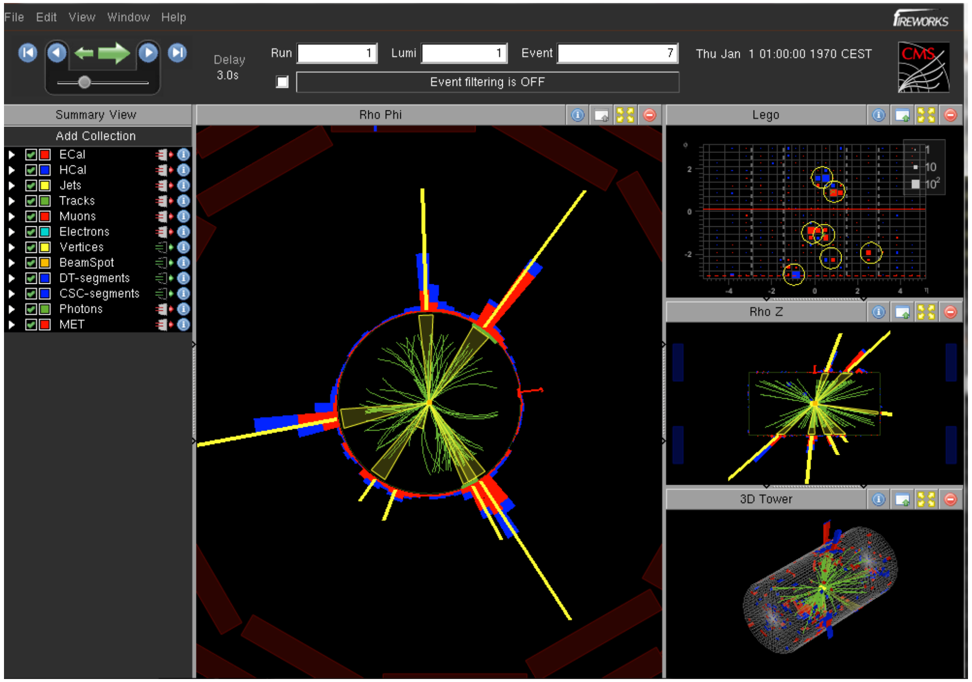
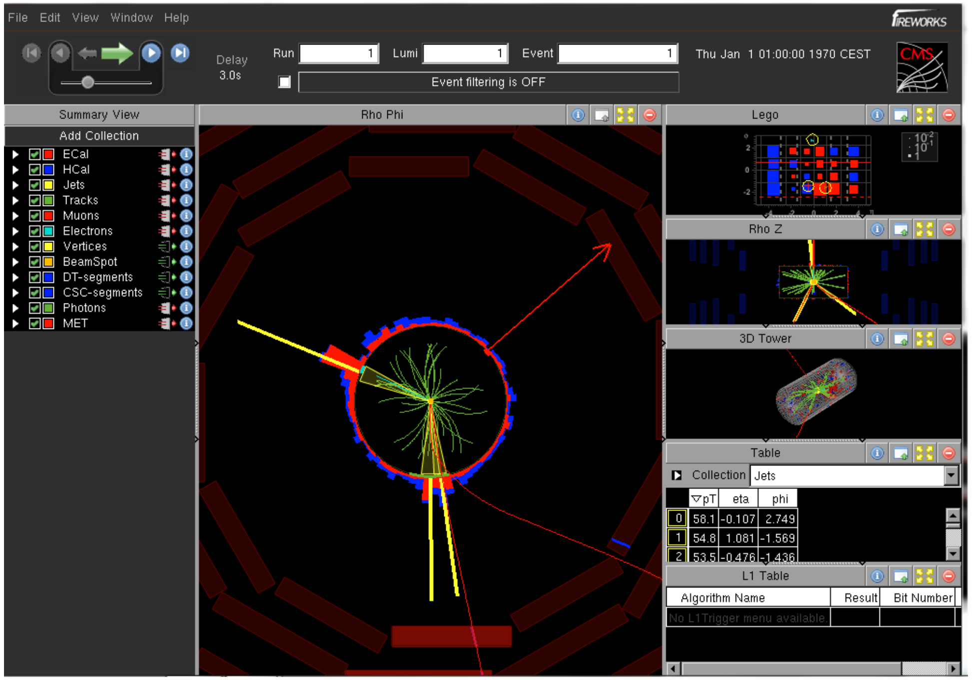
The CMS Fast Simulation

<u>A.Giammanco</u> (Louvain & NICPB)
A.Perrotta (INFN-Bologna)



Delay 3.0s
Run 1 Lumi 1 Event 7
Thu Jan 1 01:00:00 1970 CEST
Event filtering is OFF



Delay 3.0s
Run 1 Lumi 1 Event 1
Thu Jan 1 01:00:00 1970 CEST
Event filtering is OFF

Collection: Jets

	▽pT	eta	phi
0	58.1	-0.107	2.749
1	54.8	1.081	-1.569
2	53.5	-0.476	-1.436

Algorithm Name	Result	Bit Number
No L1Trigger menu available.		

One of these is GEANT, the other is CMS Fast Simulation;
can you guess which is which?

Why do we need a fast simulation?

- Because we need very large amounts of MC:
  - To evaluate **backgrounds** with large cross sections and small survival probability (e.g., multi-jet QCD); we can filter at RECO level!
  - To **scan a model's parameters space** or evaluate **systematics**
  - To **train MVAs** (e.g., NN) with sufficient statistics
  - To develop and test efficiently reconstruction and analysis algorithms
- Some examples of crucial use in CMS:
  - Top mass extraction in 2l final states, CMS-PAS-TOP-10-006 and J. High Energy Phys. 07 (2011) 049; used for mass templates
  - Black Hole search, CMS-PAS-EXO-10-017 and CMS-PAS-EXO-11-021 and Phys. Lett. B 697 (2011) 434-453; used for signal samples with different BH models and masses
  - Most 2011 SUSY analyses (1 submitted to PRL, 8 public PAS, 6 more in the pipeline); used for scans of Simplified Model Signatures

The CMS Fast Simulation

- OO subsystem of the CMS C++ based software
- Alternative to GEANT-based approach (aka Full Simulation)
- Much more ambitious than a typical fast simulation (à la PGS, DELPHES or the old ATLFAST):
  - We do a realistic simulation of low-level objects (hits, clusters)
  - On these we apply the same high-level modules (trigger, lepton ID, jet finding, b/τ-tagging, isolation ...) as in FullSim and data, keeping a comparable level of accuracy as FullSim
  - The only case where reconstruction is customized is tracking
- CPU time for **ttbar + "early 2011" pile-up**:
  - ~120x gain in the pure simulation part (much more for simpler events)
  - ~2.5x gain in reconstruction, thanks to FastTracking

Interactions

The interactions simulated in the Fast Simulation are

1) electron Bremsstrahlung;

2) photon conversion;

3) charged particle energy loss by ionization;

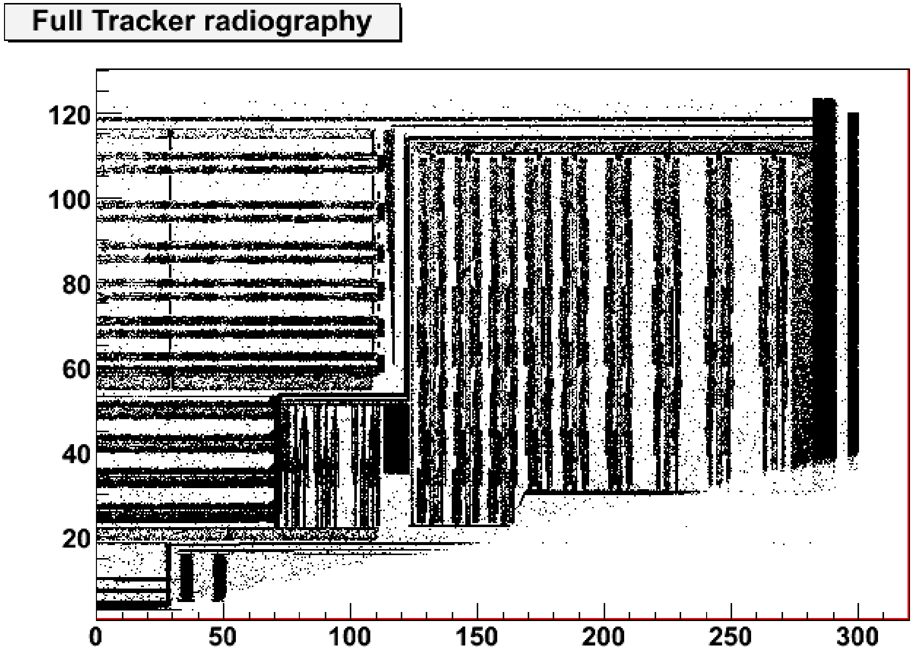
4) charged particle multiple scattering

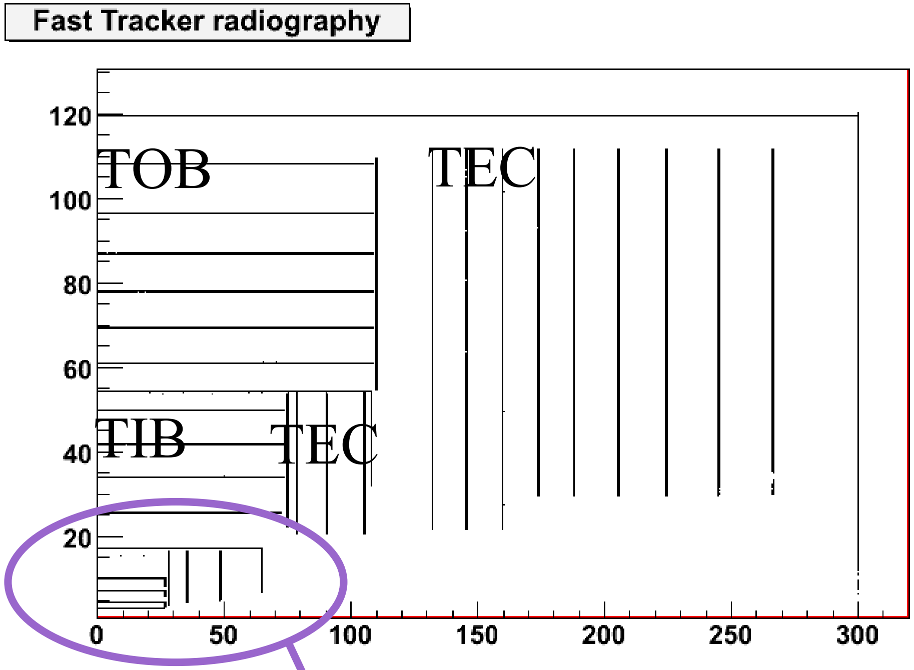
5) nuclear interactions;

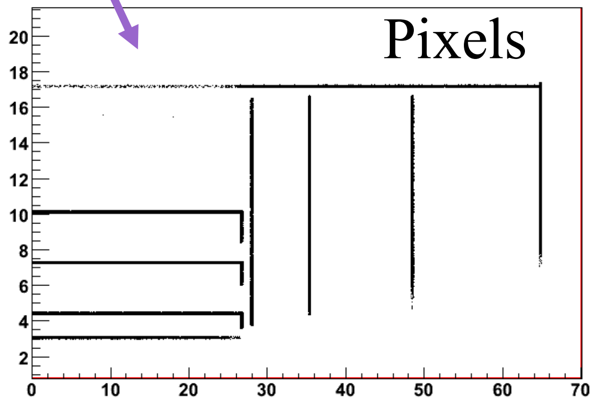
6) electron, photon, and hadron showering.

The first 5 are applied to particles crossing the thin layers of the tracker, while the latter is parameterized in the electromagnetic and hadron calorimeters. Muons propagate through the tracker, the calorimeters and the muon chambers, with multiple scattering and energy loss by ionization taken into account in the propagation.

Tracker material, track propagation


Full Tracker radiography

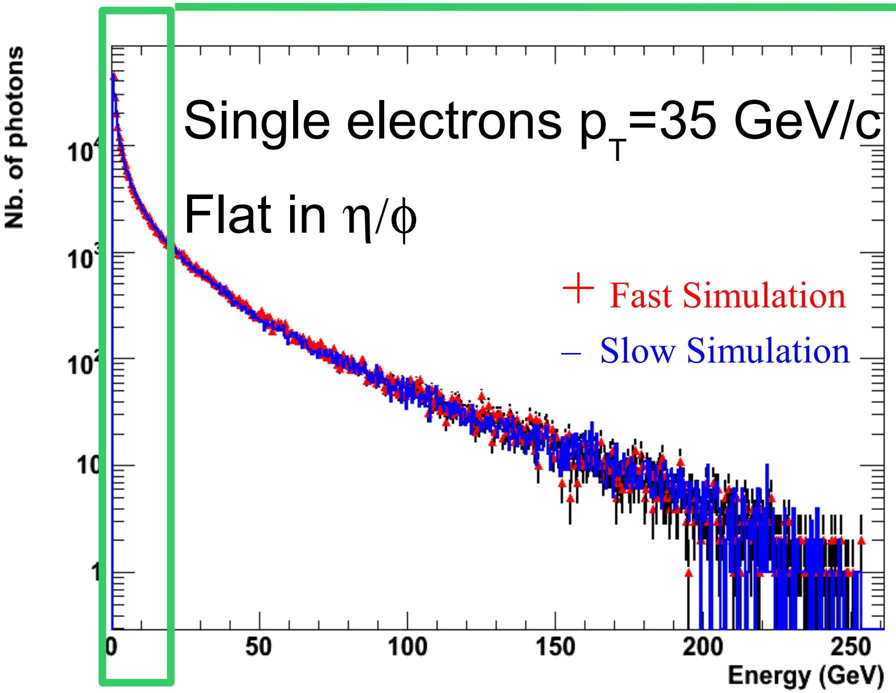

Fast Tracker radiography



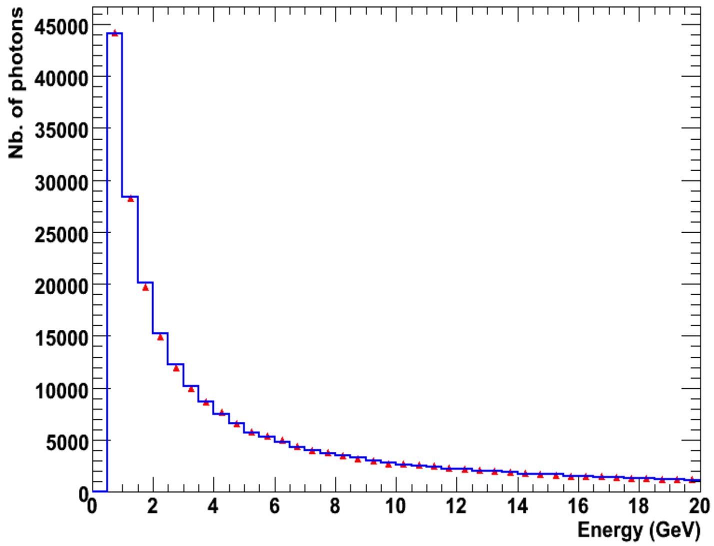
- A simplified interaction geometry is used !
  - with some details though!
    - Active and passive layers are modelled
- The complete magnetic field map is used for the track propagation between two surfaces
- Dead modules in the Pixels have been recently added; taken from the same database as full sim (not included in the plots shown today)

Tuning layer thickness

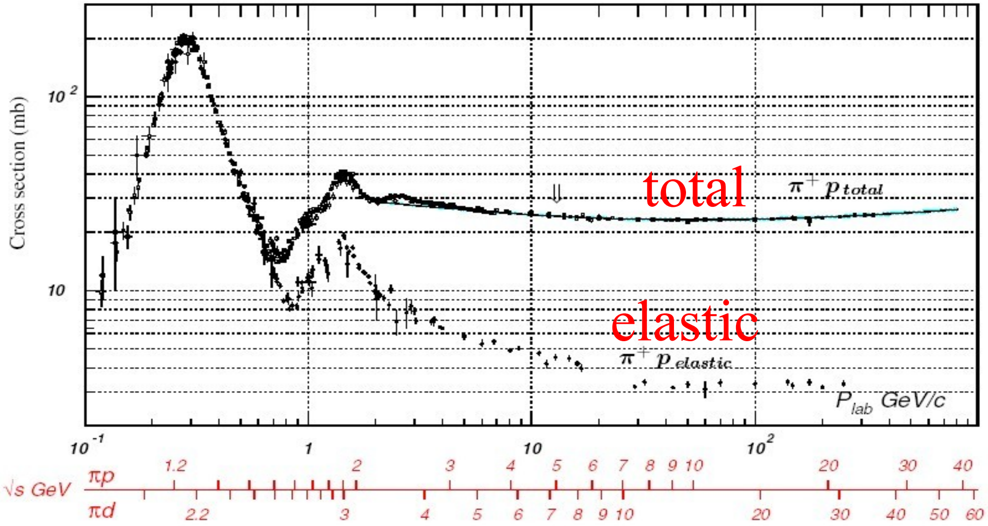
- The Brem photon emission probability and spectrum are calculated analytically, layer by layer
- The layer thickness is tuned to reproduce the number of photons in the GEANT-based simulation:
  - the photon energy spectrum is beautifully reproduced...
  - (incidentally , this tuning reproduces the actual layer thickness in x/x_0)



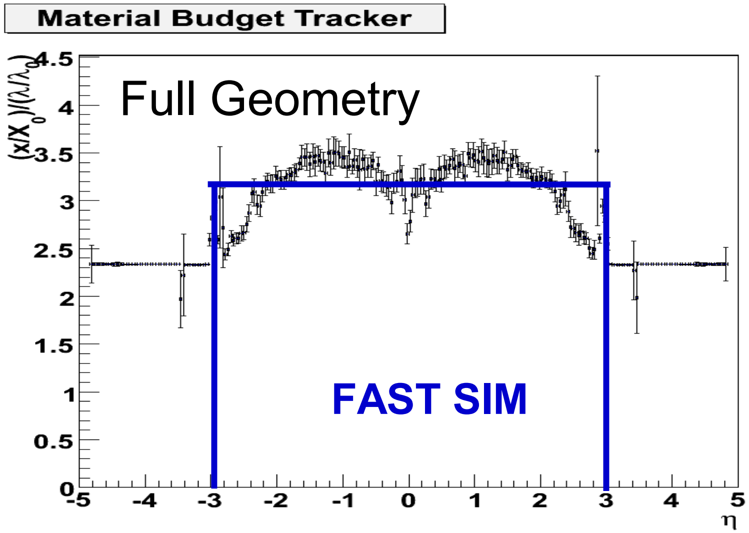
Absolute normalization !



Nuclear interactions (1)



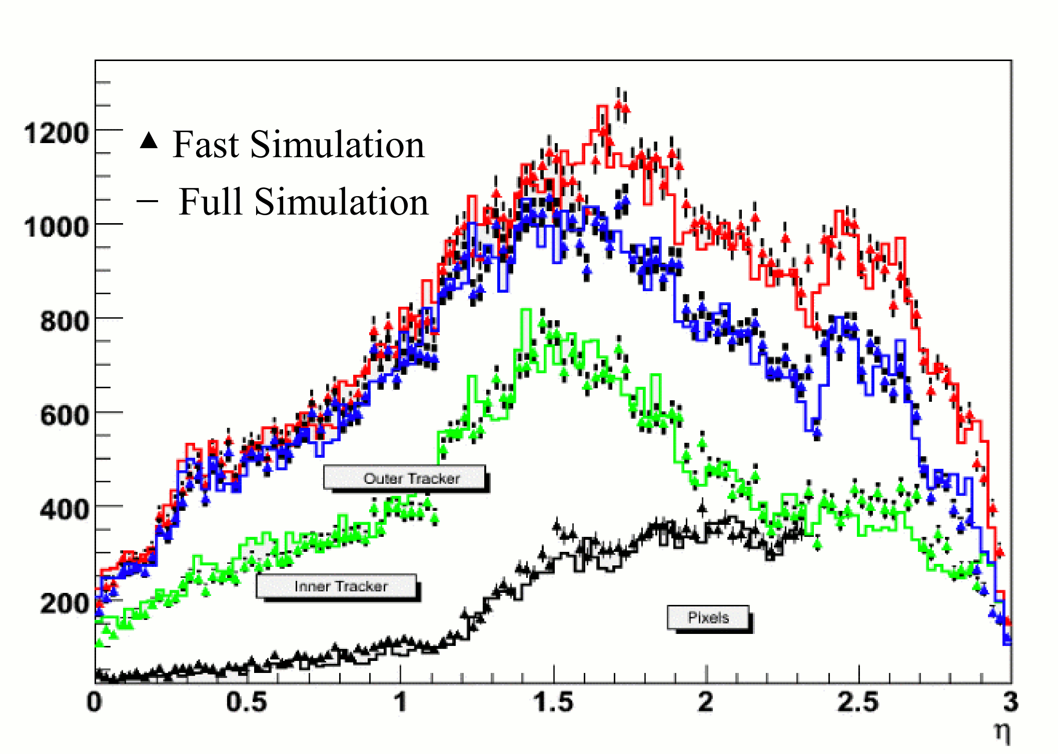
The elastic and inelastic cross sections come from experimental measurements (PDG)

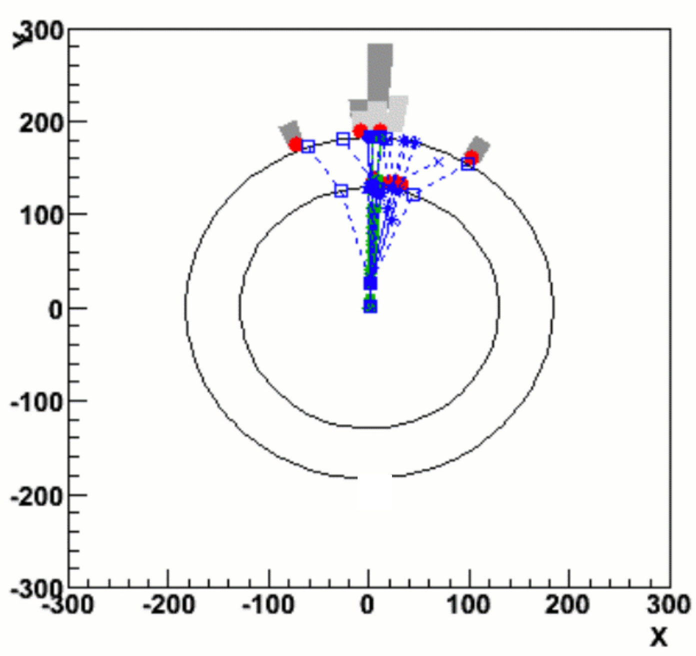


- The tracker layer thickness is expressed in terms of λ/λ_0
  - 0.31 x/X_0 (total) or 0.25 x/X_0 (inelastic)
  - (not strictly true, but good approximation in the tracker acceptance)
- Data files of inelastic N.I have been created
  - 2.5 million N.I saved, 9 different hadrons, 1<E<1000 GeV
  - when a N.I occurs, a N.I is picked up randomly in the relevant energy range
  - a rotation around the particle direction is made (extra randomness)

Nuclear interactions (2)

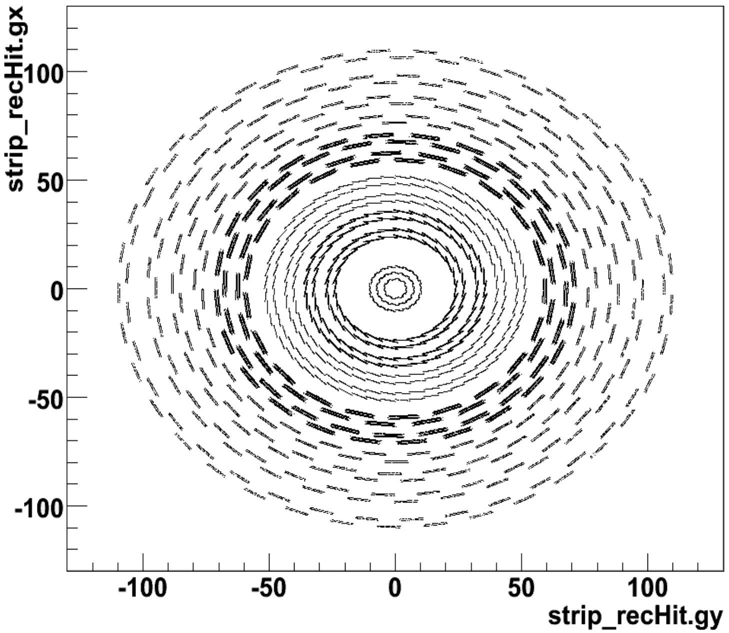
Number of nuclear interactions for 500K 15 GeV pions

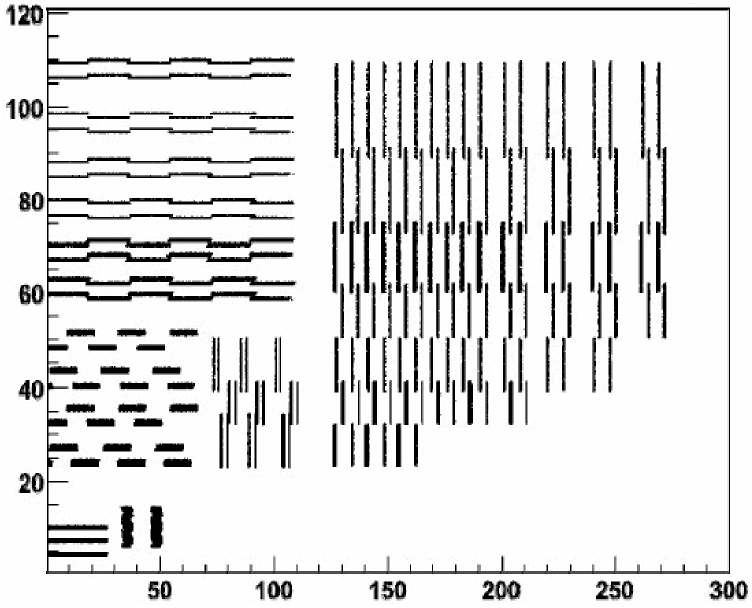




A single tau event where a pion undergoes a nuclear interaction in the tracker (fast simulation)

Tracking SimHits

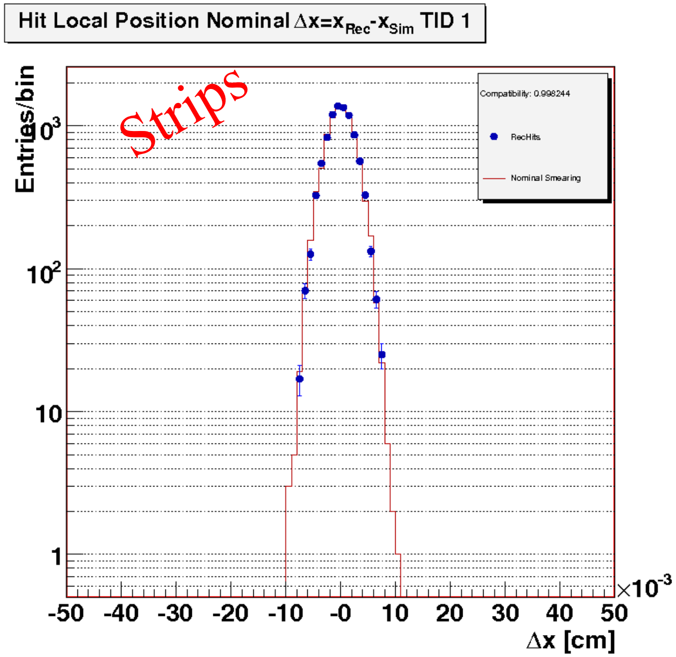




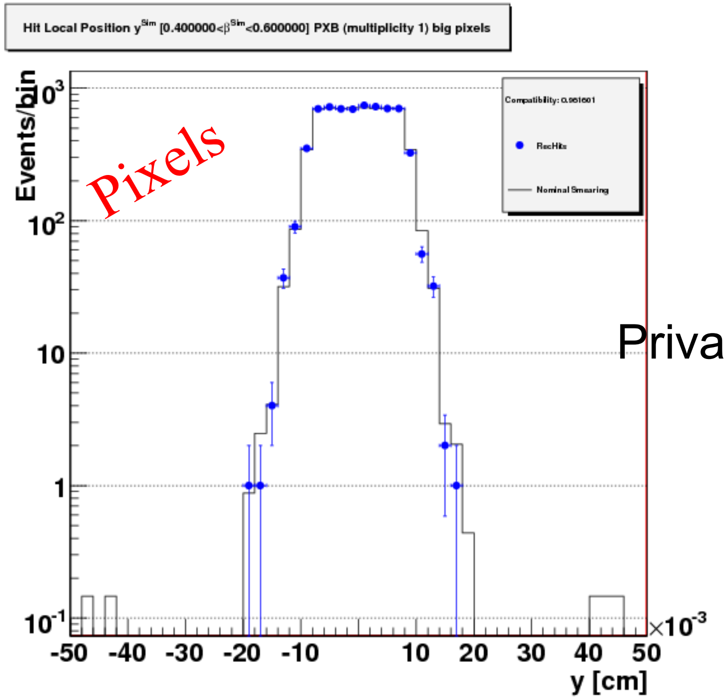
- The hits are located on the detailed tracker module geometry (propagation to the closest active layer modules)
  - create a SimHit if an intersection exists
  - this allows the mis-alignment to be simulated

Tracking RecHits smearing

- The SimHits are then smeared
  - a layer-dependent Gaussian smearing is applied in the strips
  - in the pixels, the smearing is done according to cluster-multiplicity- and incidence-angle-dependent position resolution distributions (obtained from the Full Sim, might be taken from data)
- ➡ the result is turned into tracking RecHits



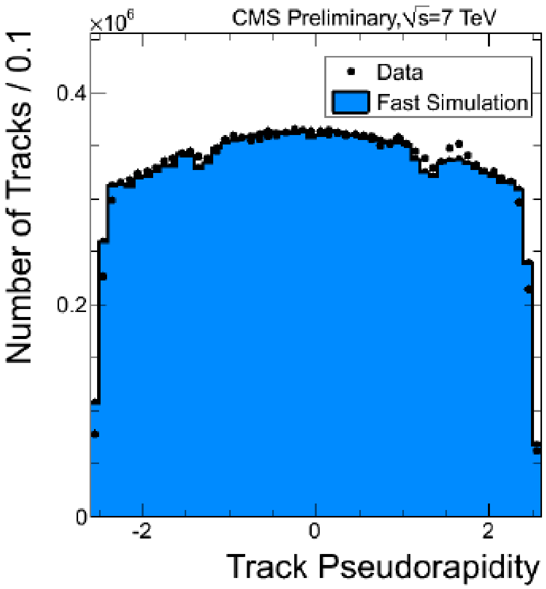
Resolution function vs. Fast Sim

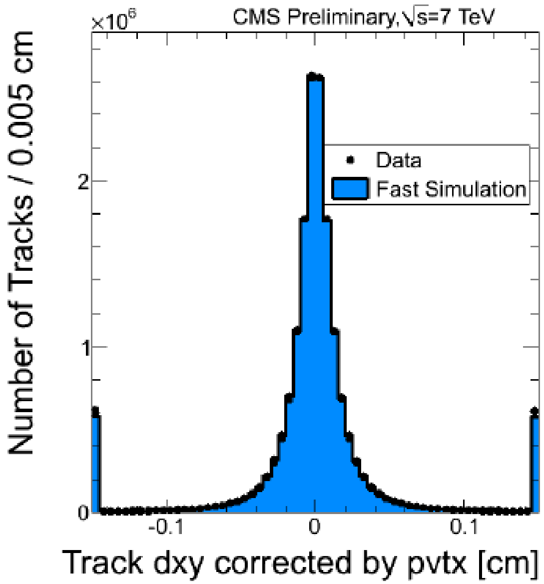


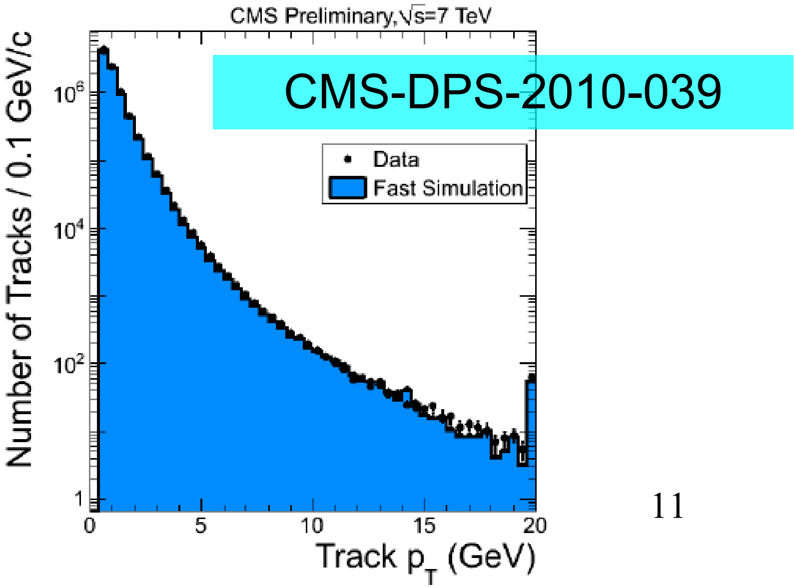
Private plots

Tracking

- We save reconstruction time with **Fast Tracking**
  - It emulates seeding efficiency (based on the hits of the MC-truth charged particle), performs fit, rejects outlier hits
  - Final track selection uses same modules as real tracking
  - No fake tracks (<1% of high-quality tracks)
  - Excellent agreement with data after basic quality cuts
- (Possibility to use standard tracking if desired)







CMS-DPS-2010-039

b-tagging (1)

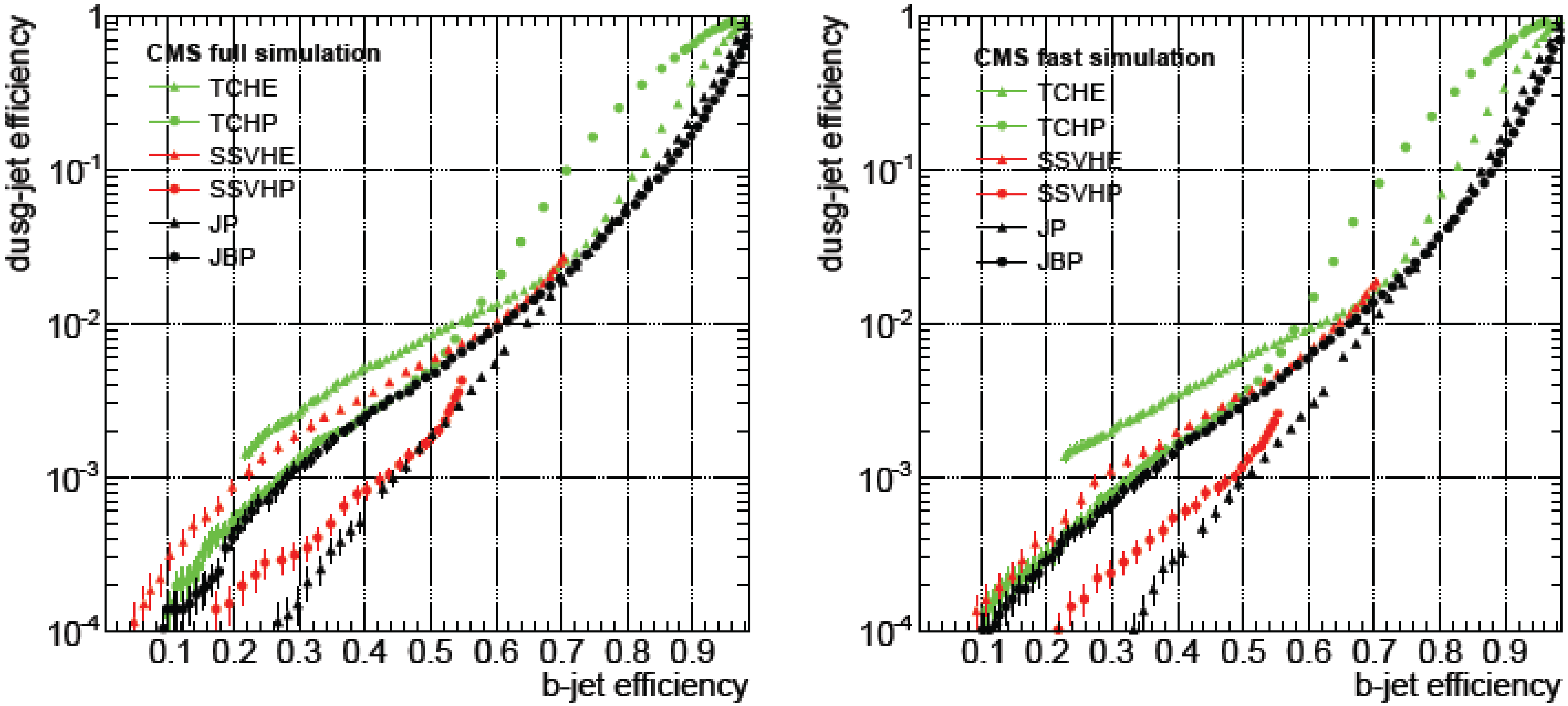
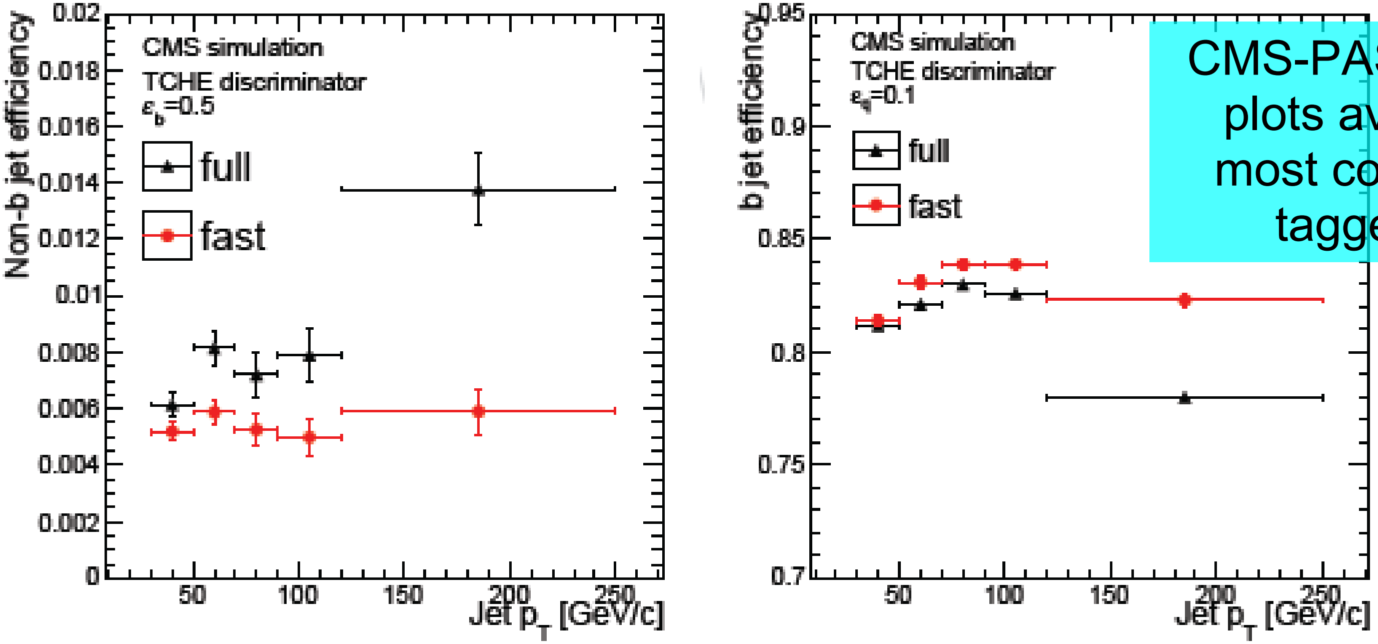


Figure 22: Light flavor mistag efficiency versus b-tagging efficiency in comparison for several b-tagging algorithms. On the left: full simulation, on the right: fast simulation.

CMS-PAS-BTV-11-002

b-tagging (2)

Fake rate: generally lower **Efficiency: generally higher**



CMS-PAS-BTV-11-002; plots available for all most commonly used taggers in CMS

Figure 24: Comparison of the b-tagging performance between full and fast detector simulation for the track counting high efficiency algorithm. Left: mistag rate versus jet p_T at fixed b-tag efficiency of 50%. Right: b-tag efficiency versus jet p_T at fixed light flavor mistag rate of 10%.

- Discrepancies attributed to:
  - No fake tracks
  - No cluster merging/splitting (important in dense high-momentum jets)
  - No dead channels, especially in the pixels (but now we added them)

Nota bene: corrections to data needed also in full simulation

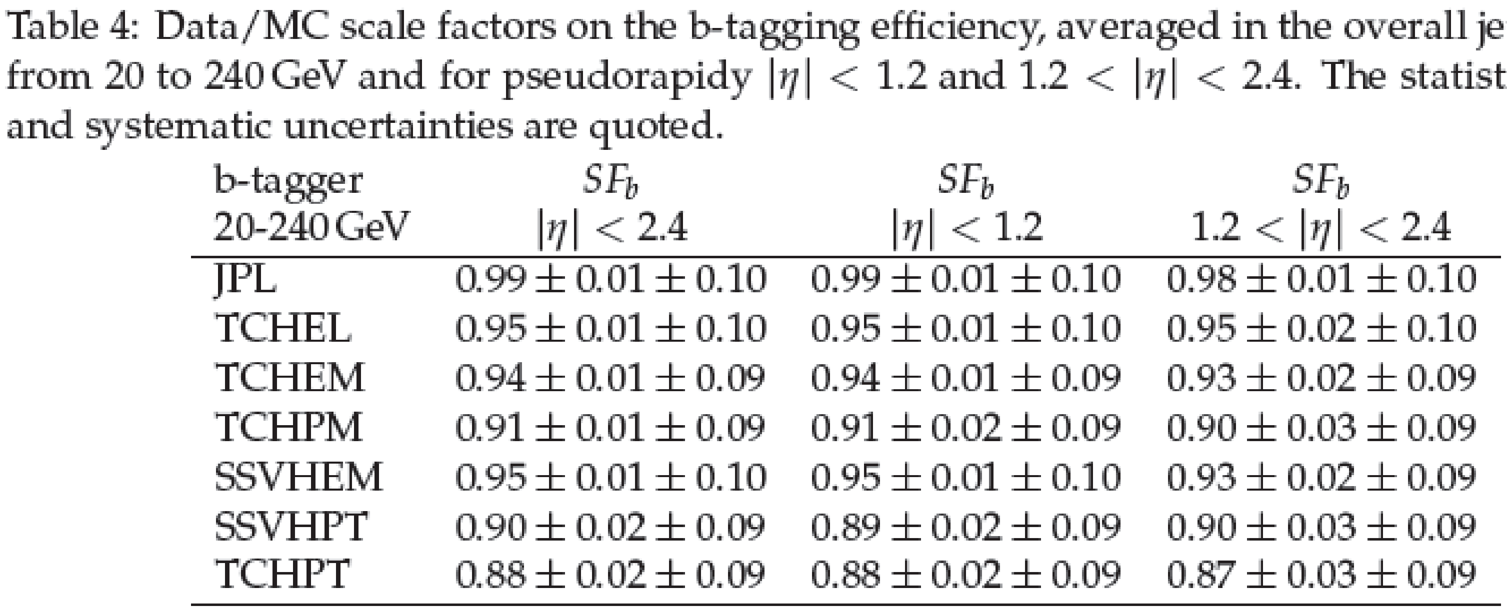


Table 4: Data/MC scale factors on the b-tagging efficiency, averaged in the overall jet p_T range from 20 to 240 GeV and for pseudorapidy $|\eta| < 1.2$ and $1.2 < |\eta| < 2.4$. The statistical errors and systematic uncertainties are quoted.

b-tagger 20-240 GeV	SF_b $\lvert\eta\rvert < 2.4$	SF_b $\lvert\eta\rvert < 1.2$	SF_b $1.2 < \lvert\eta\rvert < 2.4$
JPL	$0.99 \pm 0.01 \pm 0.10$	$0.99 \pm 0.01 \pm 0.10$	$0.98 \pm 0.01 \pm 0.10$
TCHEL	$0.95 \pm 0.01 \pm 0.10$	$0.95 \pm 0.01 \pm 0.10$	$0.95 \pm 0.02 \pm 0.10$
TCHEM	$0.94 \pm 0.01 \pm 0.09$	$0.94 \pm 0.01 \pm 0.09$	$0.93 \pm 0.02 \pm 0.09$
TCHPM	$0.91 \pm 0.01 \pm 0.09$	$0.91 \pm 0.02 \pm 0.09$	$0.90 \pm 0.03 \pm 0.09$
SSVHEM	$0.95 \pm 0.01 \pm 0.10$	$0.95 \pm 0.01 \pm 0.10$	$0.93 \pm 0.02 \pm 0.09$
SSVHPT	$0.90 \pm 0.02 \pm 0.09$	$0.89 \pm 0.02 \pm 0.09$	$0.90 \pm 0.03 \pm 0.09$
TCHPT	$0.88 \pm 0.02 \pm 0.09$	$0.88 \pm 0.02 \pm 0.09$	$0.87 \pm 0.03 \pm 0.09$

CMS-PAS-BTV-11-001: data-driven efficiencies and fake rates

Efficiencies lower in data than in both MCs; error ~±10% (covers fast-full diff.)

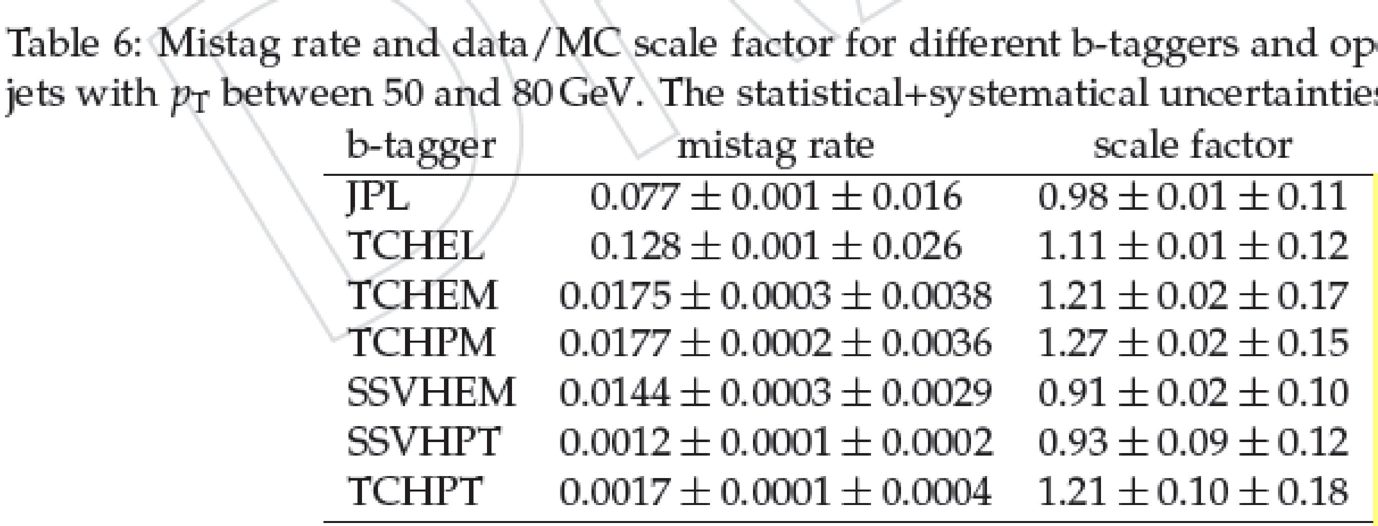


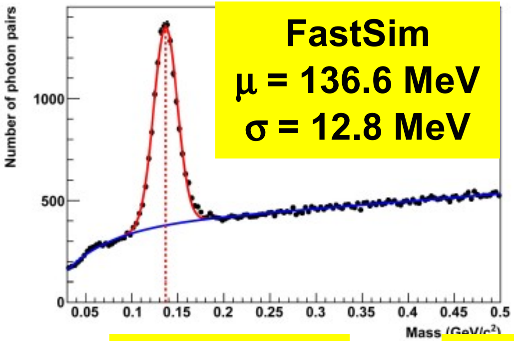
Table 6: Mistag rate and data/MC scale factor for different b-taggers and operating points for jets with p_T between 50 and 80 GeV. The statistical+systematical uncertainties are quoted.

b-tagger	mistag rate	scale factor
JPL	$0.077 \pm 0.001 \pm 0.016$	$0.98 \pm 0.01 \pm 0.11$
TCHEL	$0.128 \pm 0.001 \pm 0.026$	$1.11 \pm 0.01 \pm 0.12$
TCHEM	$0.0175 \pm 0.0003 \pm 0.0038$	$1.21 \pm 0.02 \pm 0.17$
TCHPM	$0.0177 \pm 0.0002 \pm 0.0036$	$1.27 \pm 0.02 \pm 0.15$
SSVHEM	$0.0144 \pm 0.0003 \pm 0.0029$	$0.91 \pm 0.02 \pm 0.10$
SSVHPT	$0.0012 \pm 0.0001 \pm 0.0002$	$0.93 \pm 0.09 \pm 0.12$
TCHPT	$0.0017 \pm 0.0001 \pm 0.0004$	$1.21 \pm 0.10 \pm 0.18$

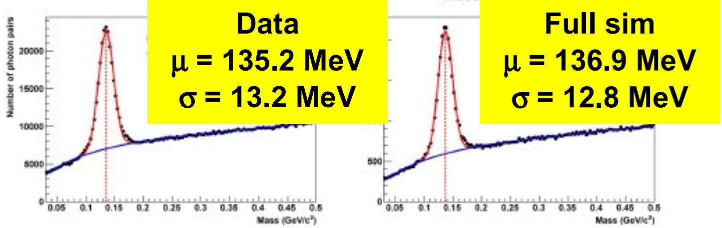
Fake rates: ~±10-20%; dedicated corrections needed (but irrelevant for main users so far: either signal has true b's, or b-tagging is not used)

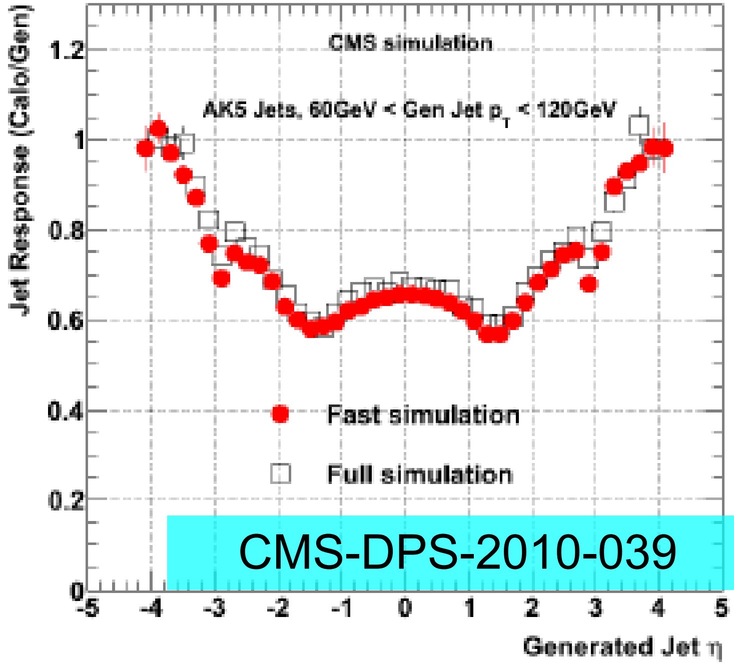
Calorimeters in FastSim

- Showers simulated à la GFLASH
- ECAL:
  - Treated as a homogeneous medium
  - Cracks, leakage, magn.field as full sim
  - Noise, zero suppression as in full sim
- HCAL:
  - Response and resolution tuned to single pions in full sim
  - Validated originally with test-beam data, now also with isolated tracks
- We also apply realistic miscalibration



Di-photon invariant mass in the ECAL barrel (photons reconstructed with the Particle Flow algorithm)

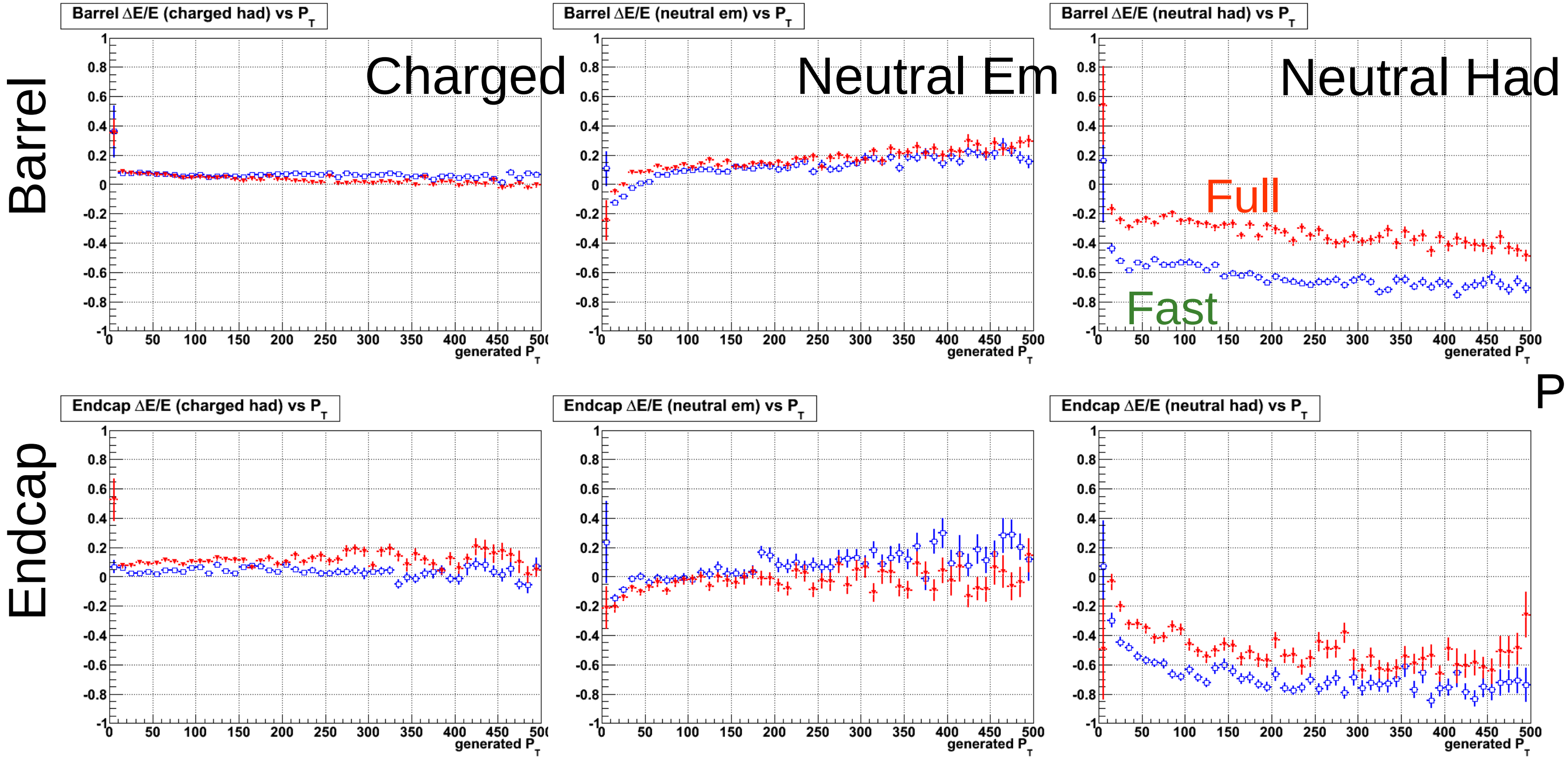




CMS-DPS-2010-039

Particle Flow

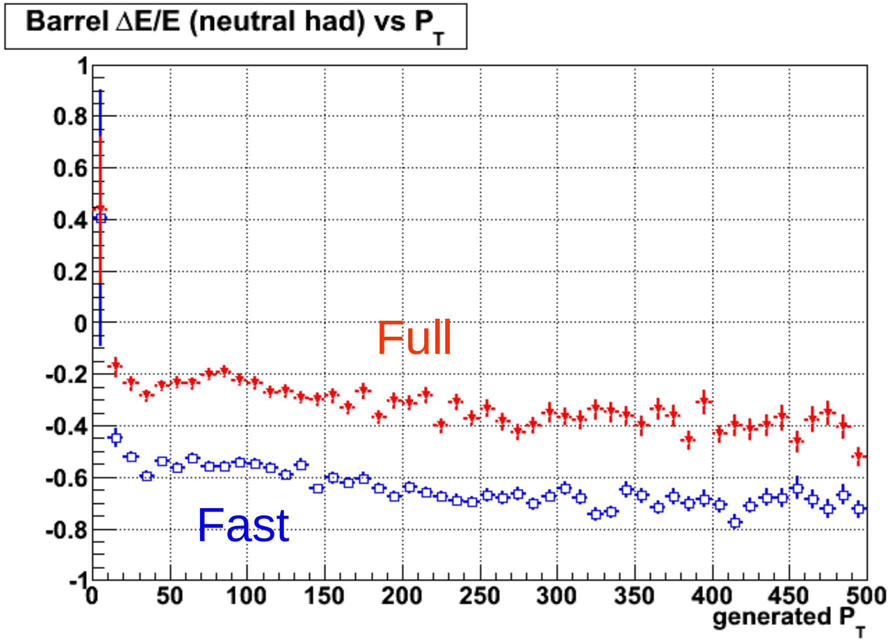
- Optimal use of tracking and ECAL for particle-id before jet clustering
- Used by default for jets, MET and isolation in high-P_T CMS analyses
- ~65% of jet energy is seen in the Tracker, only ~10% in HCAL
- A discrepancy in the neutral hadronic component was observed



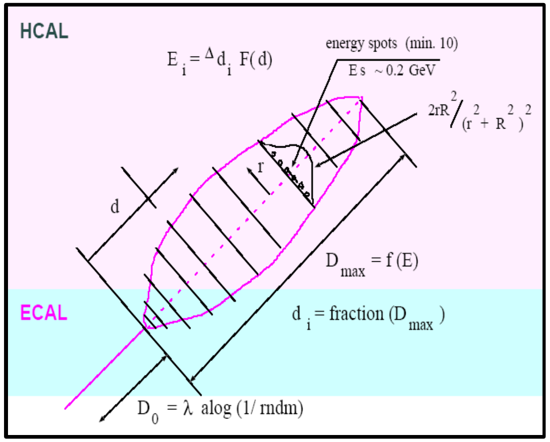
Private plots

In a nutshell

- It's not really a “missing” neutral component: it has to do with how PF accounts for far-away energy clusters, deciding whether to attribute them to a nearby charged particle or to create a “neutral had” cluster
- Lateral shower is tuned on FullSim, but cannot account for outliers
- As a short-term patch, we now can create extra clusters at PF level, fixing this variable while keeping (or improving) agreement elsewhere



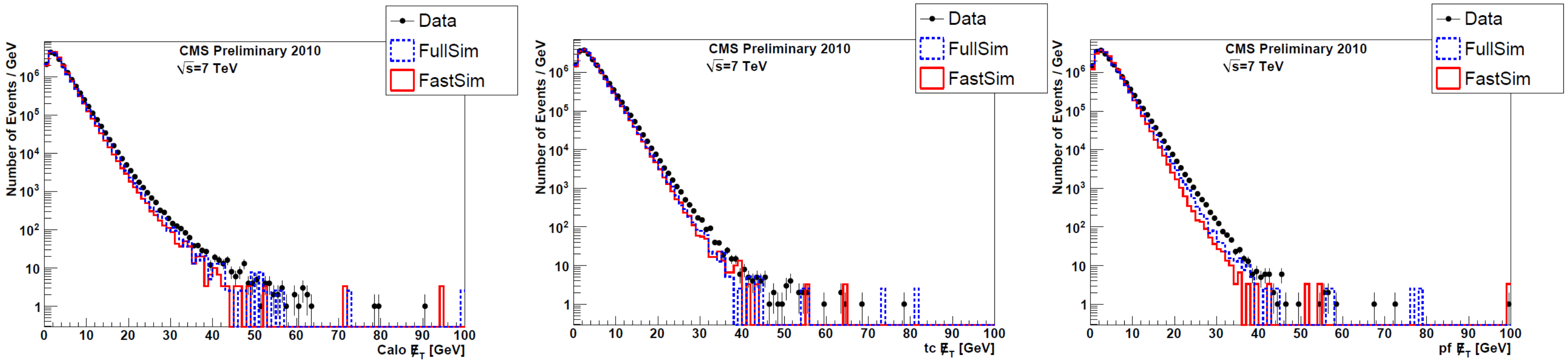
Current Hadron Shower model based on a 7 years old approximation of GFlash:



Complete GFlash implementation in FastSim could be a long-term solution

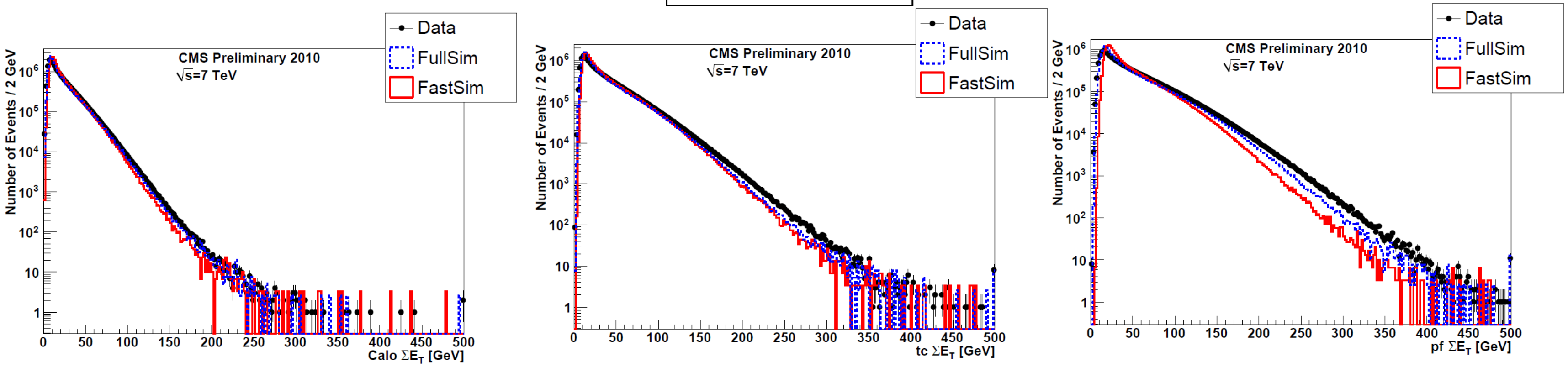
MET performance

MET



CMS-DPS-2010-039

Sum ET



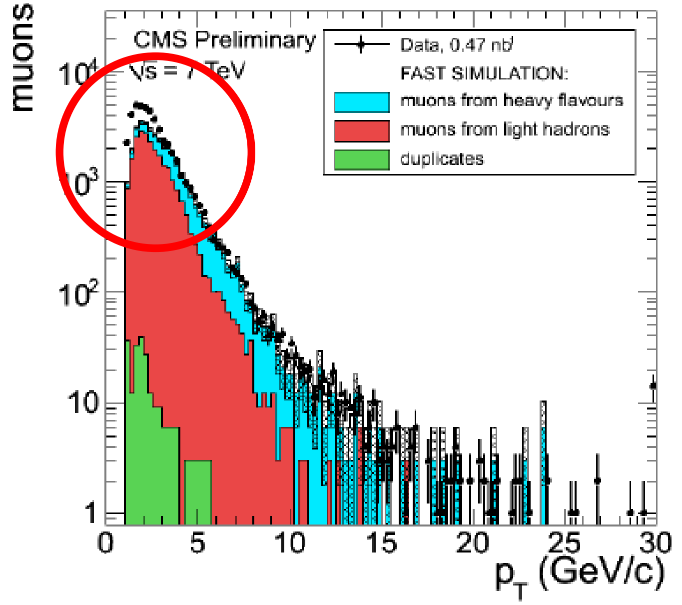
Calorimeters only Track-Corrected PFlow

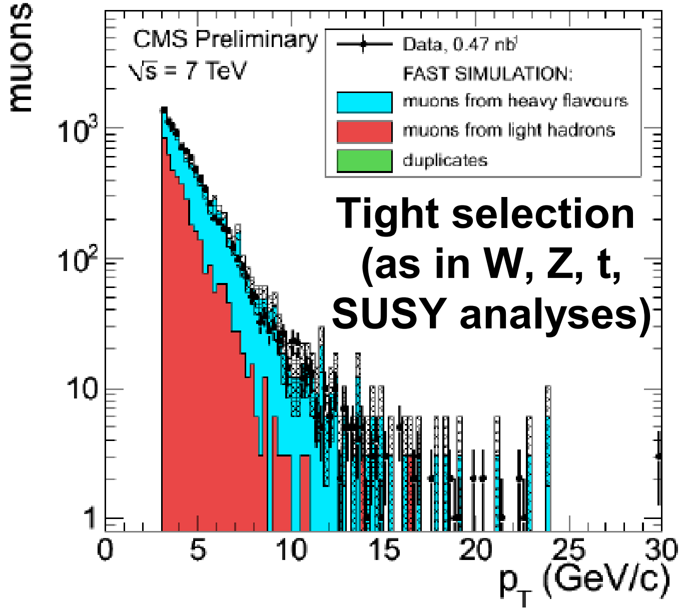
Impressive fast-full agreement over 7 orders of magnitude. Differences with data due to old Pythia tuning in early 2010 analysis; fast-full differences in PF are due to the differences in neutral hadron clusters

Muons in FastSim

- Muons are the only generated particles propagated to the muon chambers
  - Mult.scattering and dE/dx by ionization
  - Muons from hadronic decays propagated only if the decay is in the tracker volume; no late decays and no punch-through
  - Calo deposits are parametrized
  - No bremsstrahlung, no delta rays
- Same geometry as full sim
- Standard digi+reco is applied to the muon SimHits
  - No need for short-cuts in outer tracking: hit multiplicity is small

CMS-DPS-2010-039

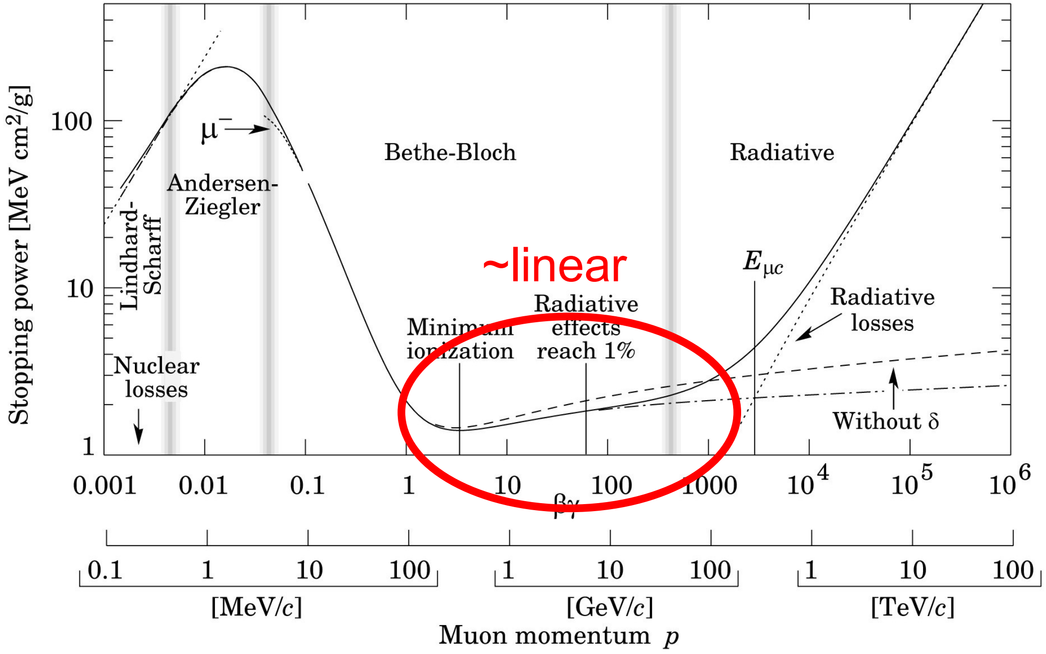




Tight selection (as in W, Z, t, SUSY analyses)

Emulation of δ-rays effects for μ

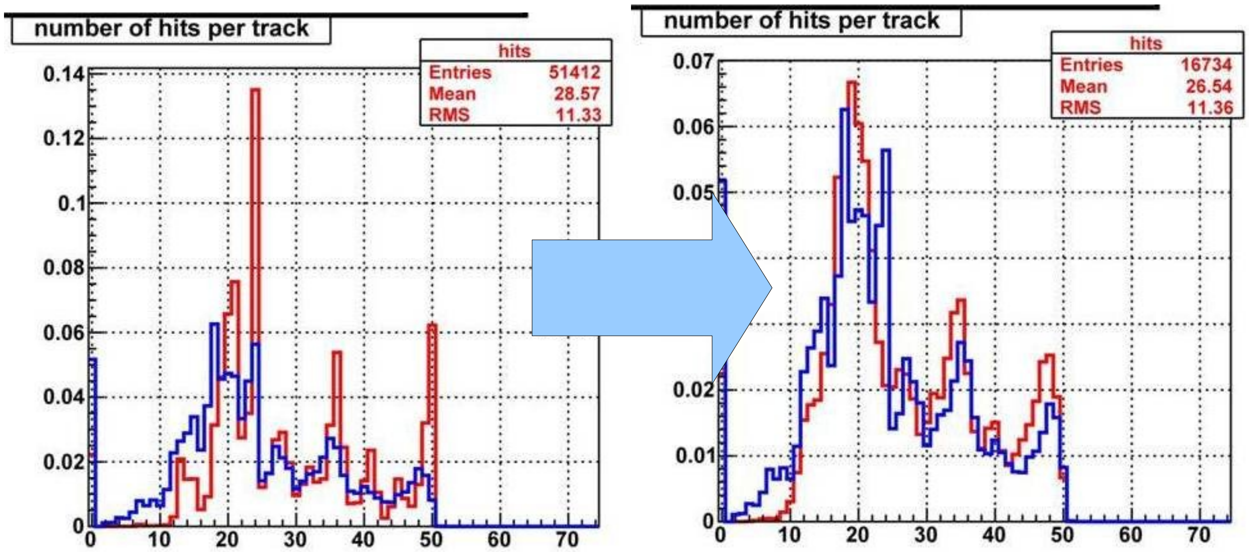
- δ-rays emitted at the entry of a cell may cause the hit to get corrupted (⇒inefficiency) or an after-pulse (~harmless)
- Log of hit inefficiency is found to be pretty linear with log(P) for DT and CSC, as expected if the cause are δ-rays; almost no P dependence found in RPCs, as expected due to their coarser spatial resolution
- Hit inefficiency has been parametrized as a function of log(P) for DT and CSC

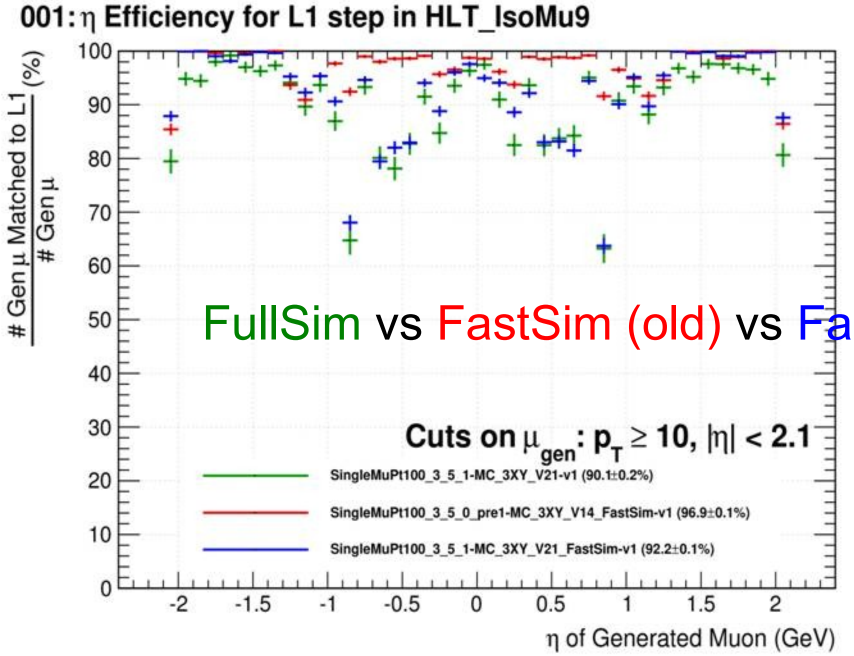


Before and after

Rec hit multiplicity in the muon chambers for L2 muon trigger, without and with the parameterization of the inefficiency due to delta rays

Taking the hit inefficiency into account yielded also a better description of reconstruction efficiencies, especially at trigger level



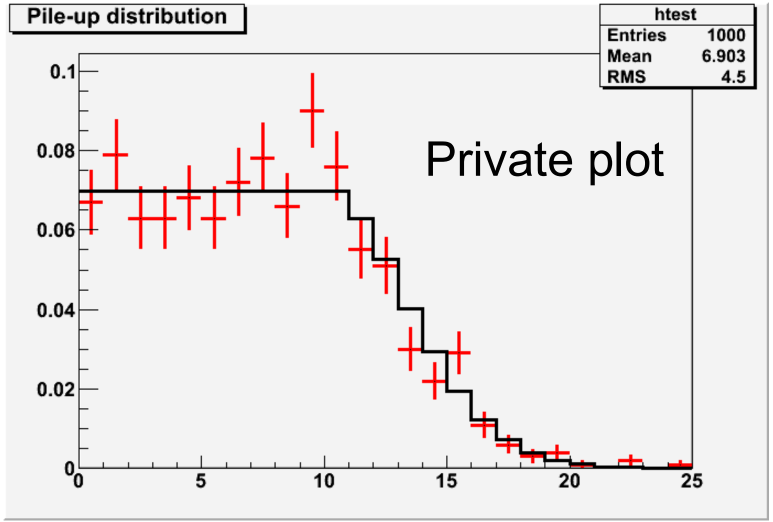


Private plots

FullSim vs FastSim (old) vs FastSim (new)

In-time pile-up

- Particles from additional minimum bias events (with vertex smearing) are added to the signal events
  - Difference with FullSim: they add SimHits, we add generated particles and we treat signal and PU particles at the same time
- Number of pile-up collisions can be diced from a Poissonian or from a user-defined distribution
  - Tools exist to reweight to a different distribution afterwards (see talk by Mike today)



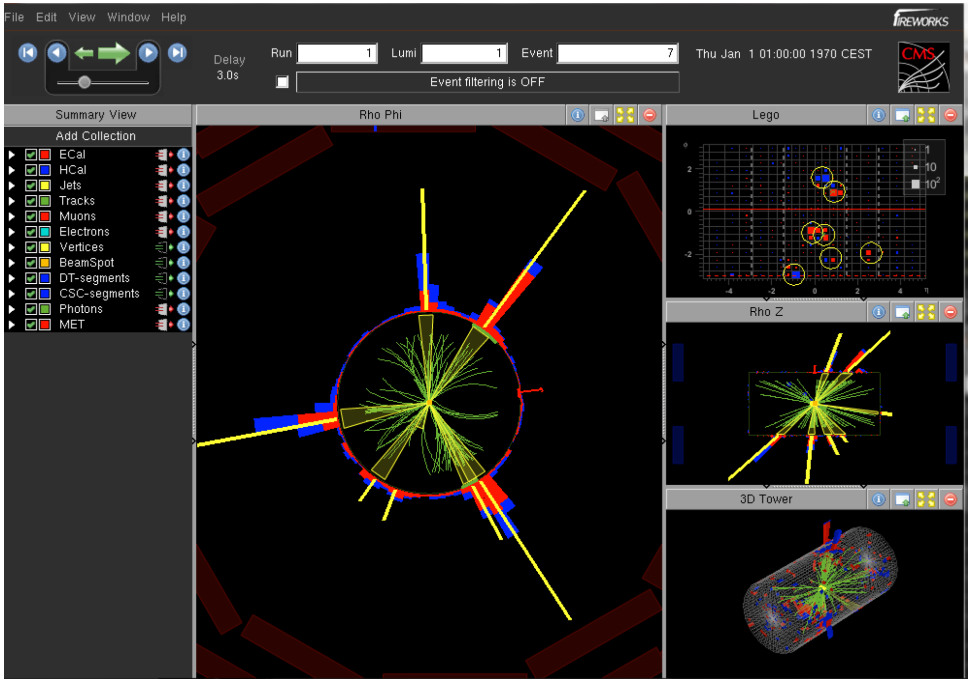
Out-of-time pile-up

- By design, there is no OOT pile-up
- Hard to ignore in 2012 with 25 ns running
- In CMS, negligible effect on inner (pixels, strips) and outer (DT, CSC, RPC) tracking systems because of the narrow pulse shapes
- Effect in the calorimeters can be treated as extra noise, and reabsorbed in noise tuning
  - Pro: extremely simple
  - Contra: different tunes needed to simulated different data-taking periods

Conclusions

- The CMS Fast Simulation is designed to achieve a O(%) accuracy with O(100) gain in speed with respect to GEANT
- Now being more and more used ($\Rightarrow$ validated $\Rightarrow$ improved) in CMS

Solution



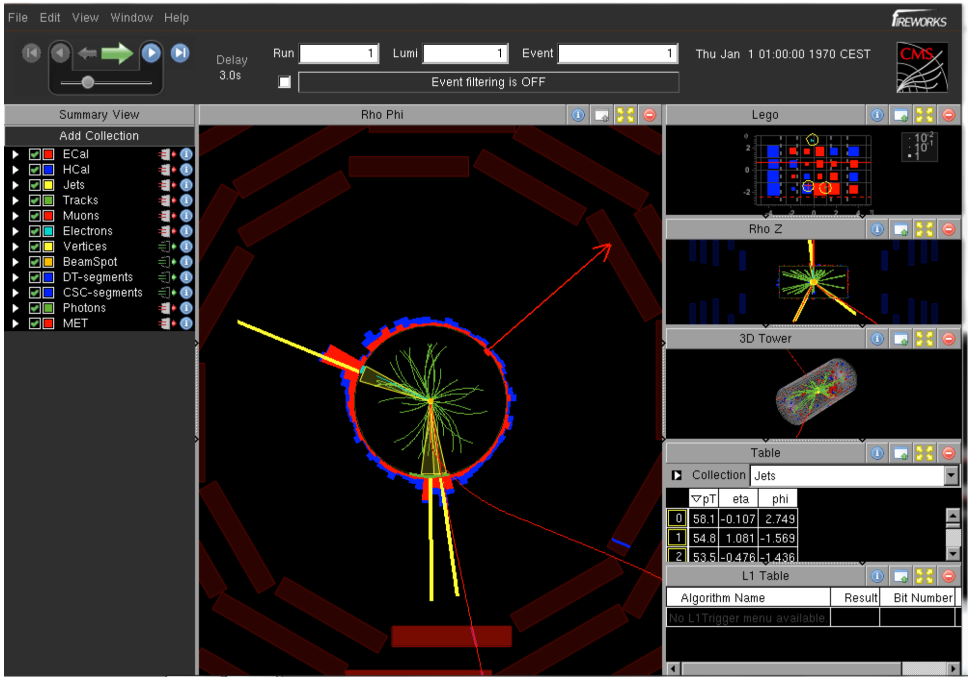
Run 1 Lumi 1 Event 7

Thu Jan 1 01:00:00 1970 CEST

Delay 3.0s

Event filtering is OFF

FullSim



Run 1 Lumi 1 Event 1

Thu Jan 1 01:00:00 1970 CEST

Delay 3.0s

Event filtering is OFF

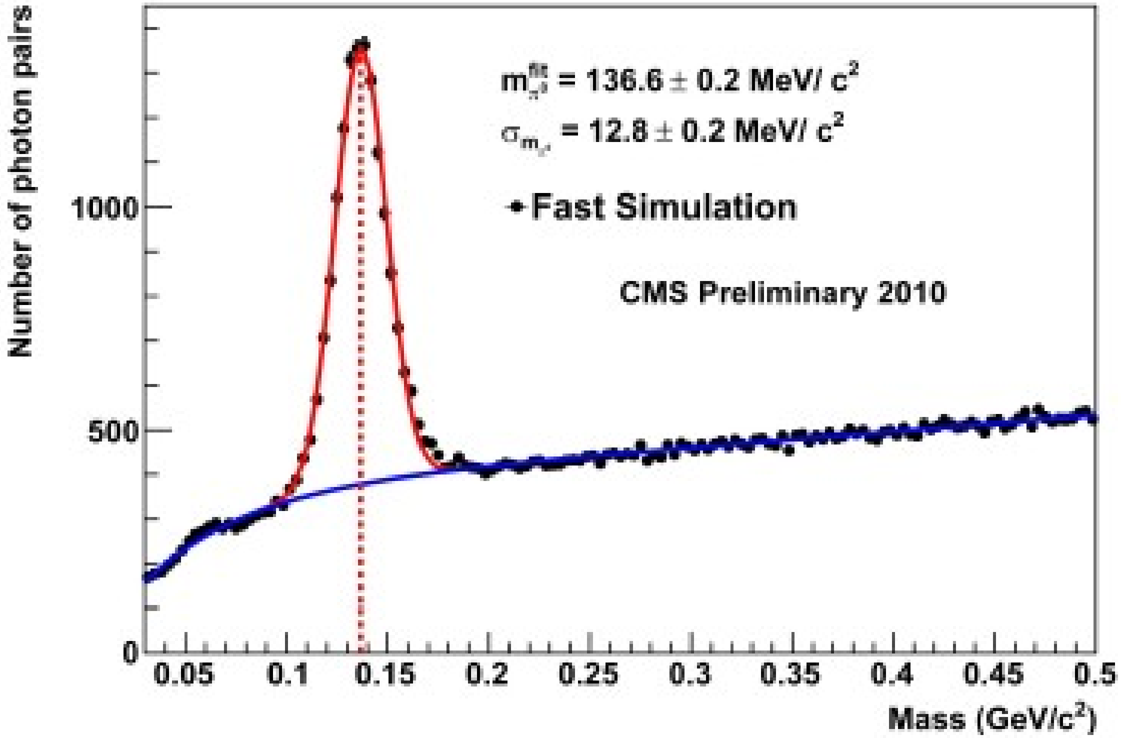
Collection: Jets

	▽pT	eta	phi
0	58.1	-0.107	2.749
1	54.8	1.081	-1.569
2	53.5	-0.476	-1.436

Algorithm Name	Result	Bit Number
No L1Trigger menu available.		

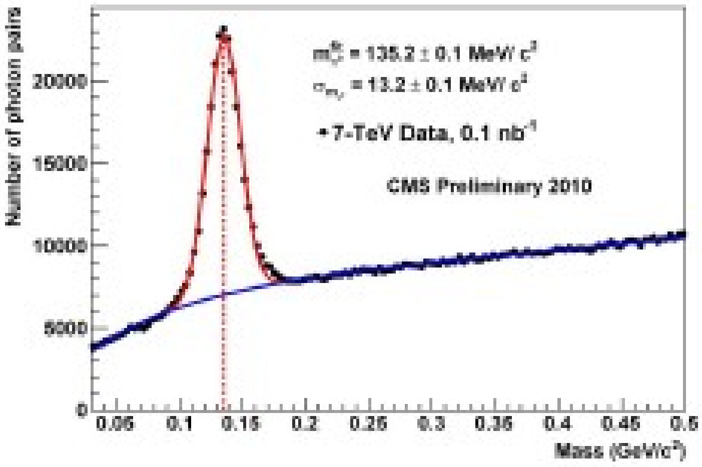
FastSim

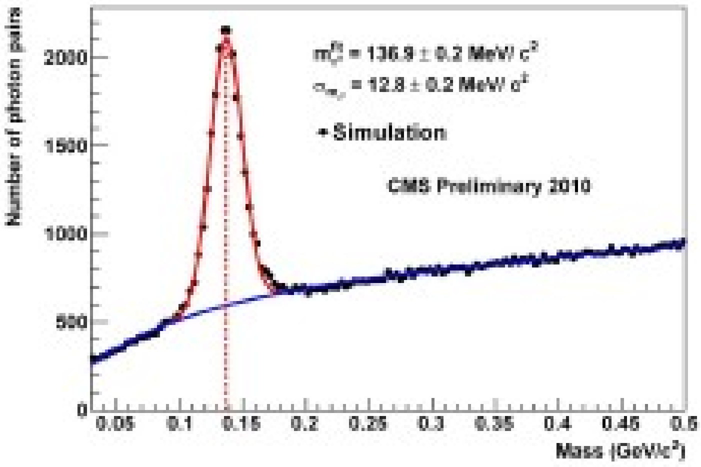
Di-photon invariant mass



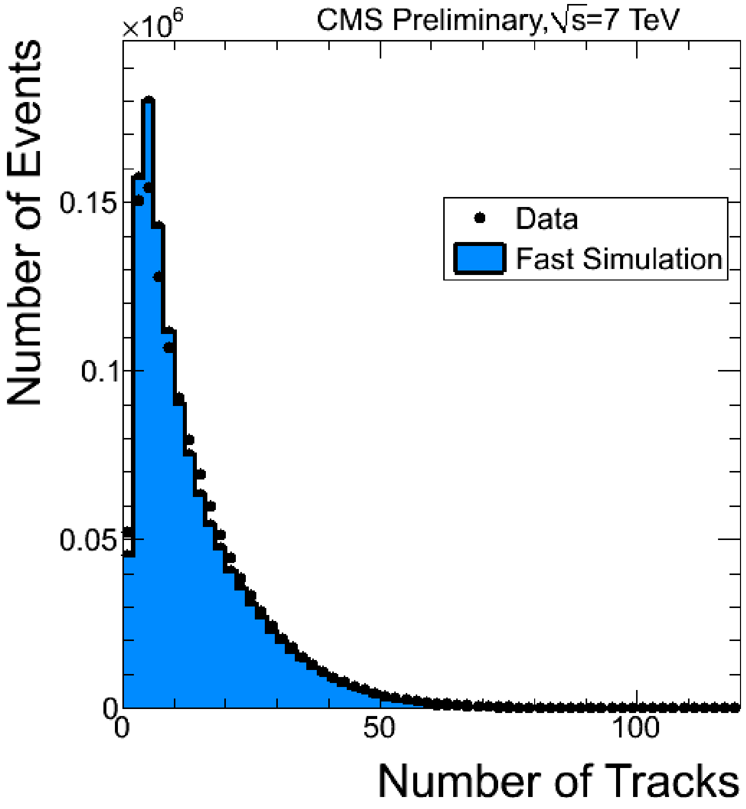
Di-photon invariant mass in the ECAL barrel (photons reconstructed with the Particle Flow algorithm)

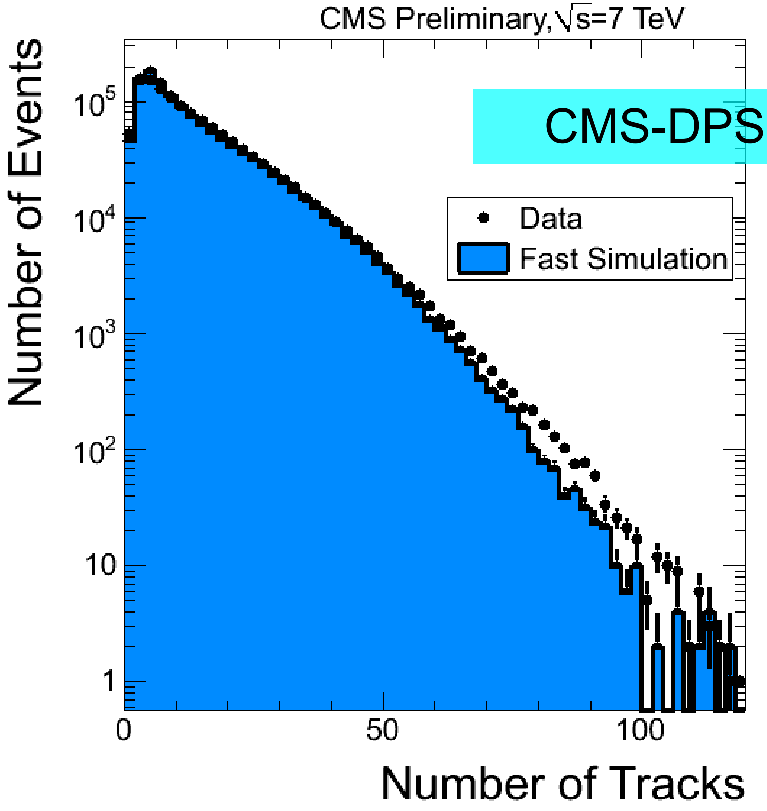
CMS-DPS-2010-039





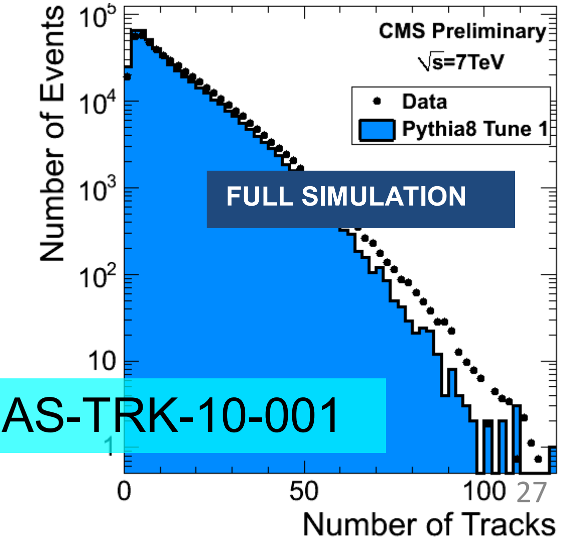
Number of reconstructed tracks





CMS-DPS-2010-039

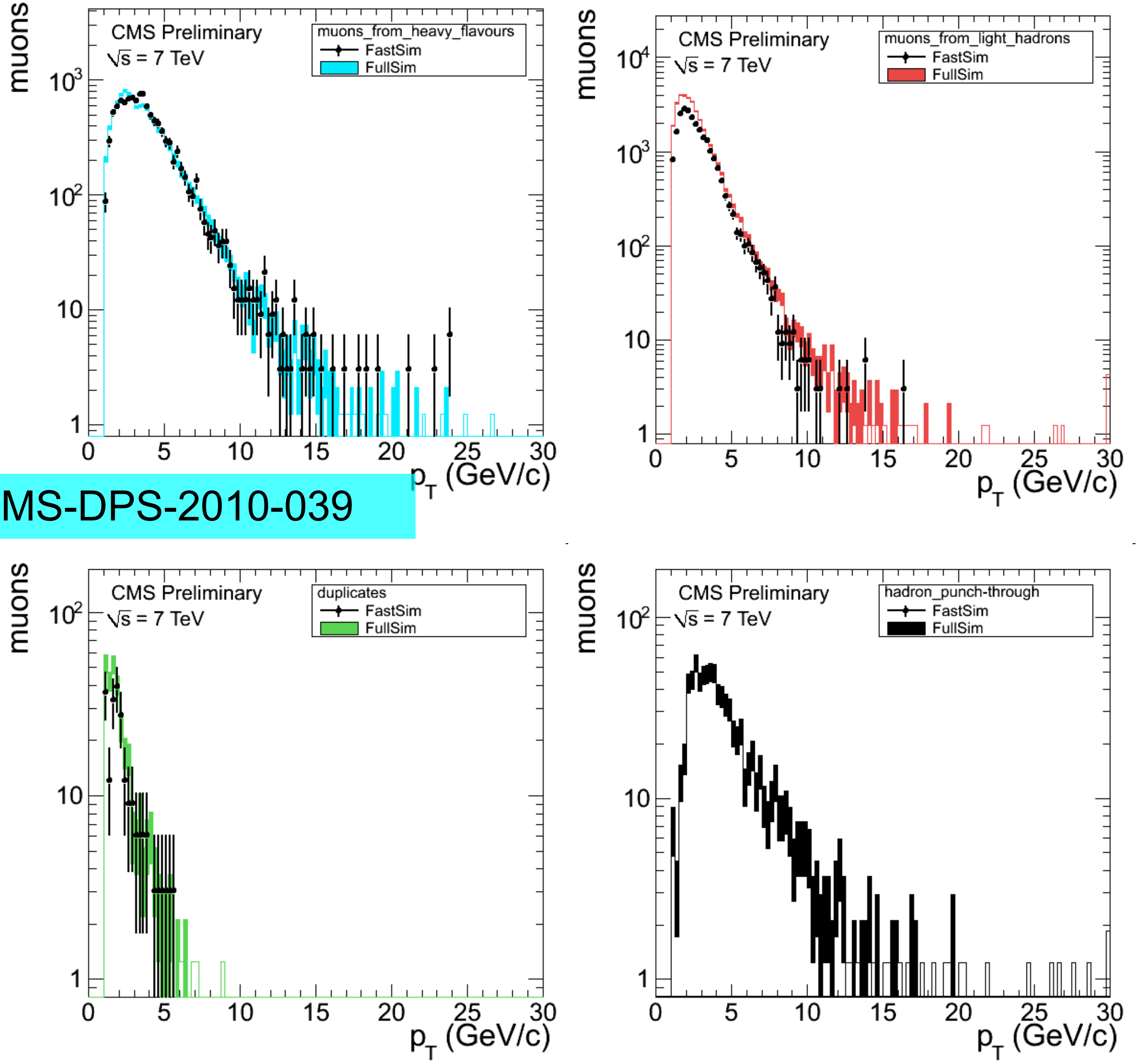
Lower number of events with high track multiplicity in the Fast Simulation with respect to the data. The same also happens in Full Simulation: Pythia tuning.



FULL SIMULATION

CMS-PAS-TRK-10-001

Breakdown of the p_T distribution according to the origin of the muon



CMS-DPS-2010-039

p_T distribution (Global Muons)

Full (histo) vs Fast (points) contributions, separated per muon type:

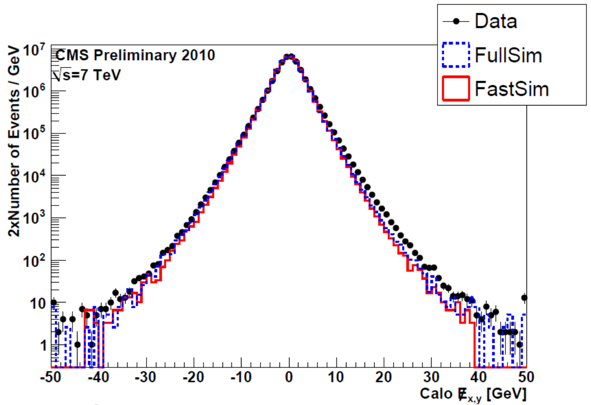
- **Prompt muons**

- **Decays in flight**

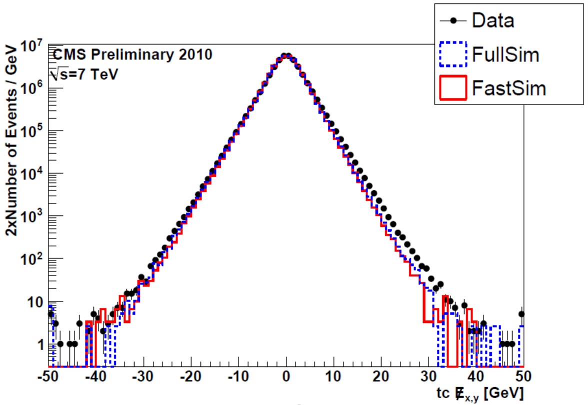
- **Fakes (ghosts)**

- **Punch-through** (only for the Full Simulation, as the Fast Simulation does not simulate it)

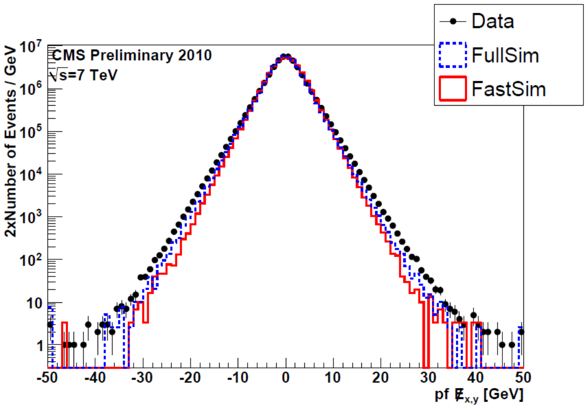
MET resolutions



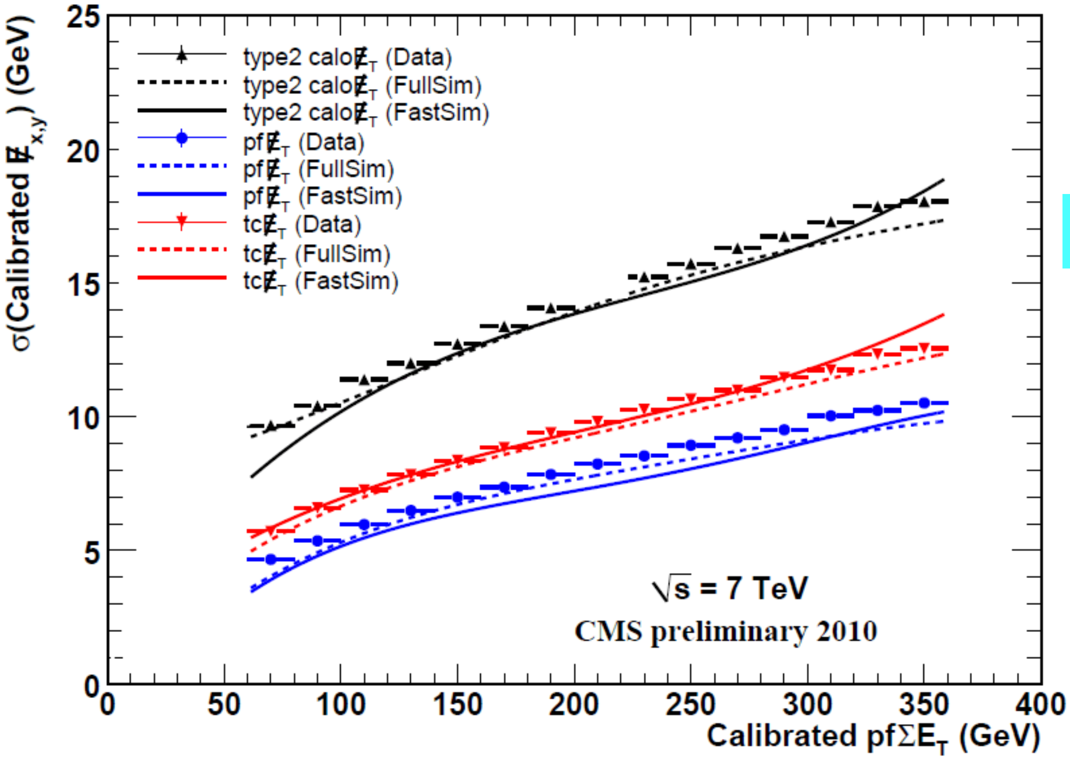
Calorimeters only



Track-Corrected

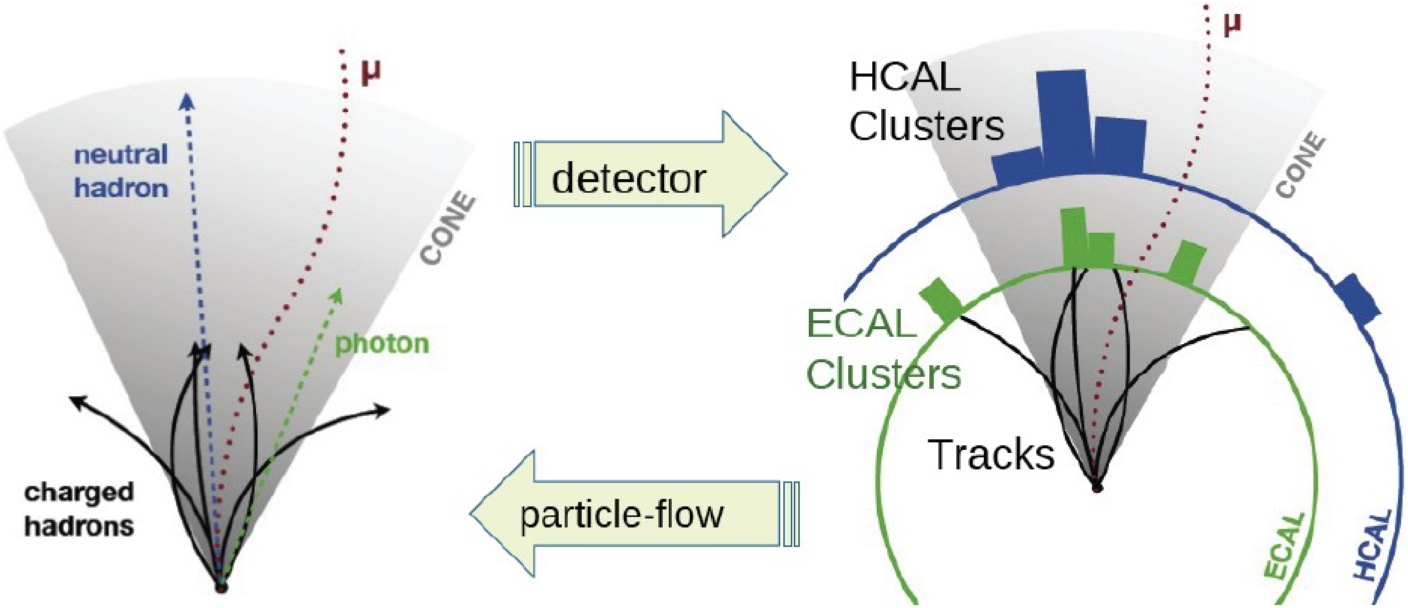


PFlow



CMS-DPS-2010-039

Particle Flow



- Idea: first perform **particle-ID** (μ,e,γ,$h^{\pm}$,h^0) and calibrate each candidate according to its identity, then cluster identified particles into jets
  - Compare with calorimetric approach: first cluster all calo deposits, then correct
- This makes optimal use of tracking and EM calorimetry: CMS strong points!
- A typical pt~50 GeV jet has **~65%** of its energy from **charged particles** (including $V^0 \rightarrow h^+h^-$), and **~25% from γ** (including $\pi^0 \rightarrow \gamma\gamma$)
  - HCAL resolution only affects the ~10% by neutral long-lived hadrons
- Extra bonus: one can ignore charged particles coming from pile-up

Particle Flow