CMS FastSim: Physics Object Validation

Lukas Vanelderen (Hamburg U.)
Andrea Giammanco (UCL Louvain-la-Neuve & FNRS)
On behalf of the CMS collaboration

Outline

- Intro: what we validate and how
- Tracks
- Electrons
- Photons
- Muons
- Jets
- MET
- b-tagging
- tau-tagging

Validation workflow

- For each CMSSW (pre)release, a set of release validation samples (*relvals*) is produced, for:
  - data
    re-reco and/or re-calibration
  - FullSim
    usually recycling generation and simulation, unless changes are expected/possible
  - FastSim
    produced from scratch and with x10 more statistics than FullSim

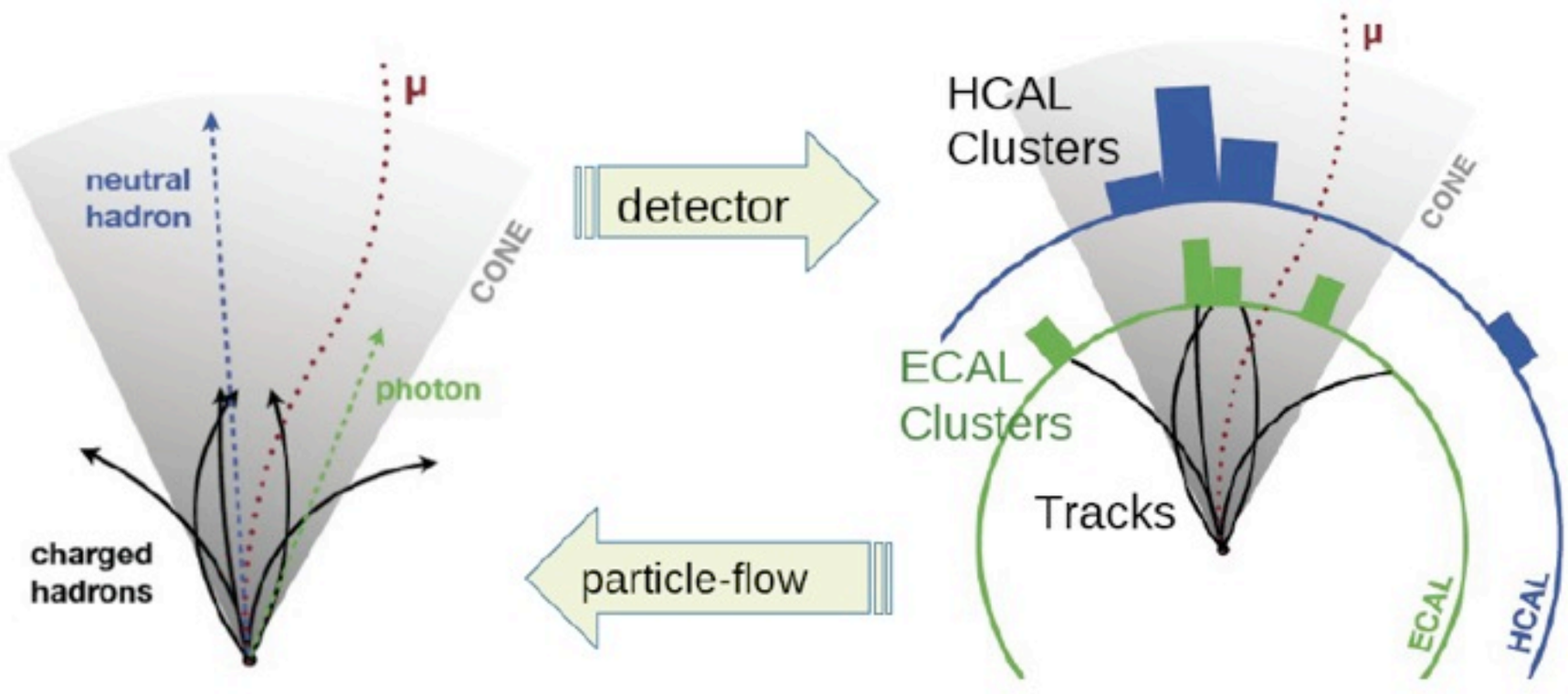
Validation Workflow

- Data Quality Monitoring histograms are produced
- A group of validators looks at those histograms They specialize per detector, per physics object, or per physics analysis group
- Comparisons run wrt a previous reference release; in addition, FastSim-vs-FullSim comparisons are performed within the same release

FastSim and Reconstruction/ Calibration

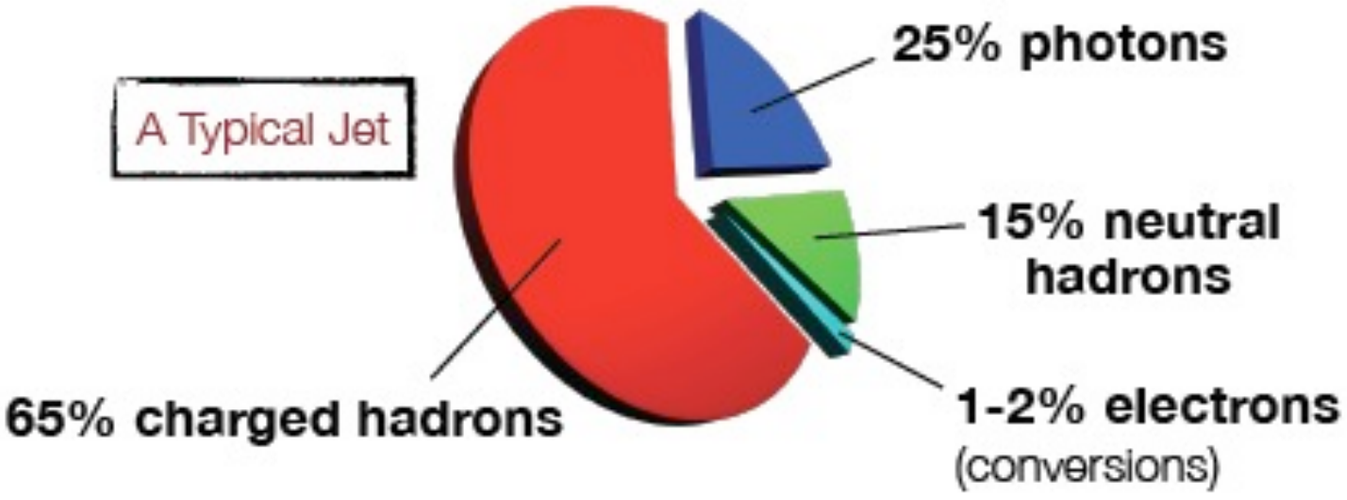
- The general aim of the CMS FastSim is to deliver low-level objects to be fed as input to high-level modules that are in common with FullSim (e.g., the digitizers) and in many cases also with data (e.g., reconstruction (*), calibration)
  - (*) most notable exception is tracking, which is emulated
- One implication is that very often, when the routine validation procedures spot an unexpected discrepancy in MC with respect to previous releases, we understand if it is related to simulation itself or to changes in reconstruction/calibration, because in the latter case both MC distributions move coherently
- Sometimes an unexpected change in reconstruction/calibration is unnoticed or unclear in the FullSim validation because of poor statistics of the validation samples; FastSim samples have 10x larger statistics and help spotting this kind of issues

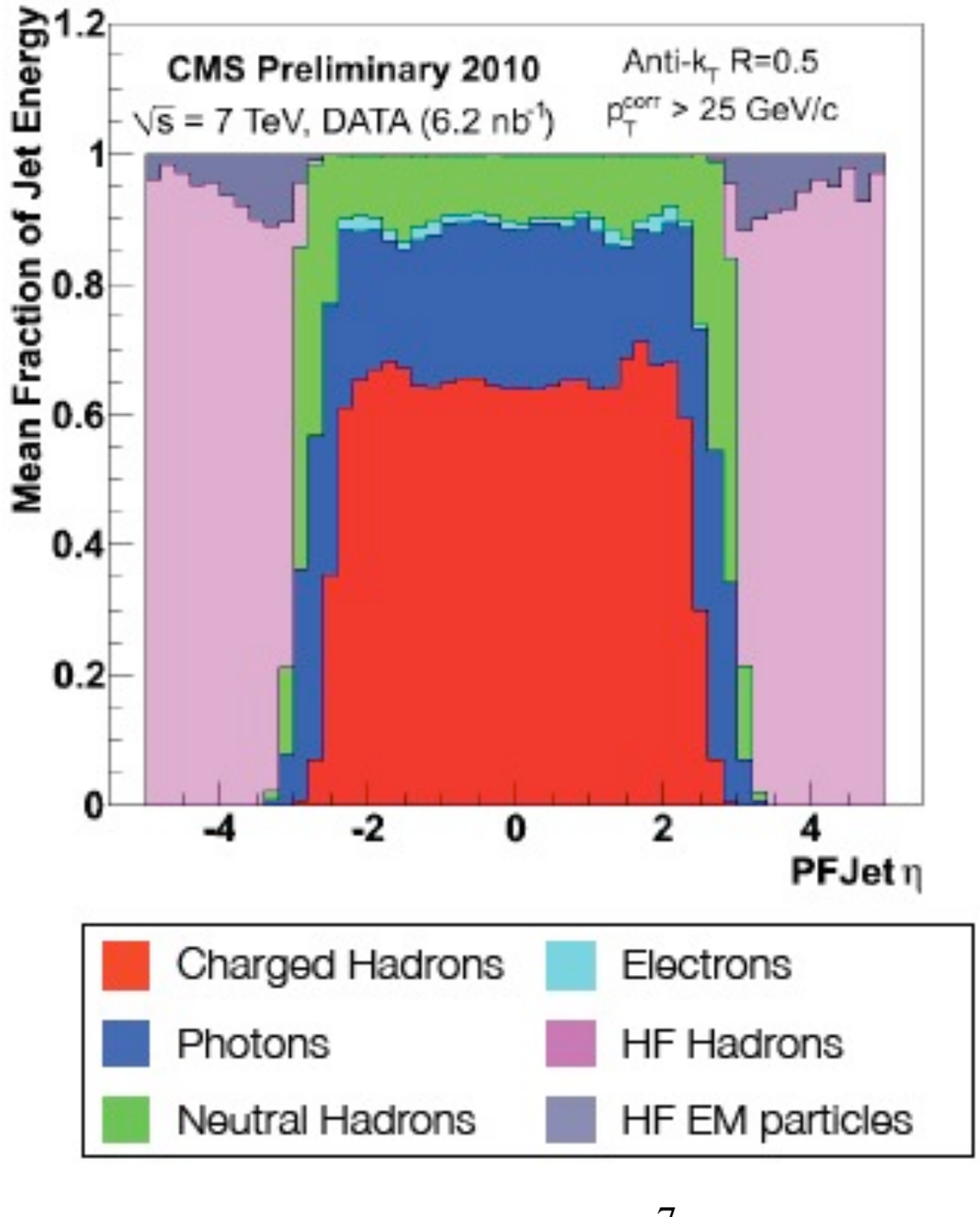
Particle Flow



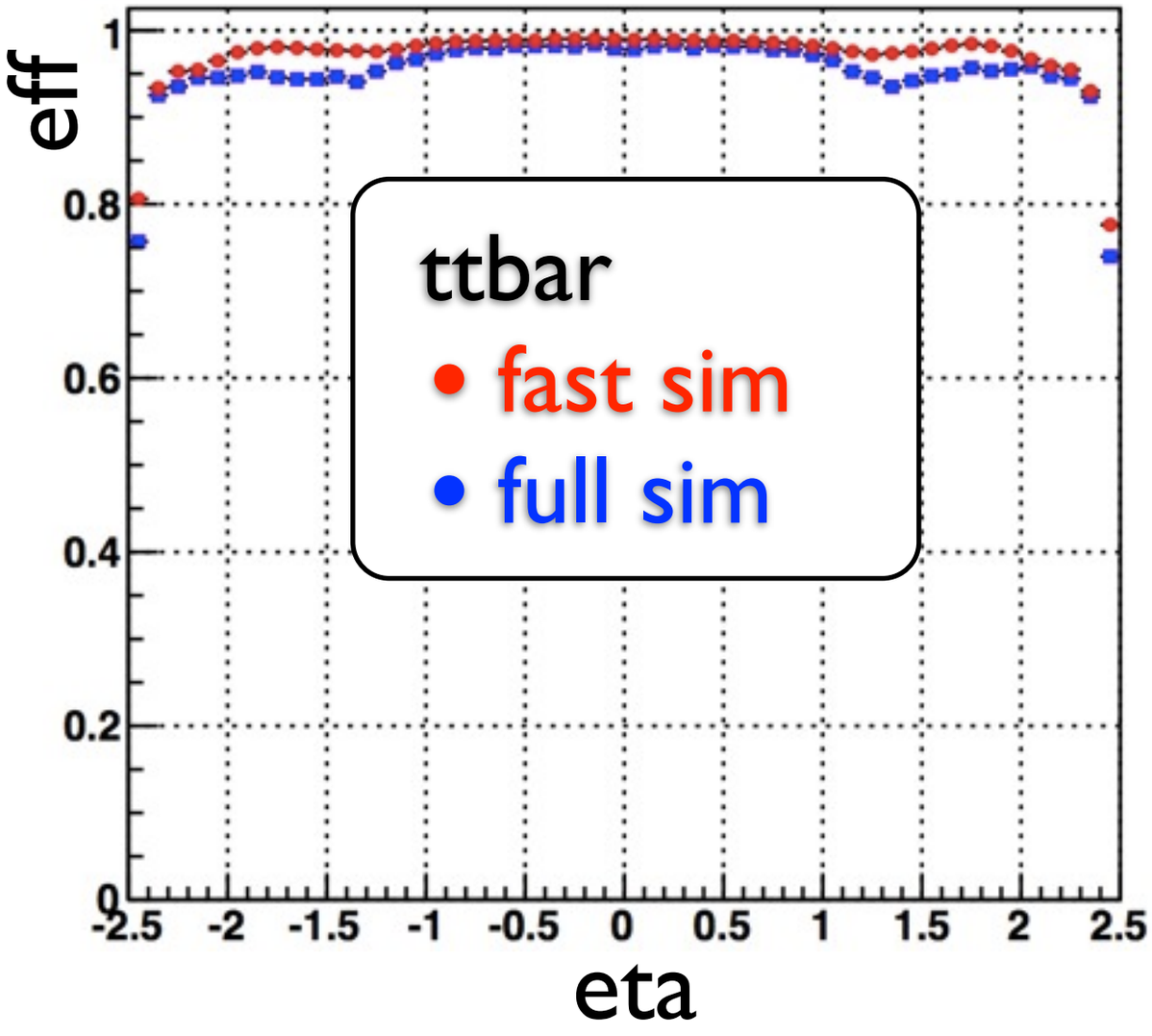
- Idea: first perform **particle-ID** (μ,e,γ,$h^{\pm}$,h^{0}, sequentially and mutually exclusive) and calibrate each candidate according to its identity, then build jets and tau-jets, calculate isolation, MET, etc., ignoring charged particles associated to pile-up
  - Compare with calorimetric approach: first cluster all calo deposits, then correct
- Because of the PF approach, the accuracy of the simulation of tracks (the only physics objects whose reconstruction takes shortcuts in FastSim) has an indirect impact on the accuracy of every object downstream!

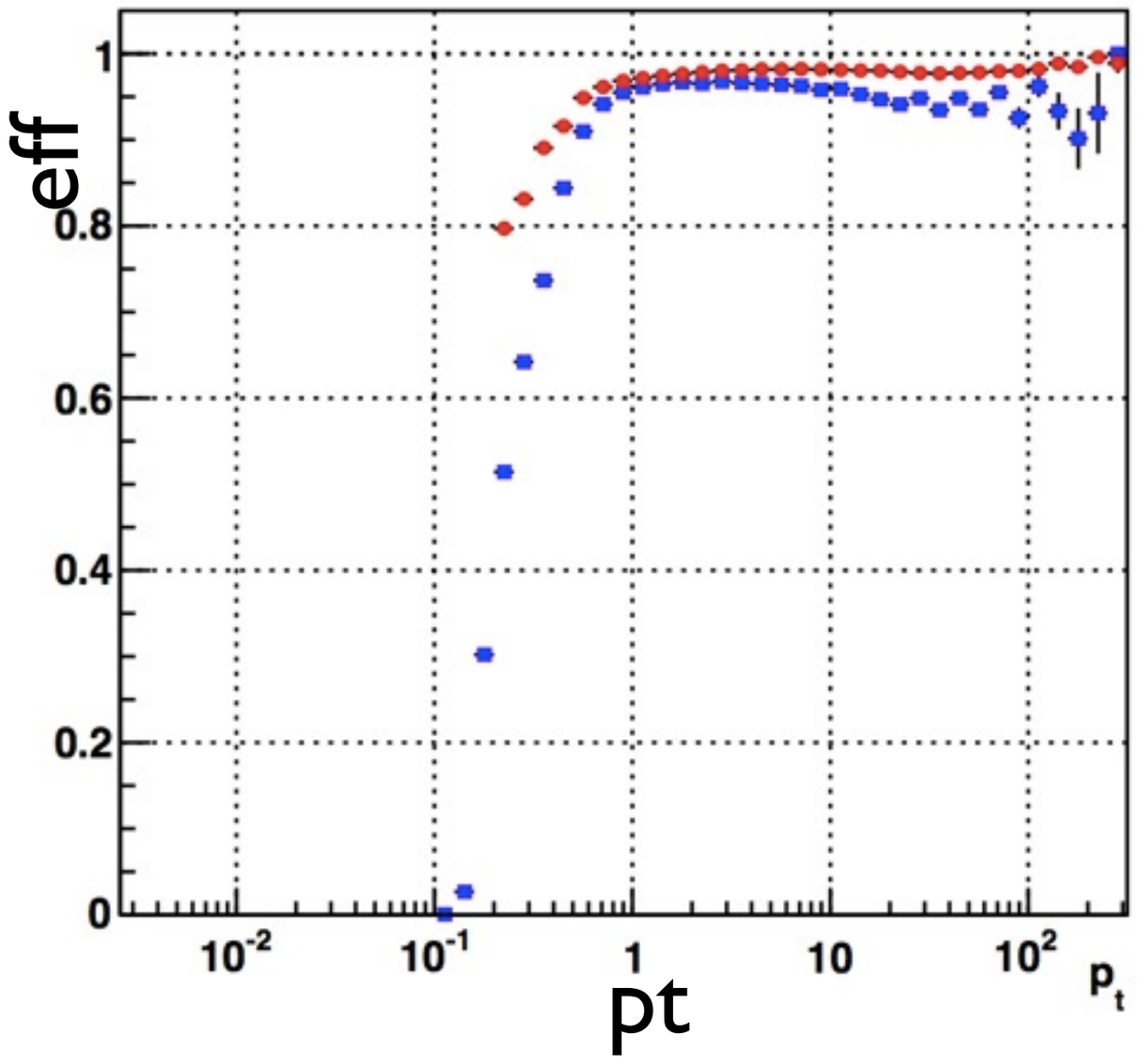
Particle Flow





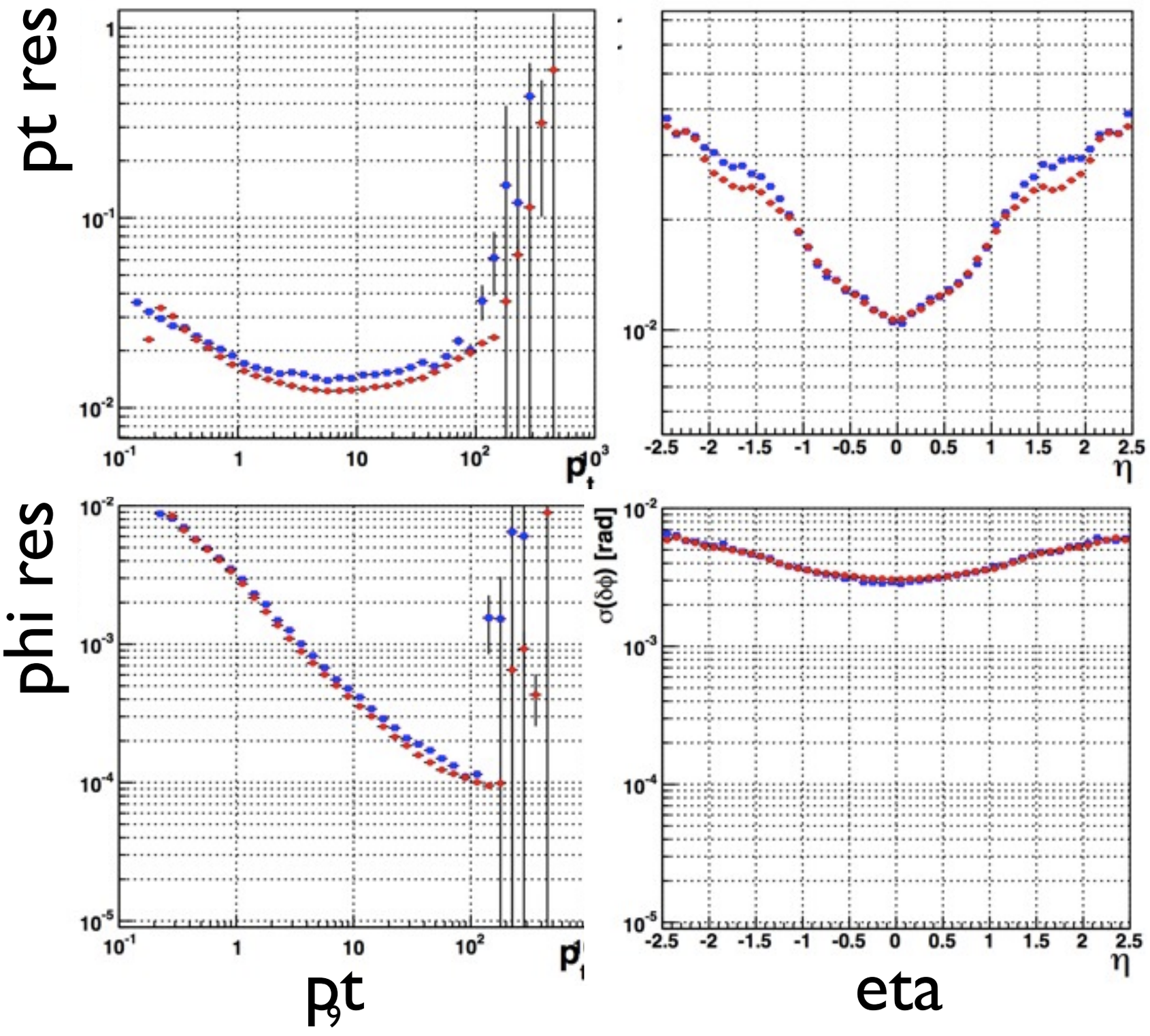
tracks: efficiency





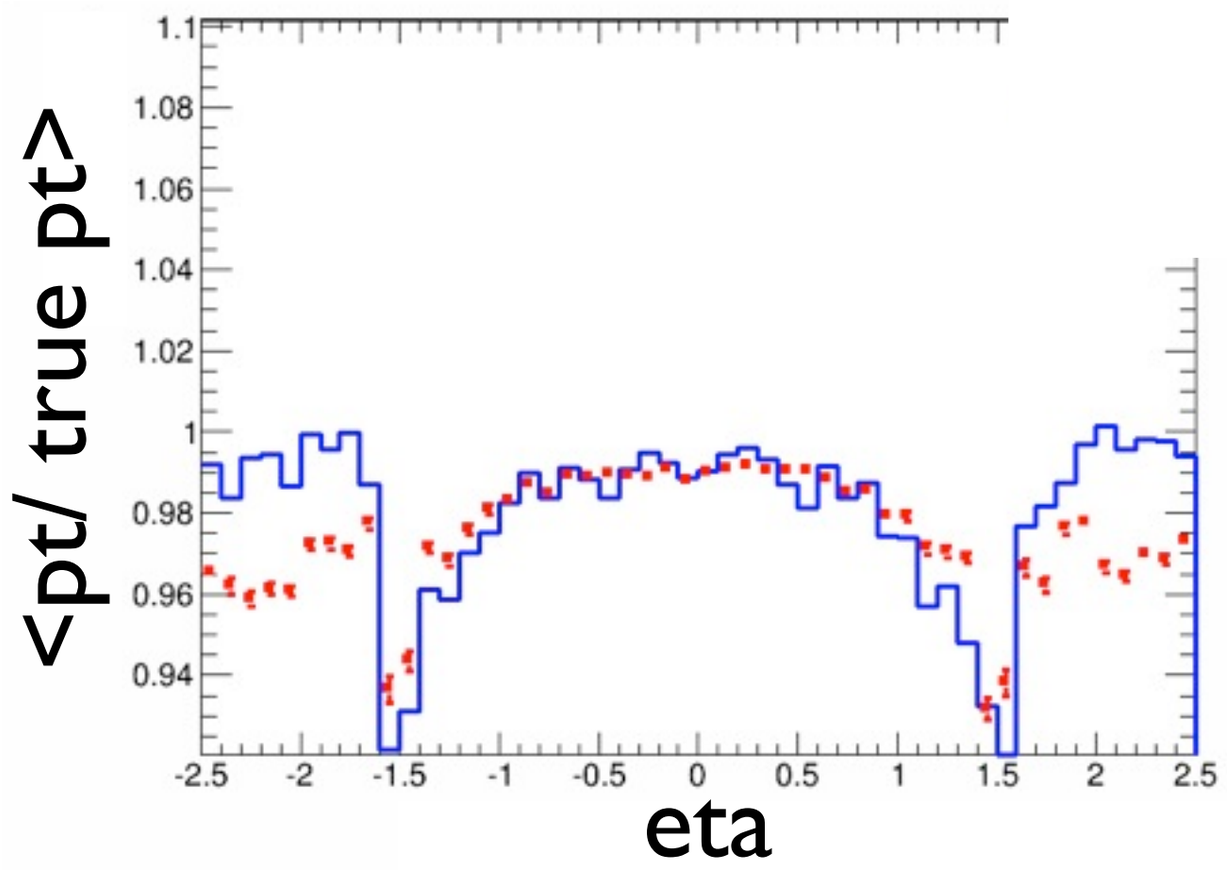
tracks: resolution

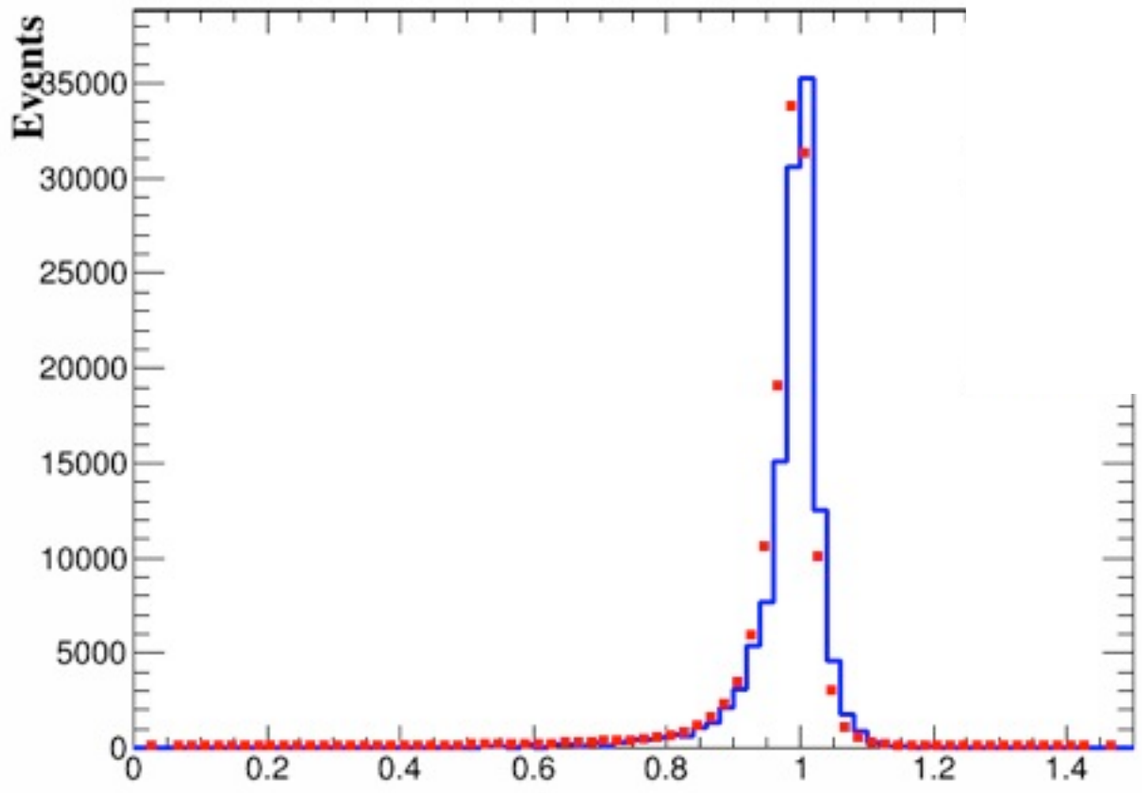
ttbar
- fast sim
- full sim



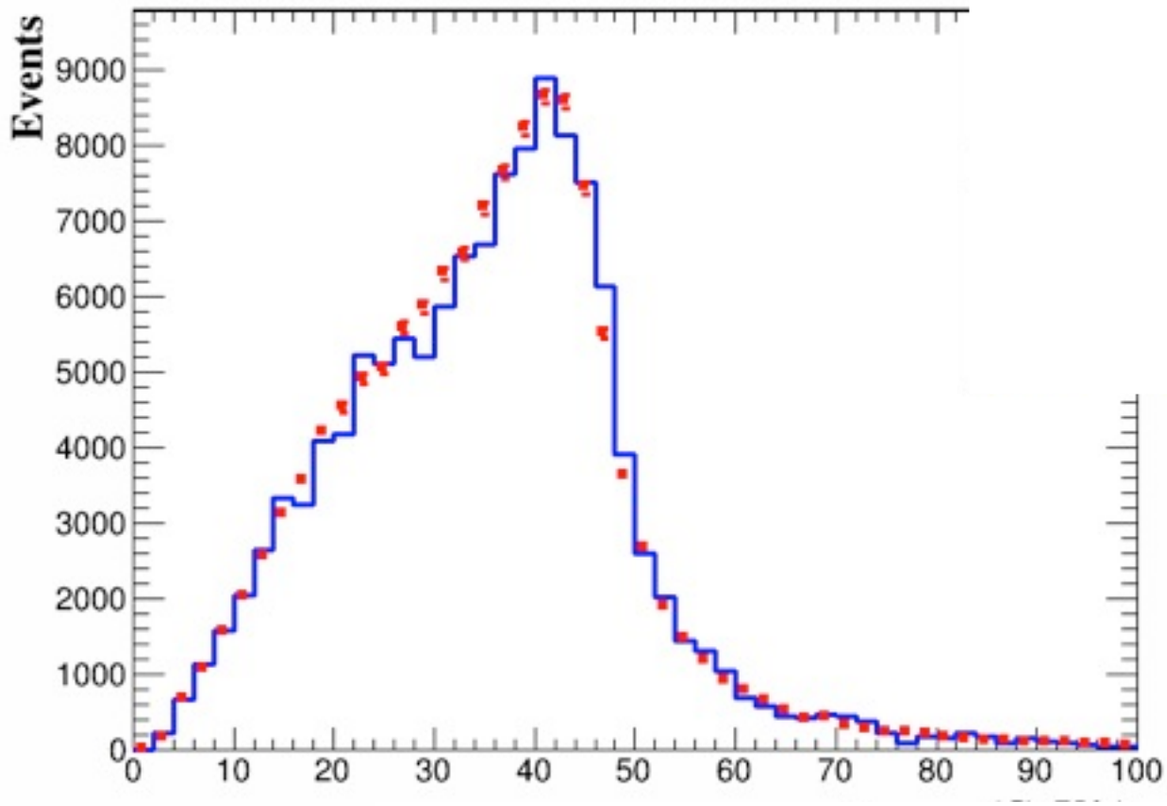
electrons: pt

Z(ee)
- fast sim
- full sim



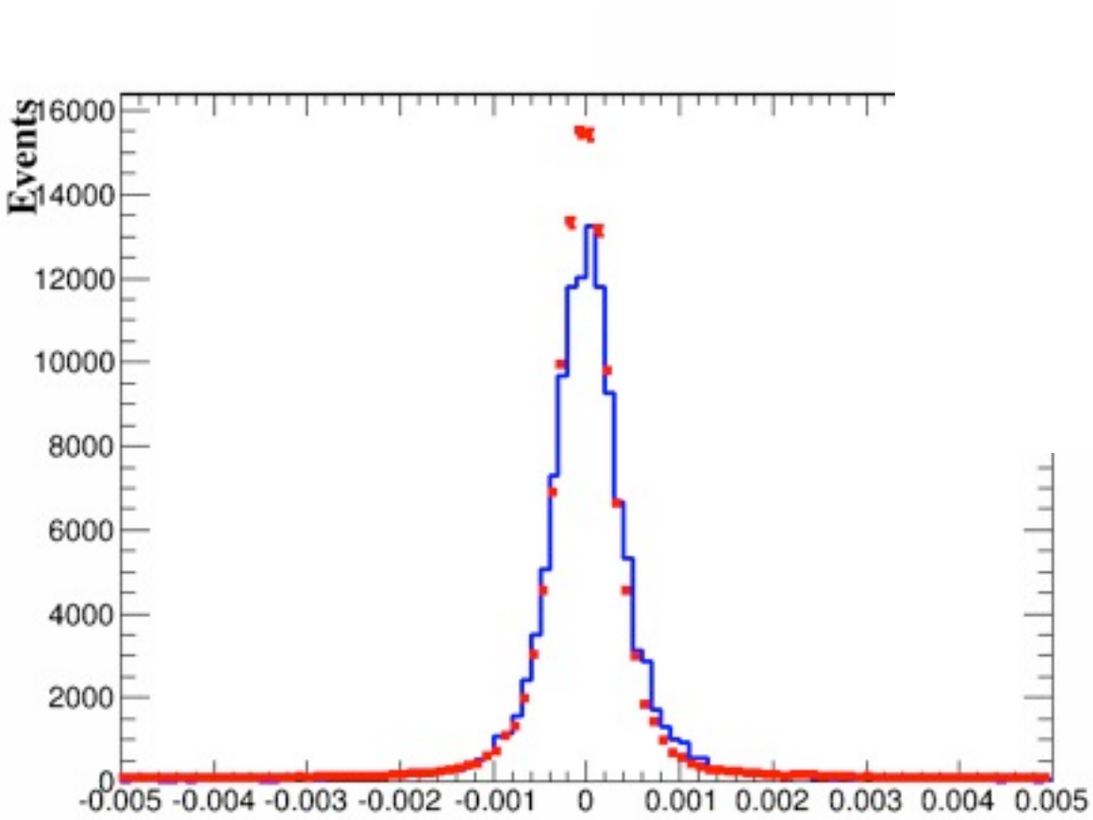


pt / true pt



pt

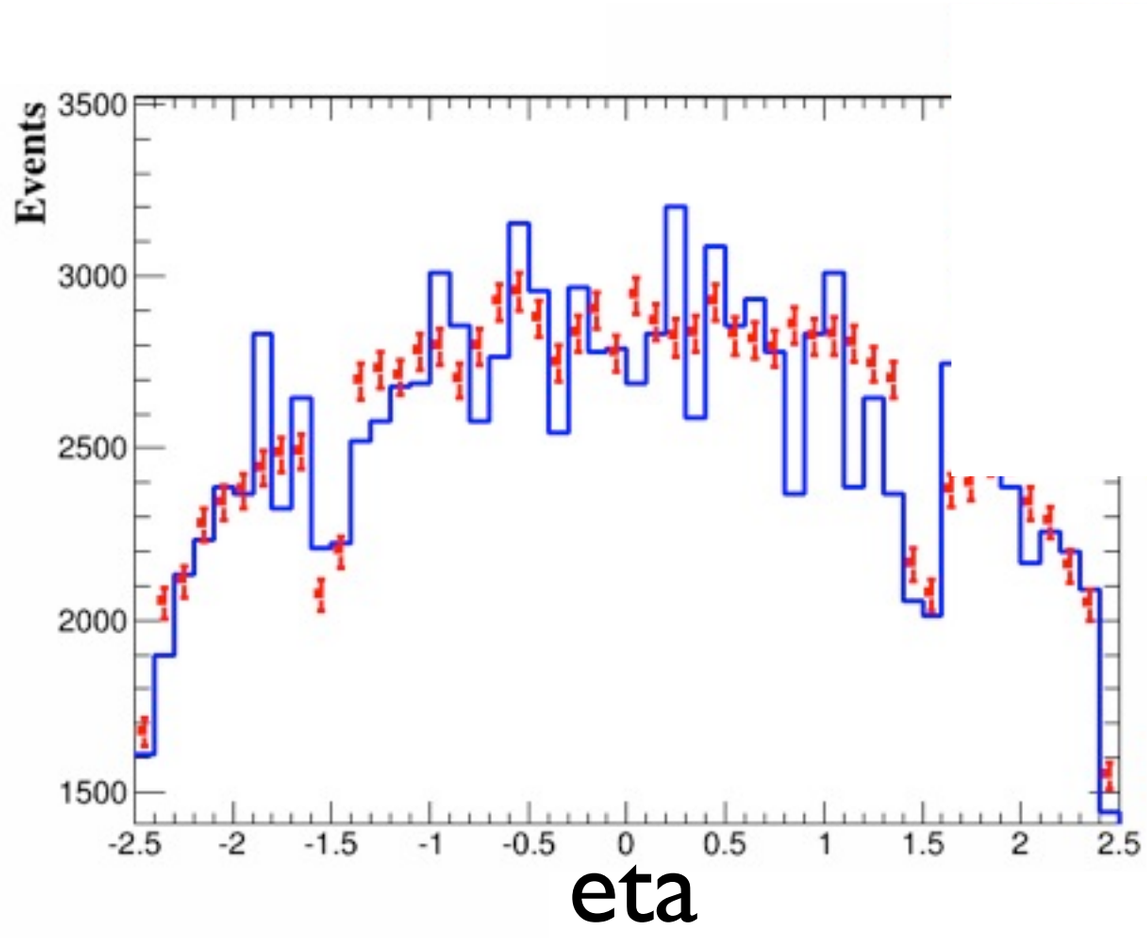
electrons: eta



eta - true eta

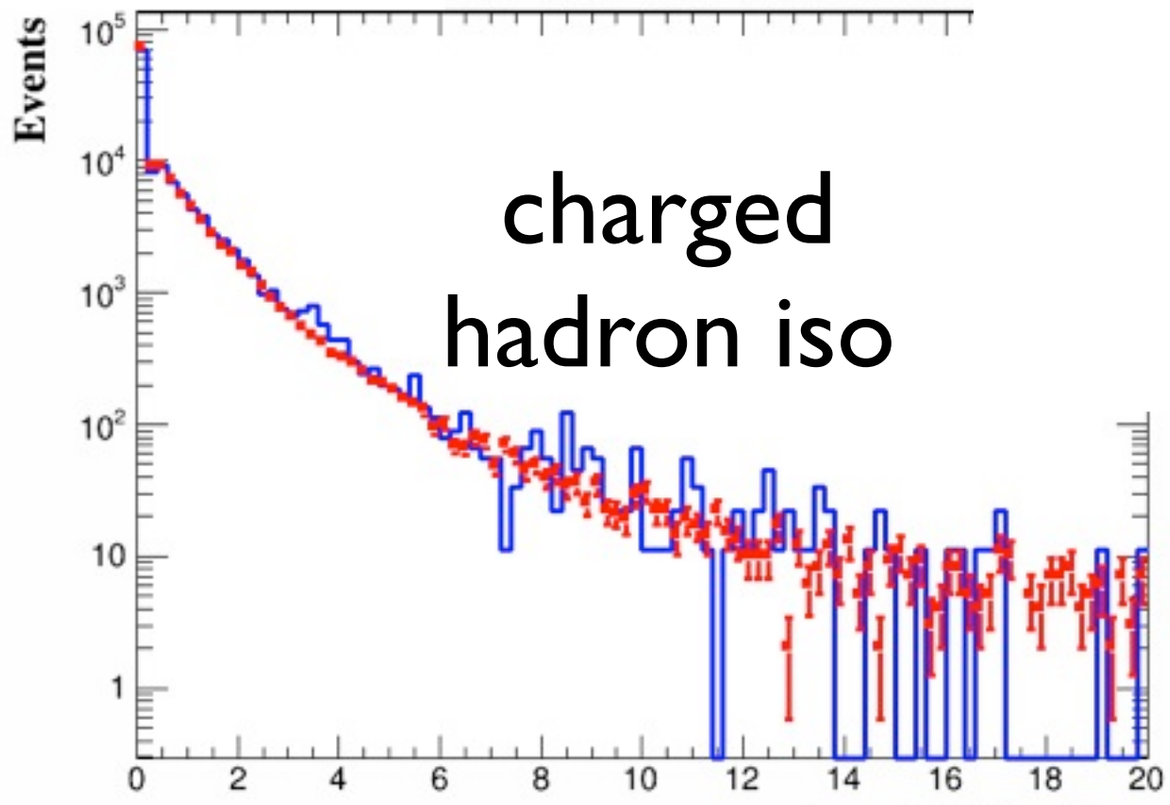
Z(ee)

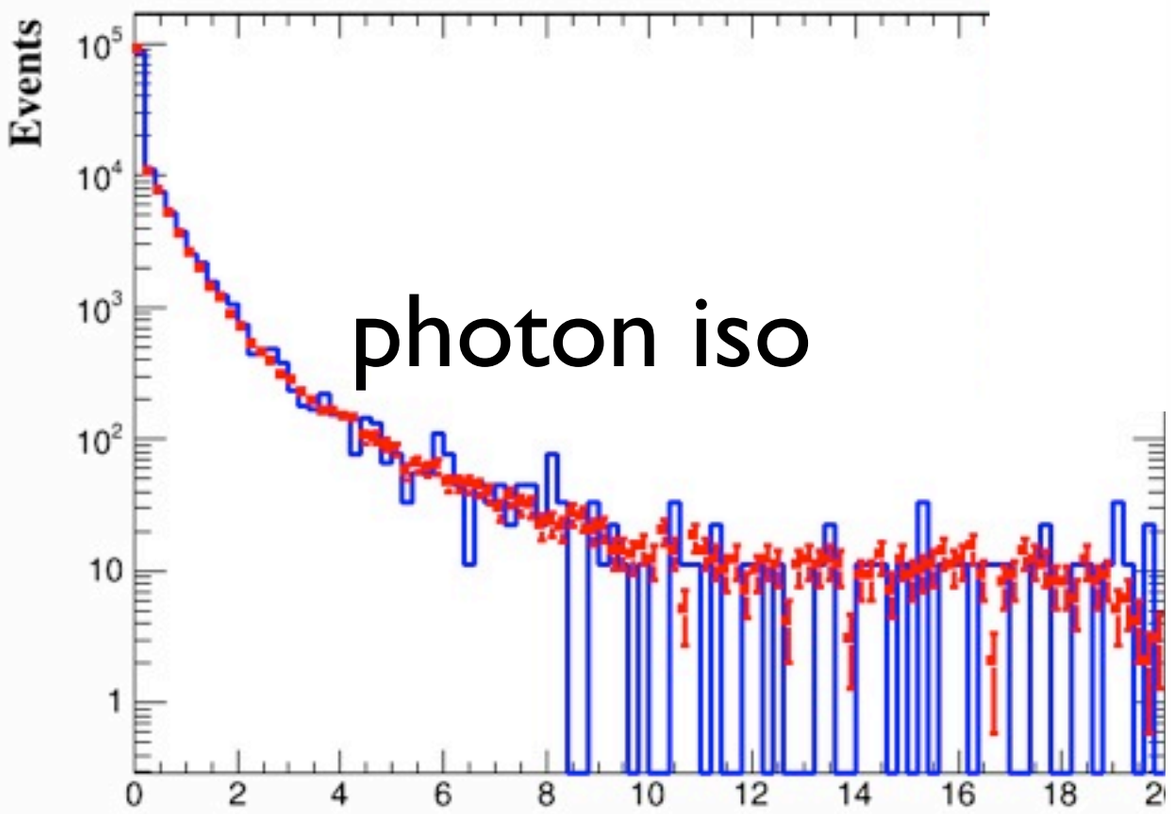
- fast sim
- full sim

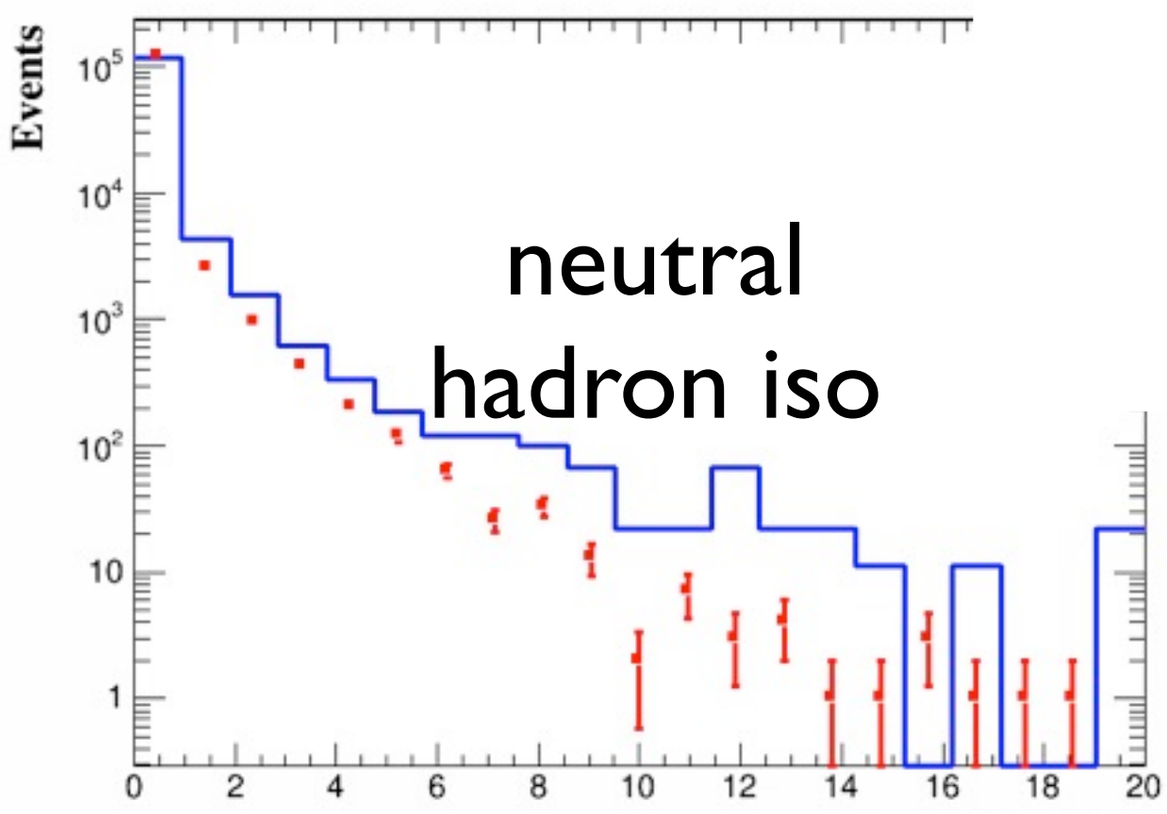


electrons: isolation

Z(ee)
- fast sim
- full sim

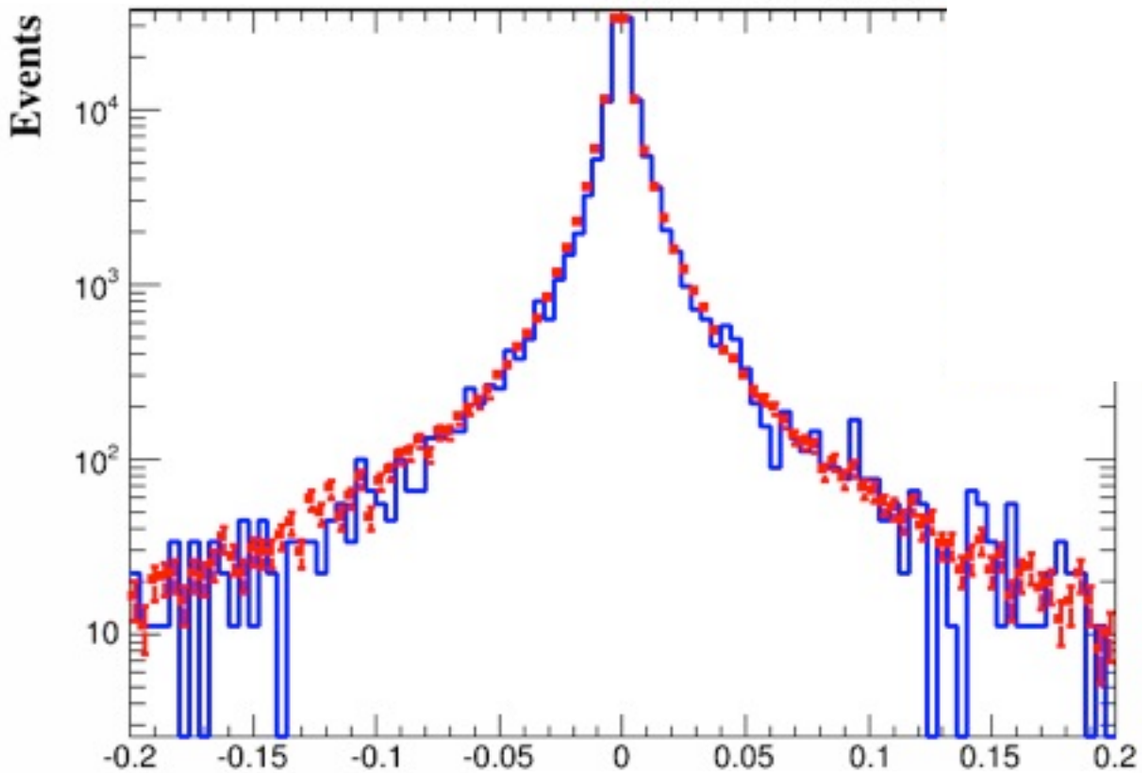




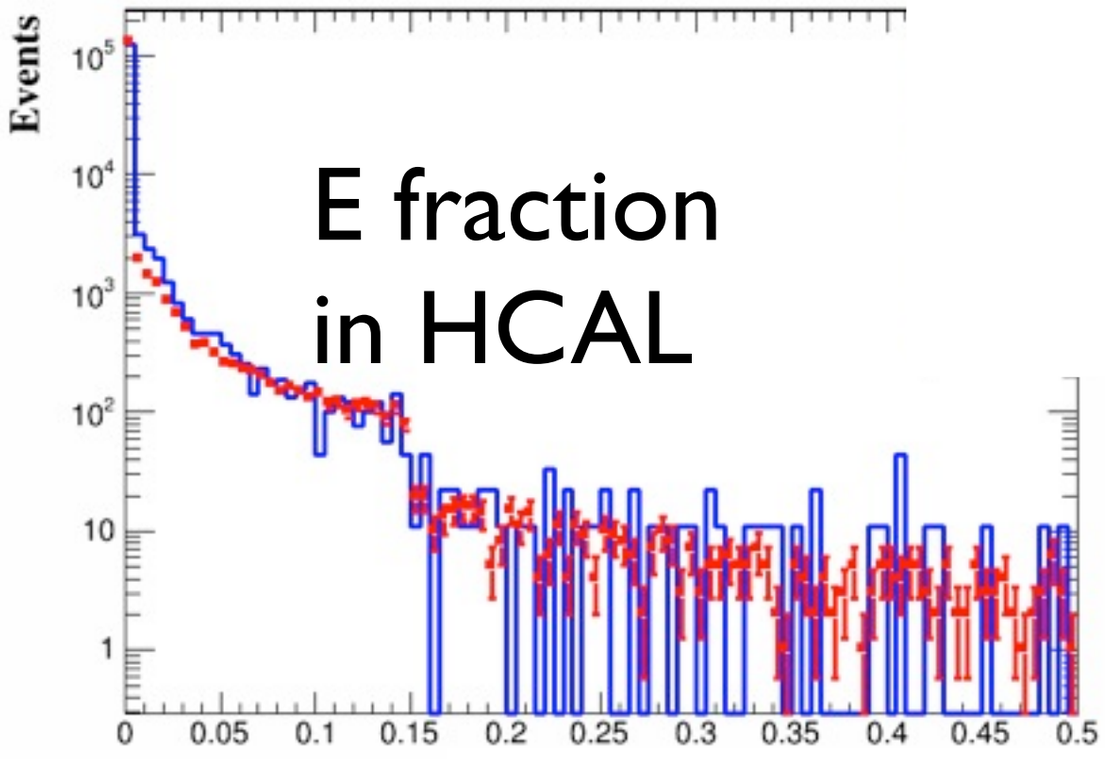


electron id

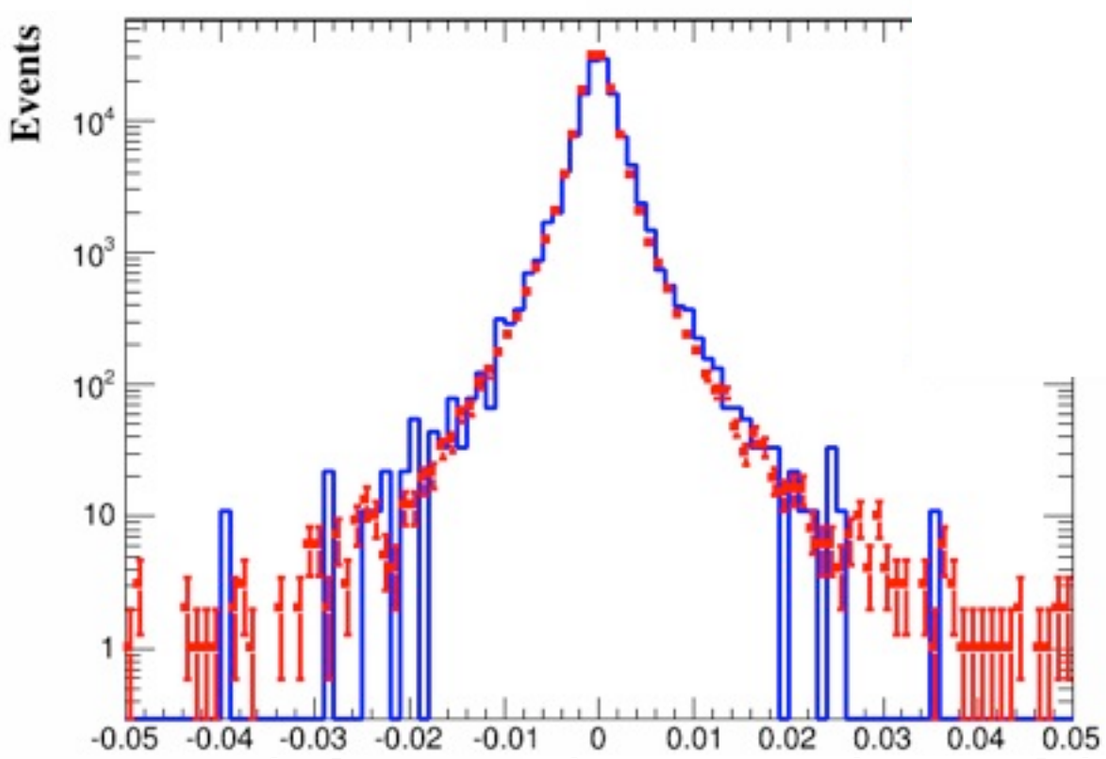
Z(ee)
- fast sim
- full sim



phi(cluster) - phi(track)

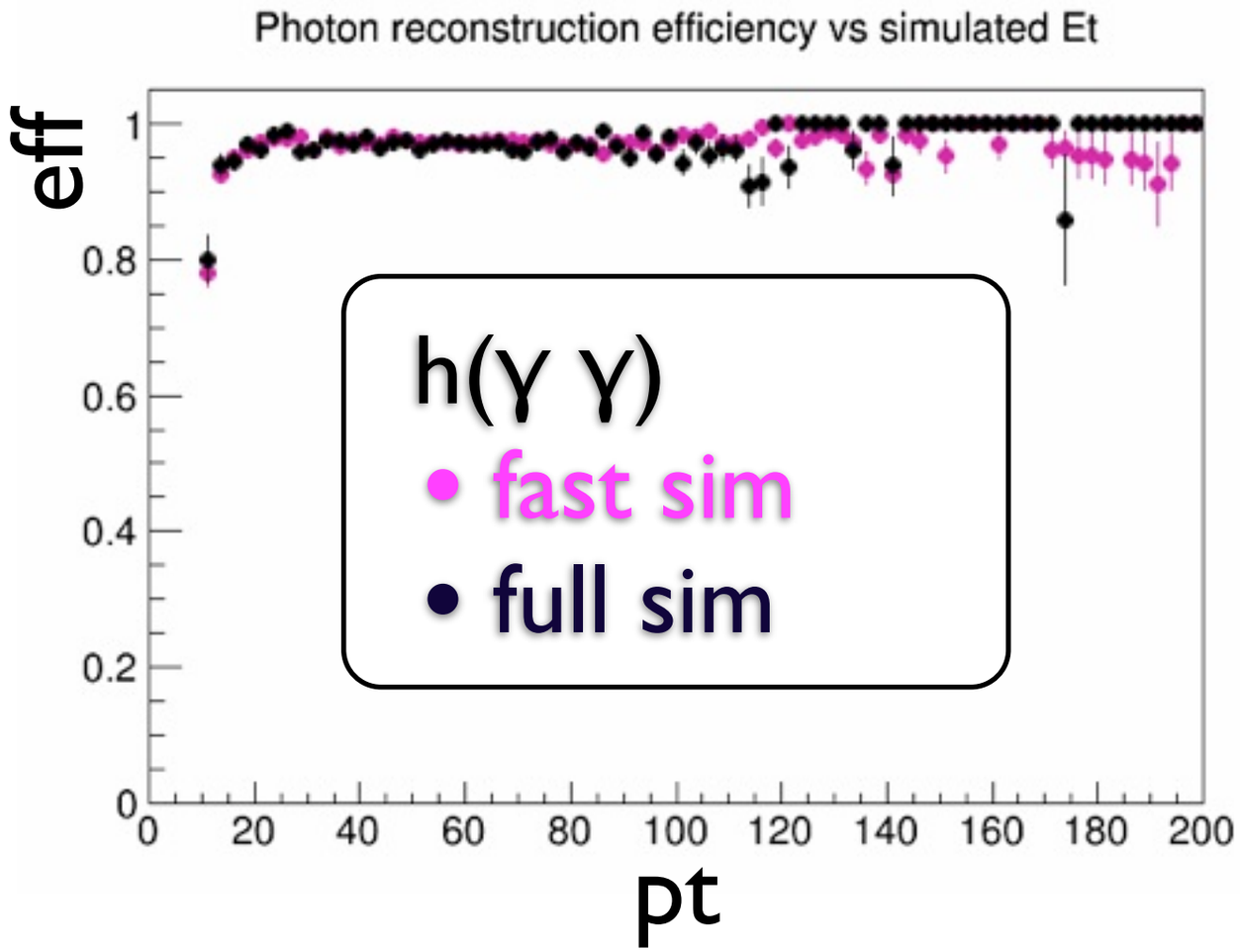


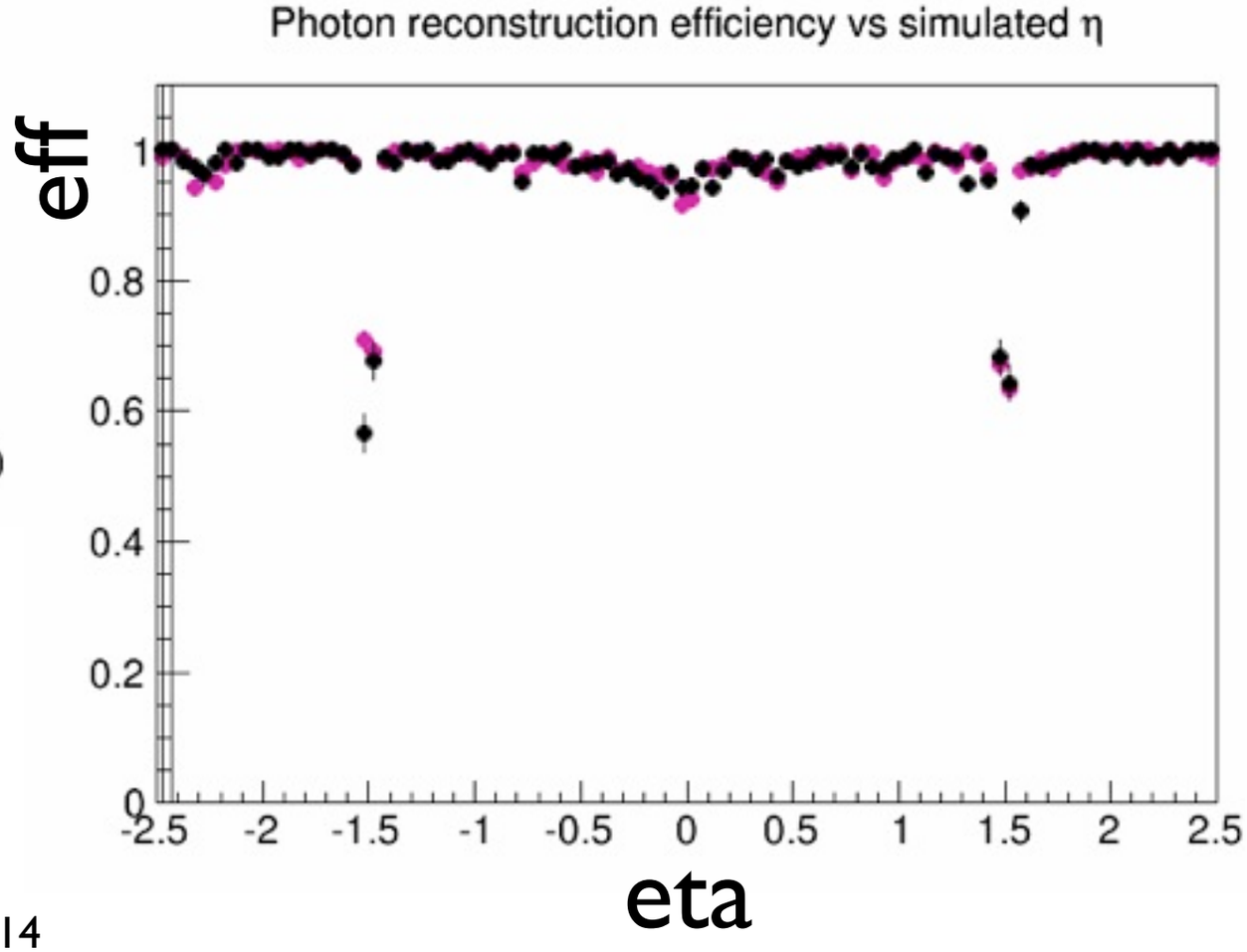
E fraction in HCAL



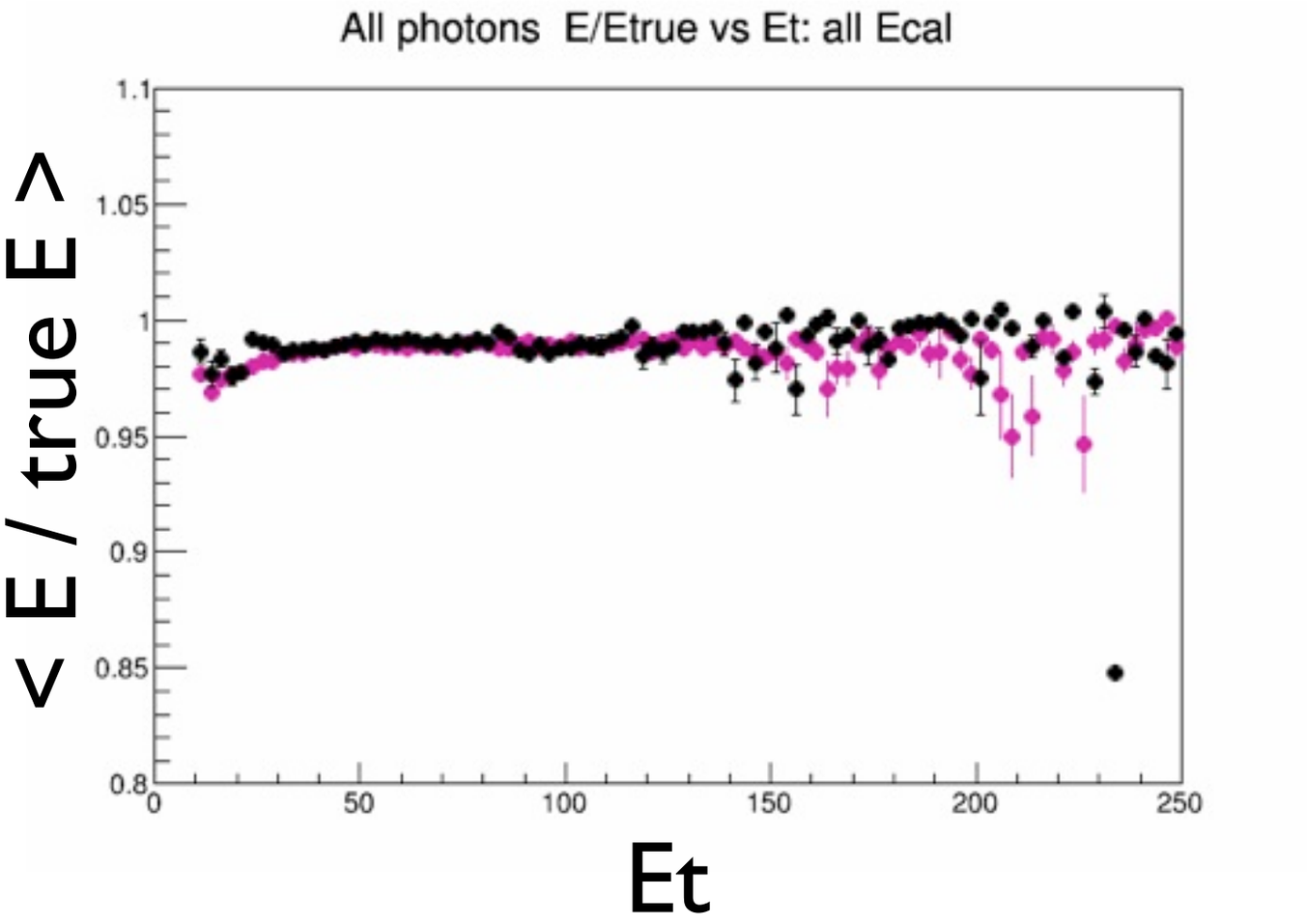
eta(cluster) - eta(track)

photons: efficiency

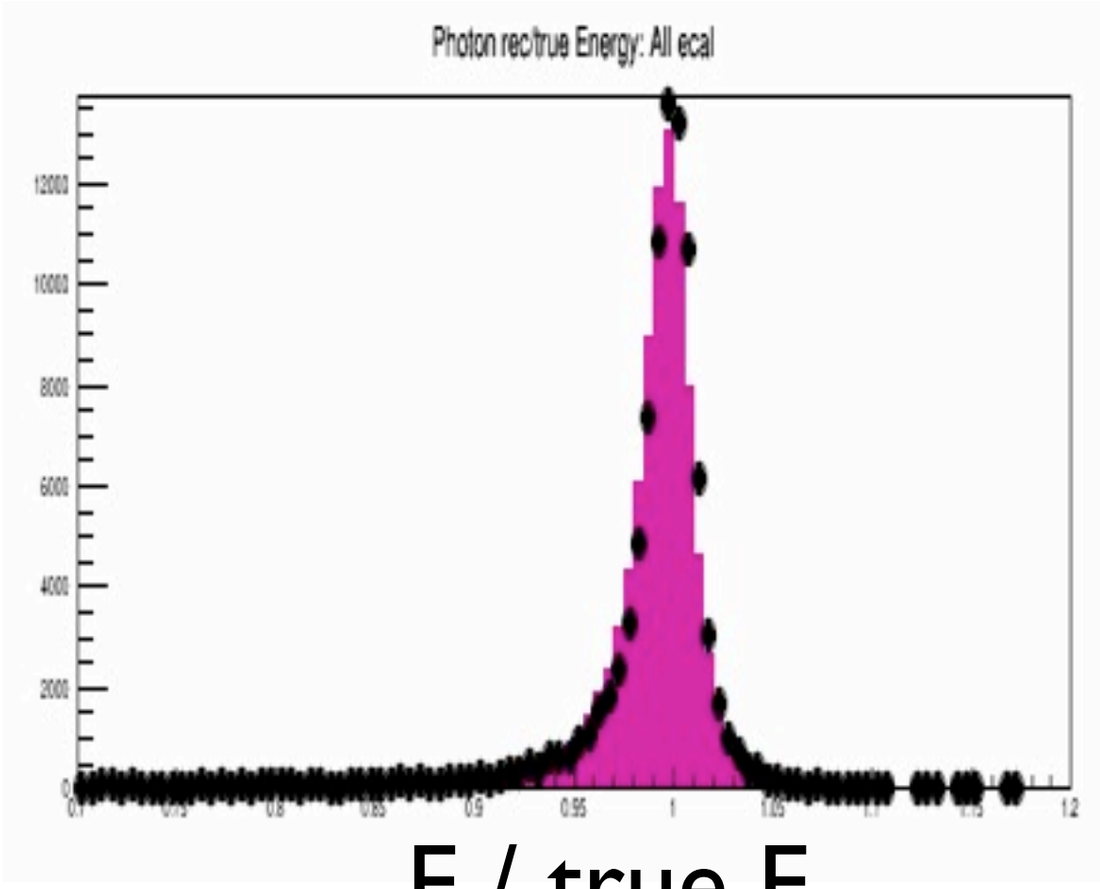




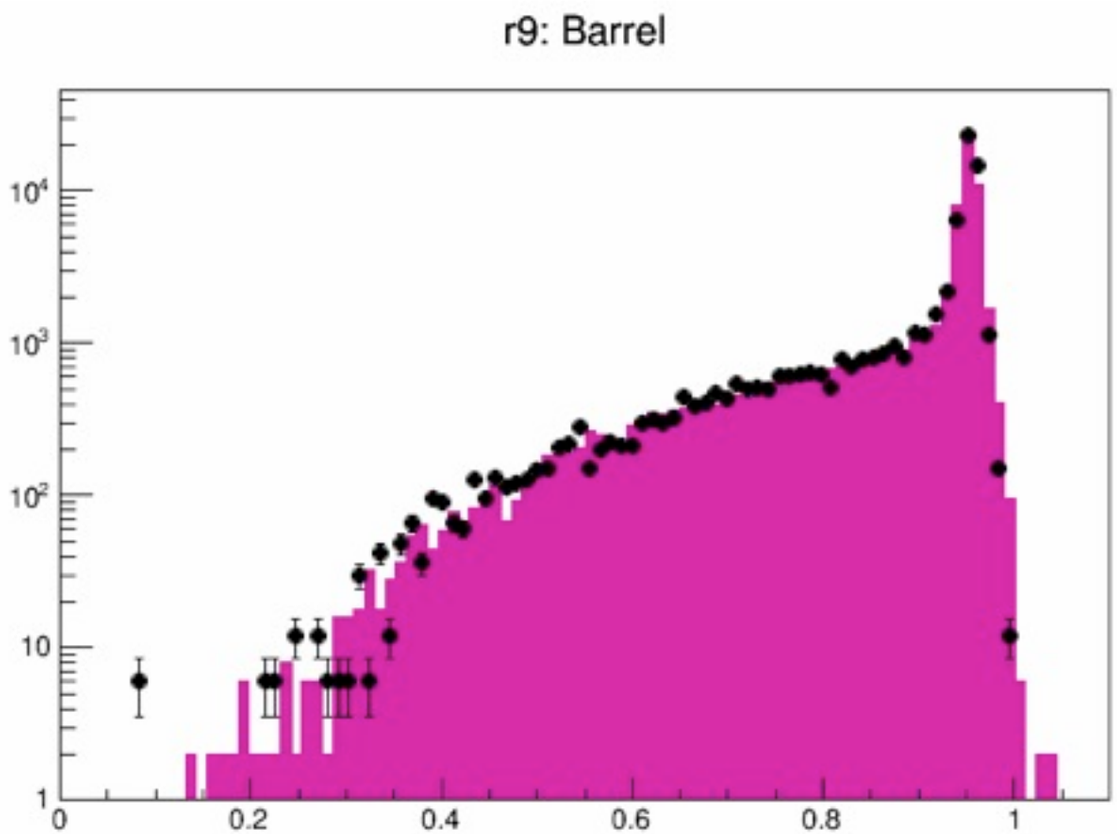
photons: energy



h(γ γ)
- fast sim
- full sim



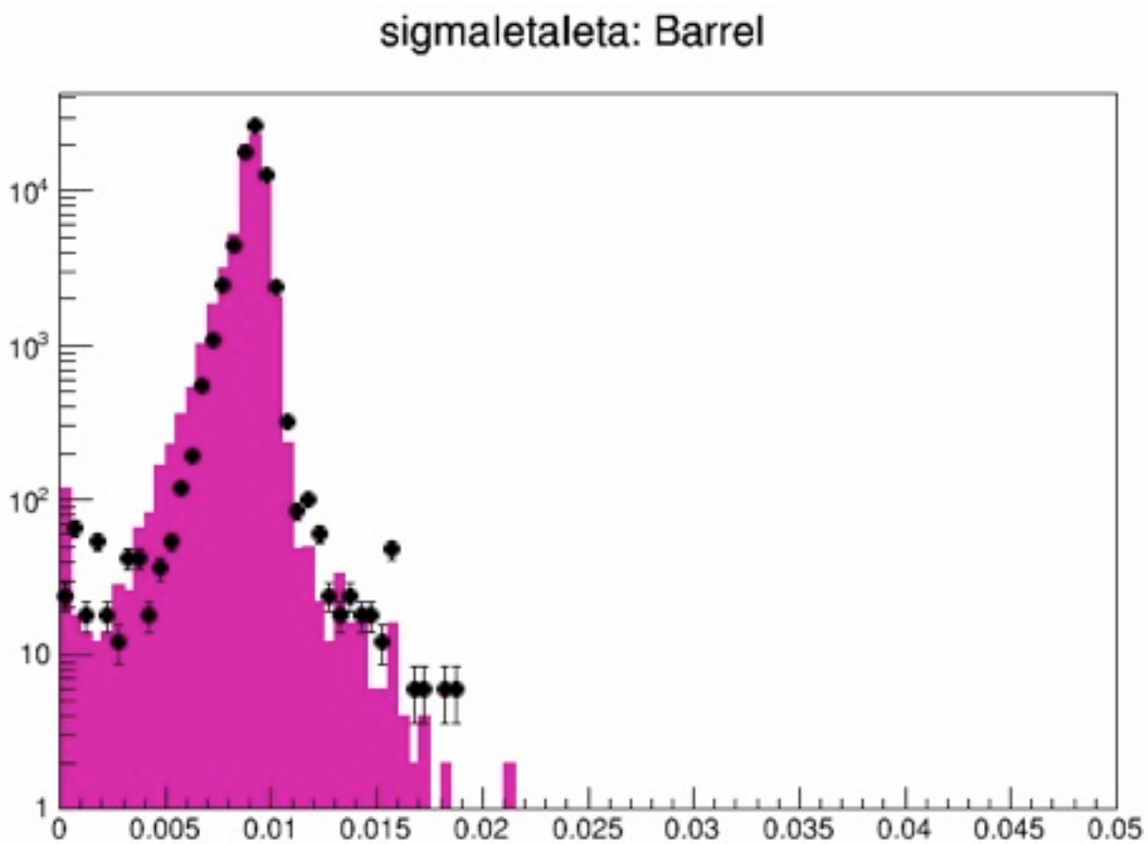
photon ID



R9

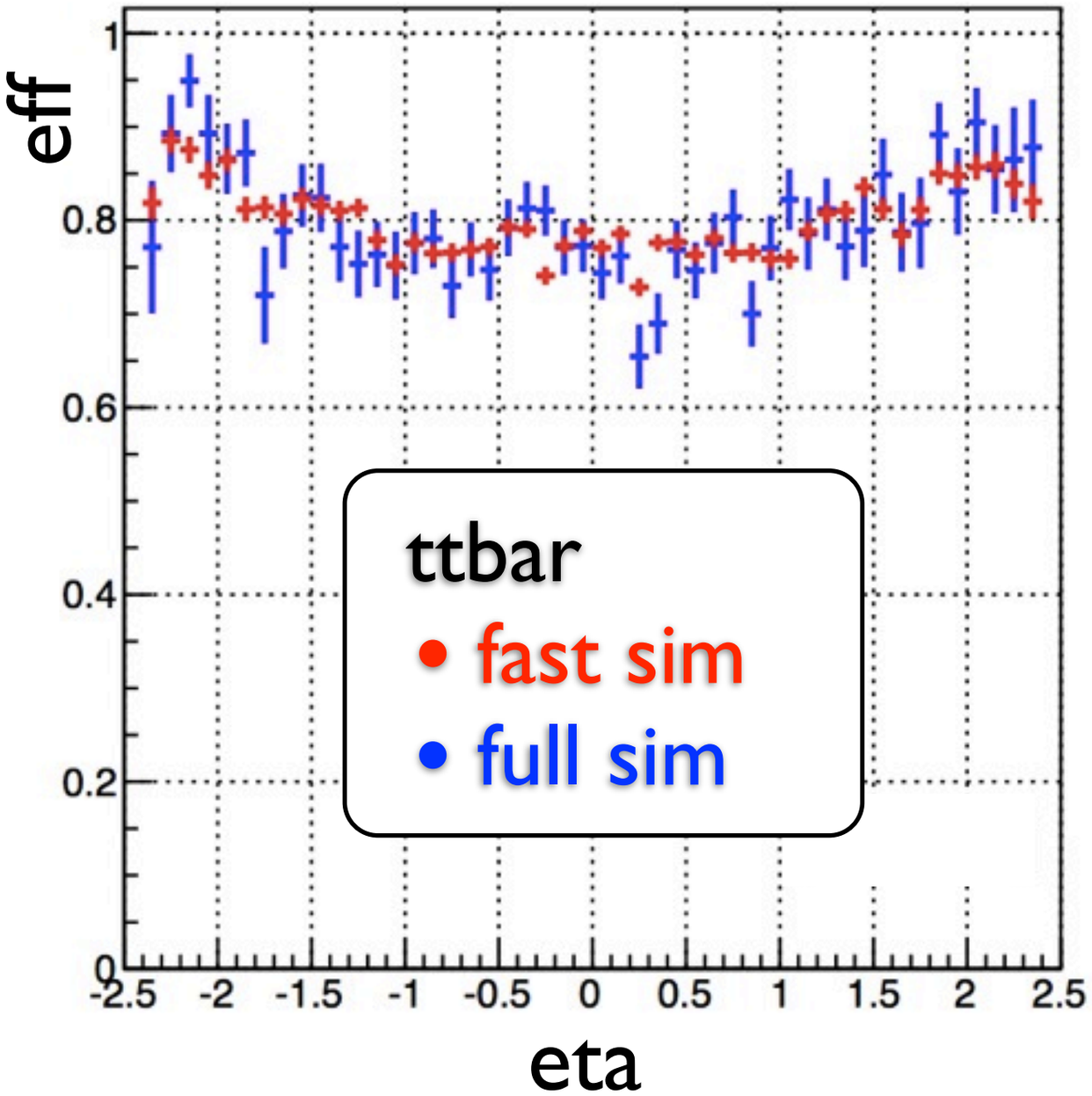
h(γ γ)

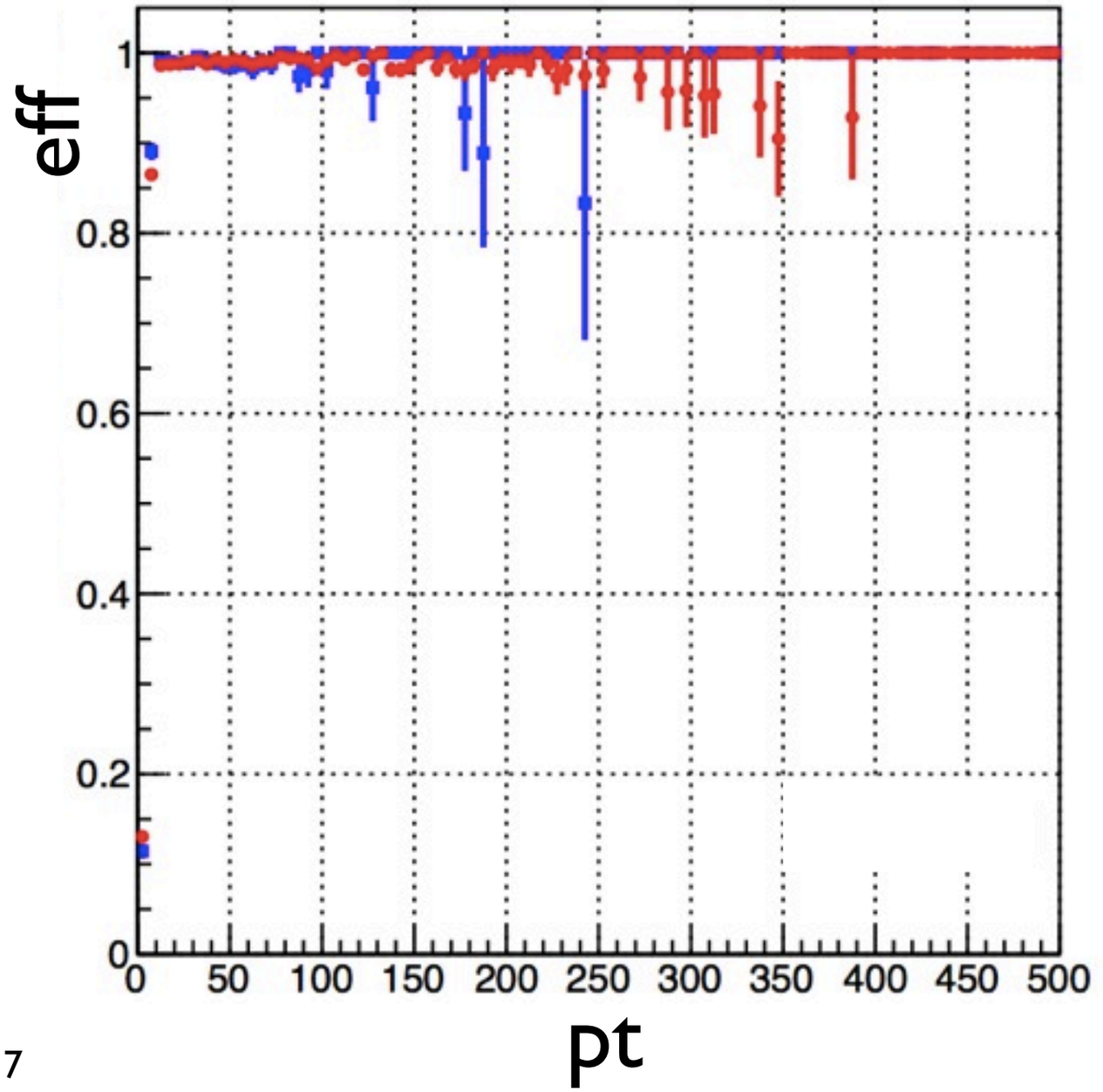
- fast sim
- full sim



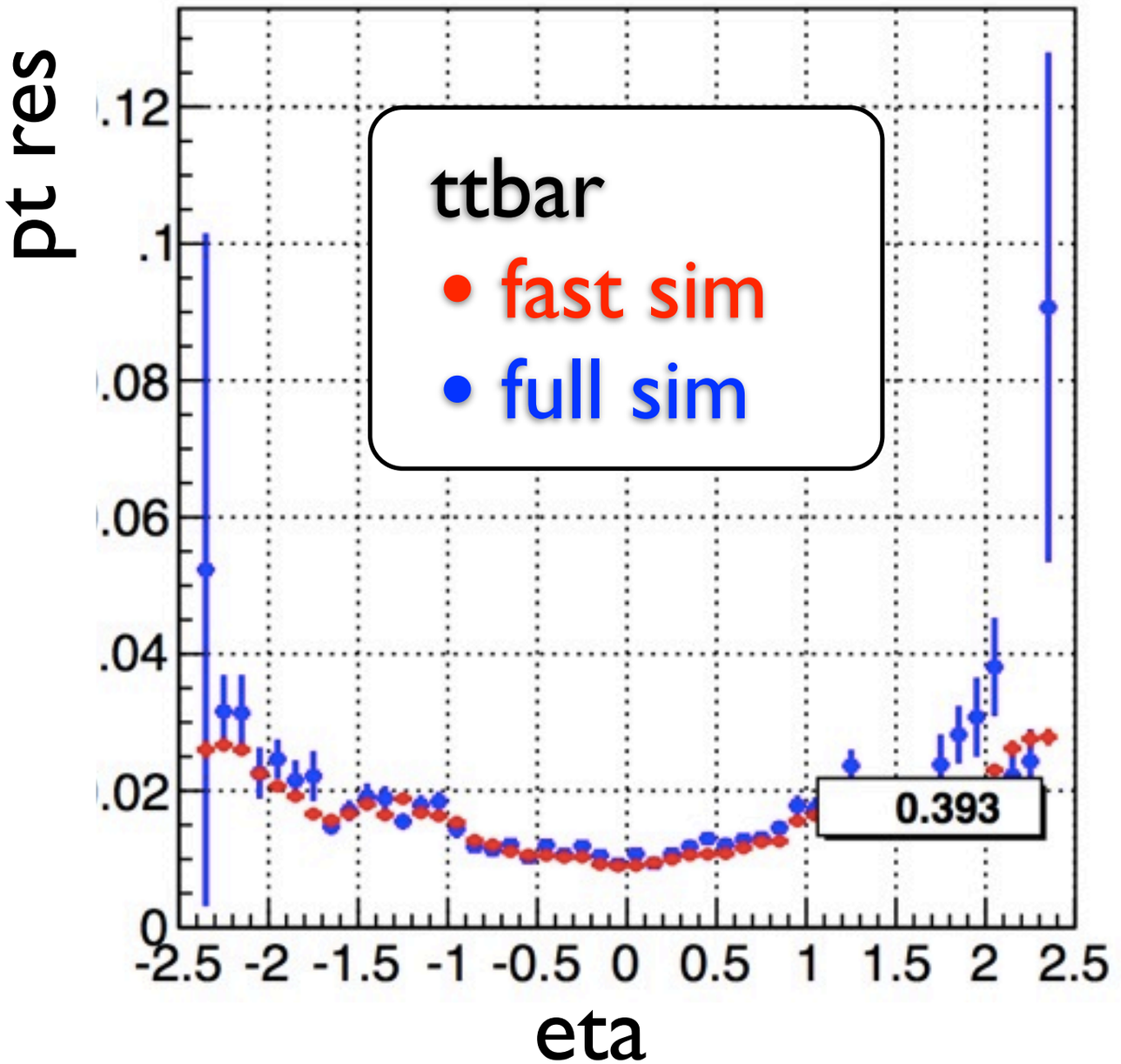
sigmaletaleta

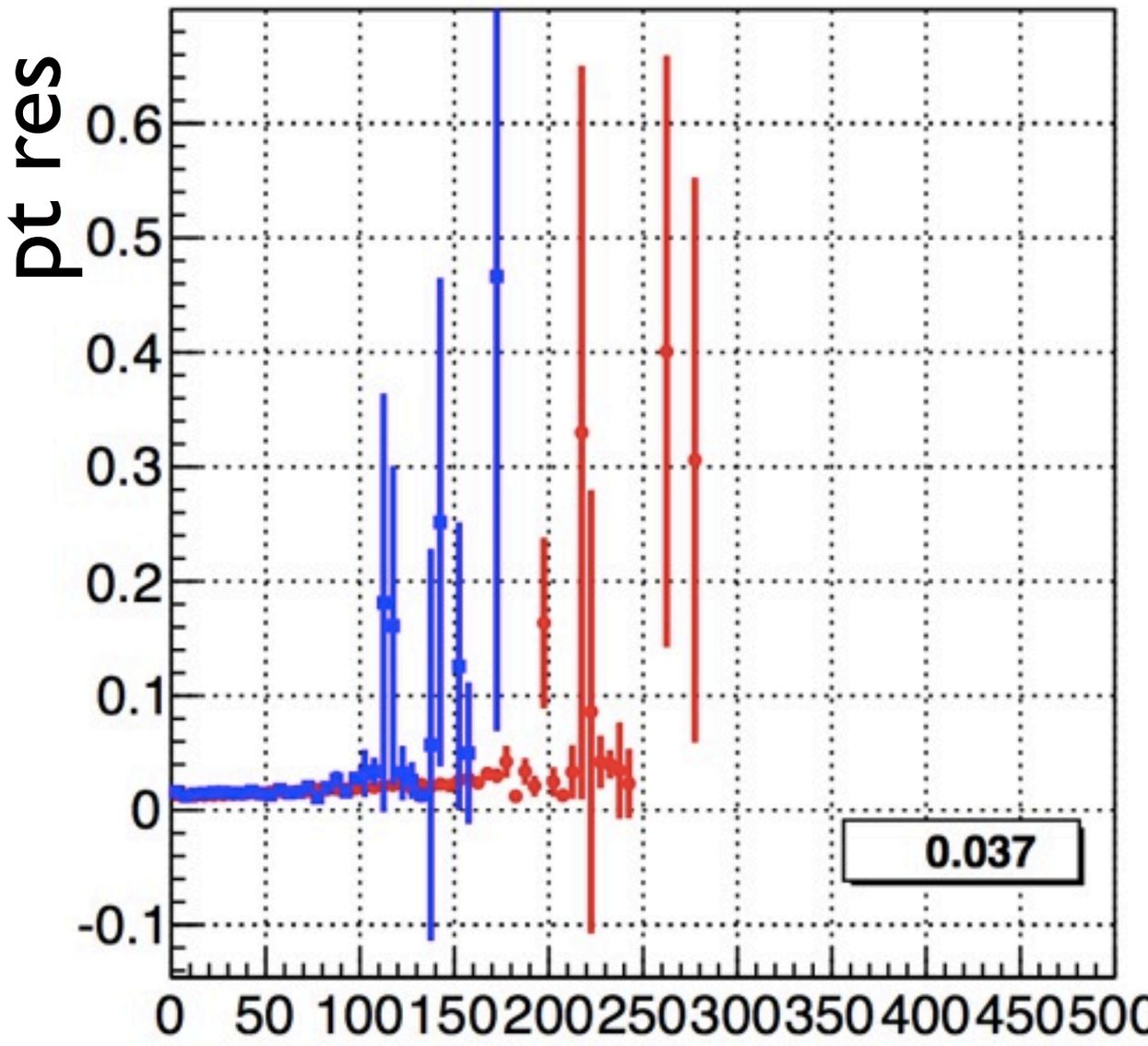
muon eff


eff
1
0.8
0.6
0.4
0.2
0
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
eta
ttbar
• fast sim
• full sim

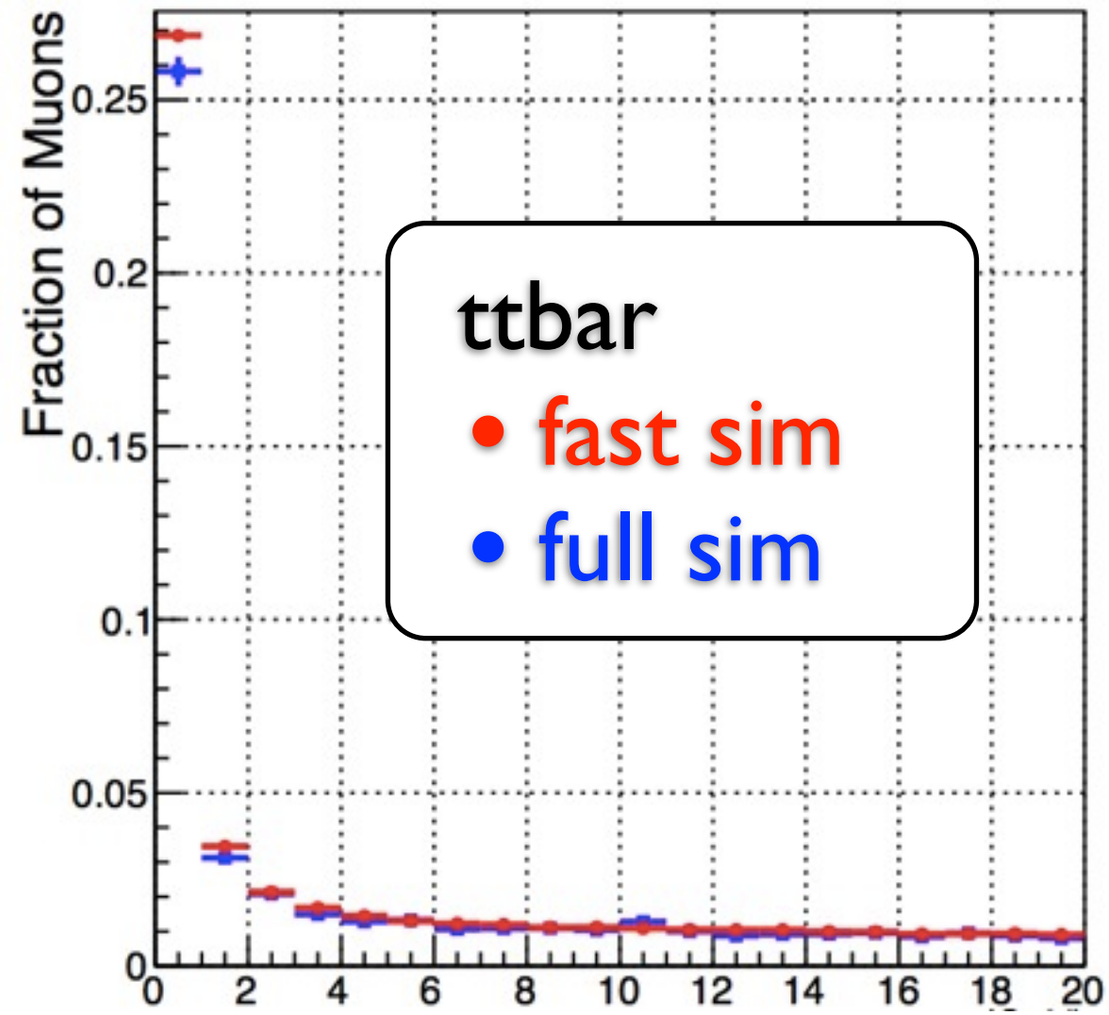

eff
1
0.8
0.6
0.4
0.2
0
0
50
100
150
200
250
300
350
400
450
500
pt

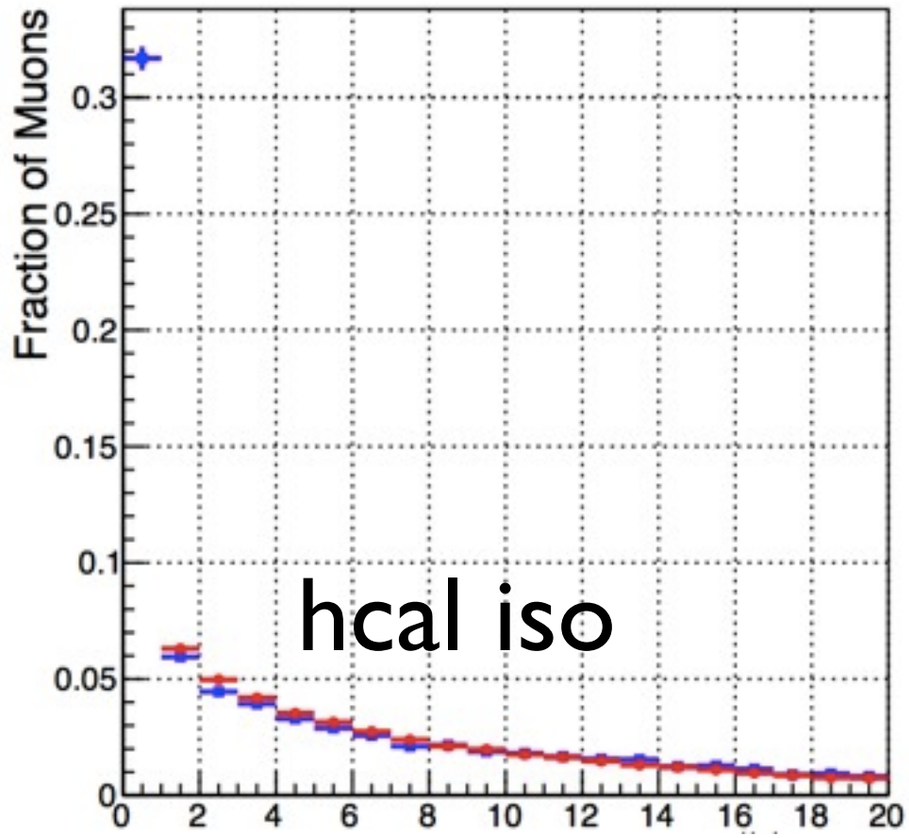
muons resolution

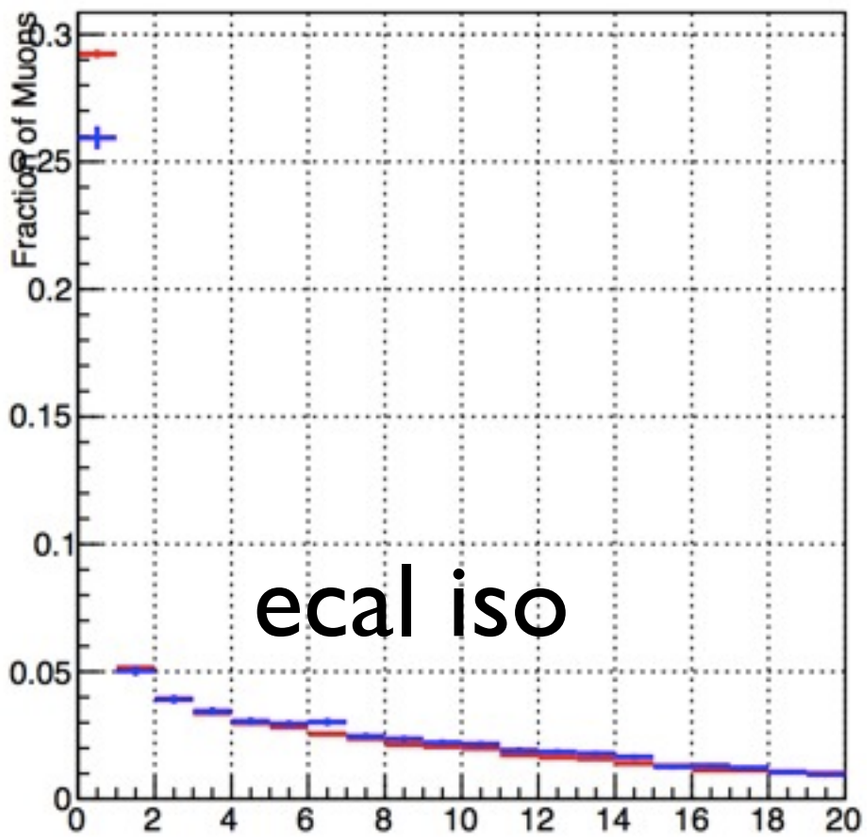




muons: isolation

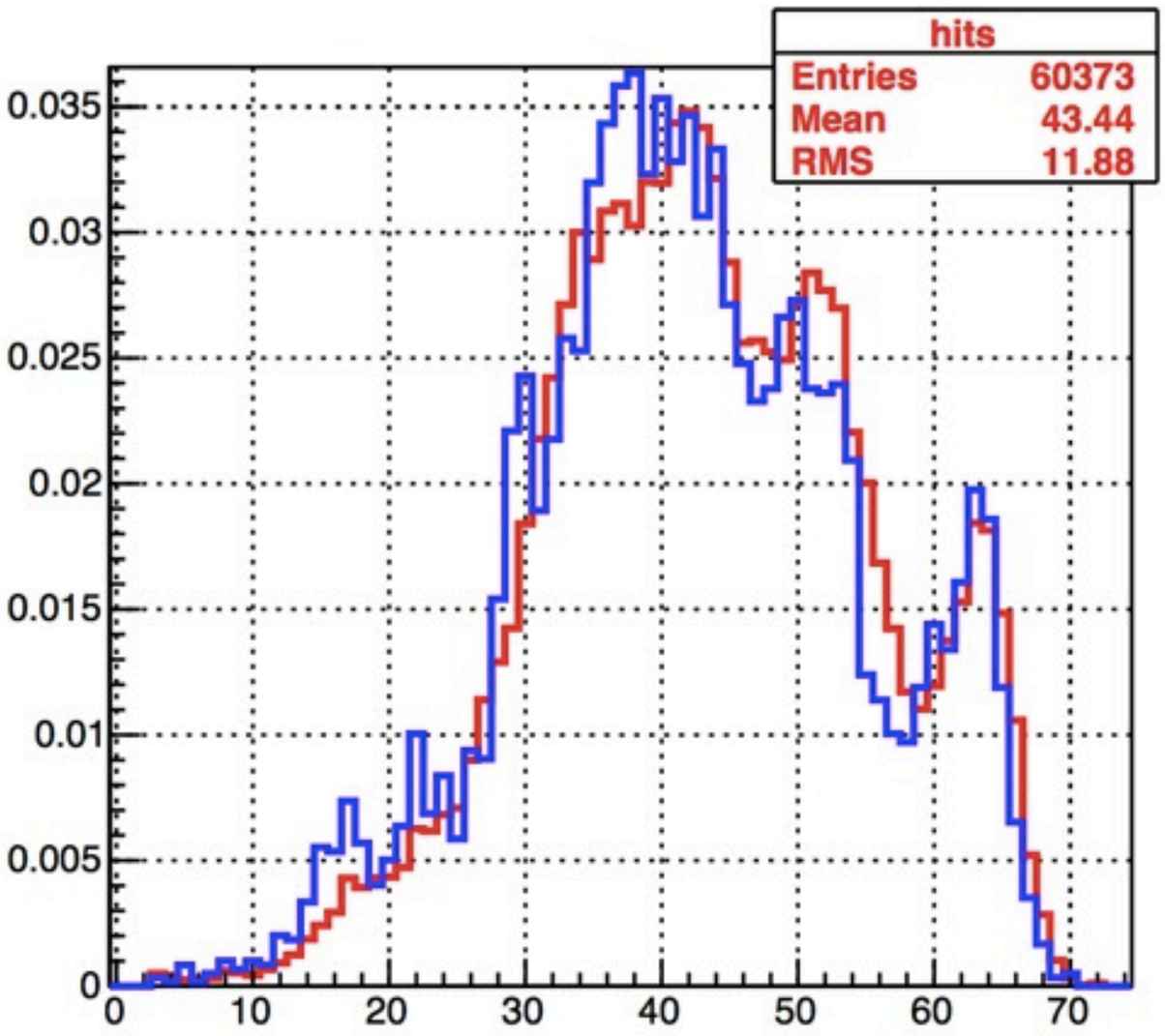




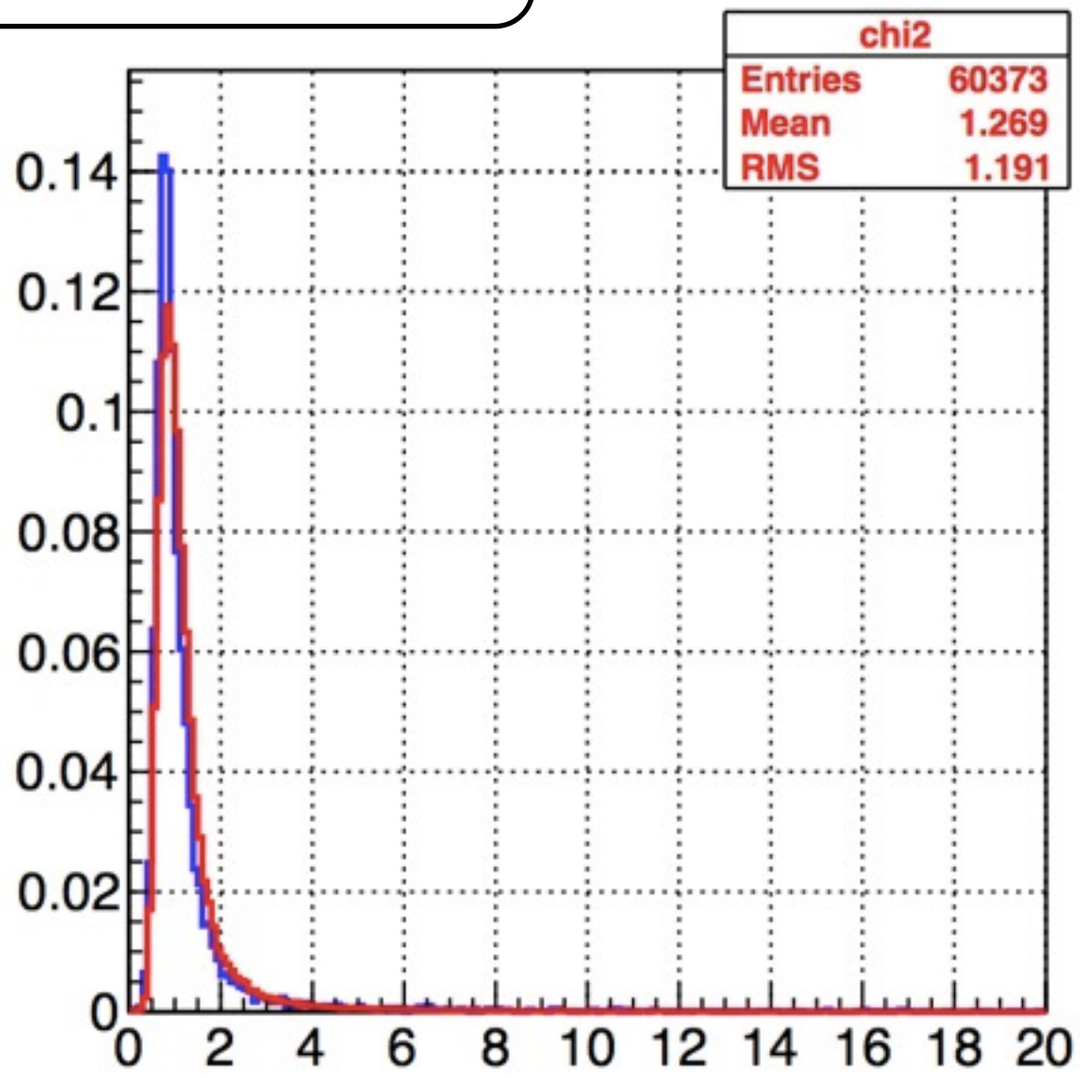


muons: id

ttbar
- fast sim
- full sim



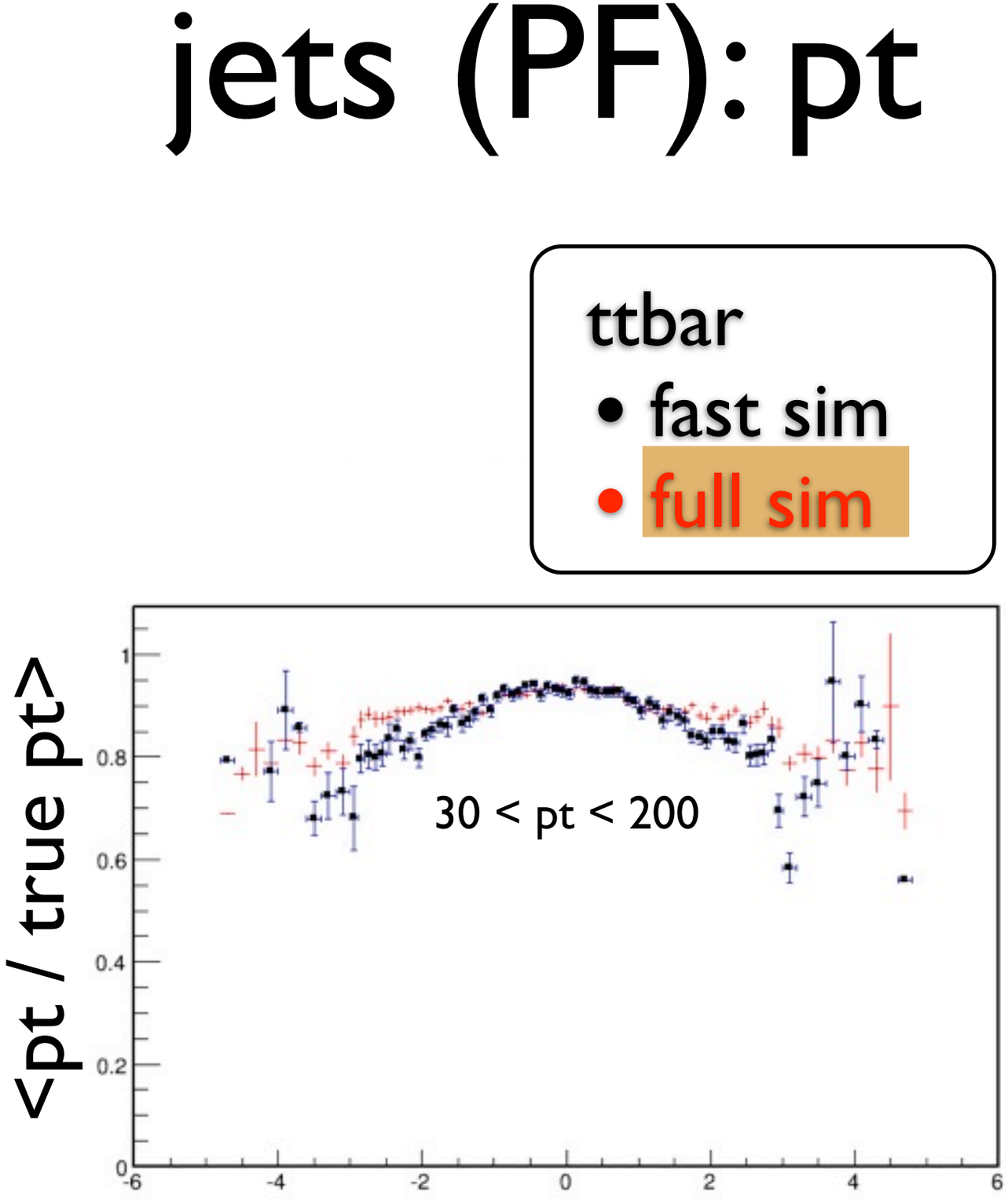
num of hits

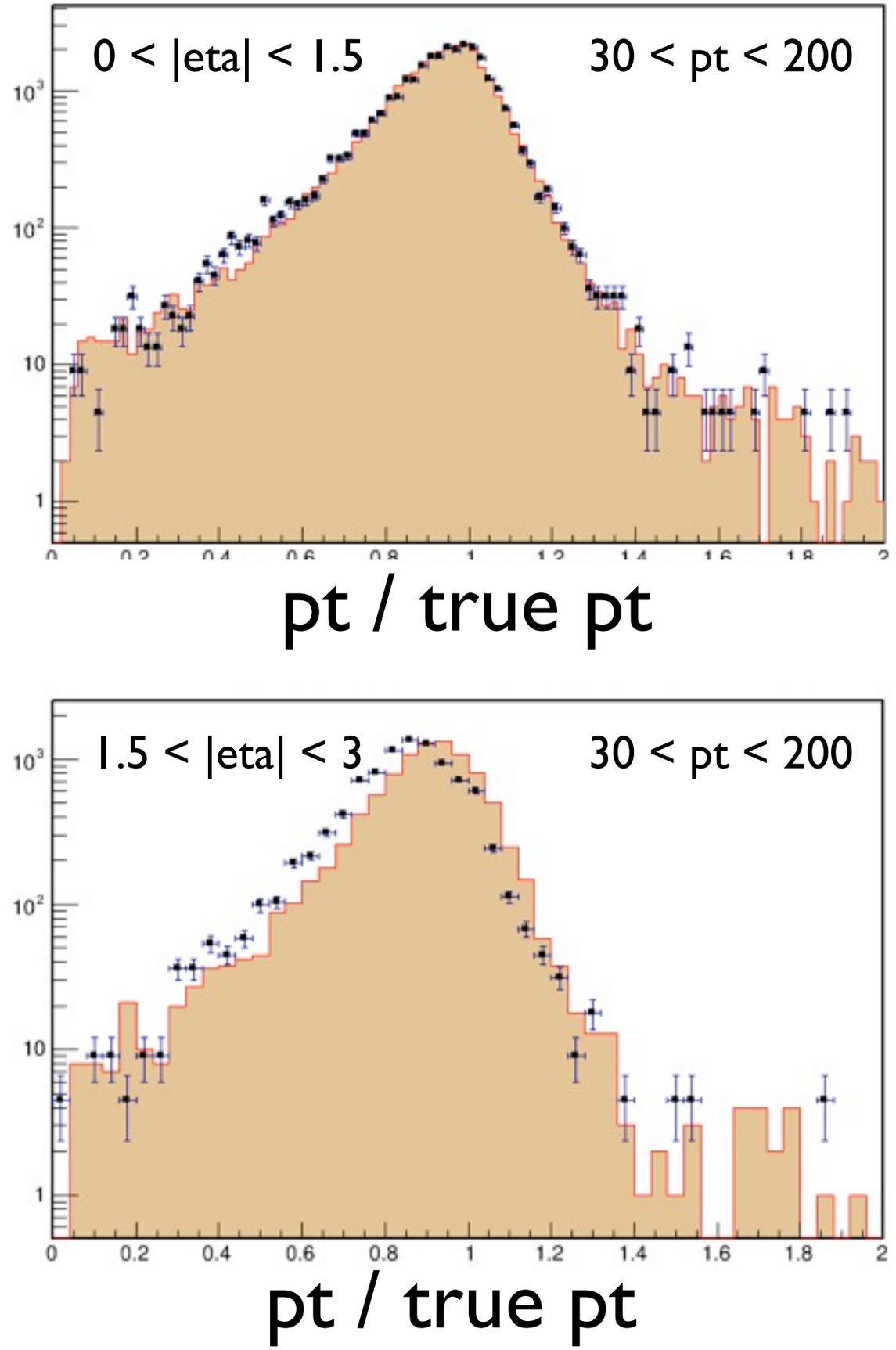


chi2 / ndof

jets (PF): pt

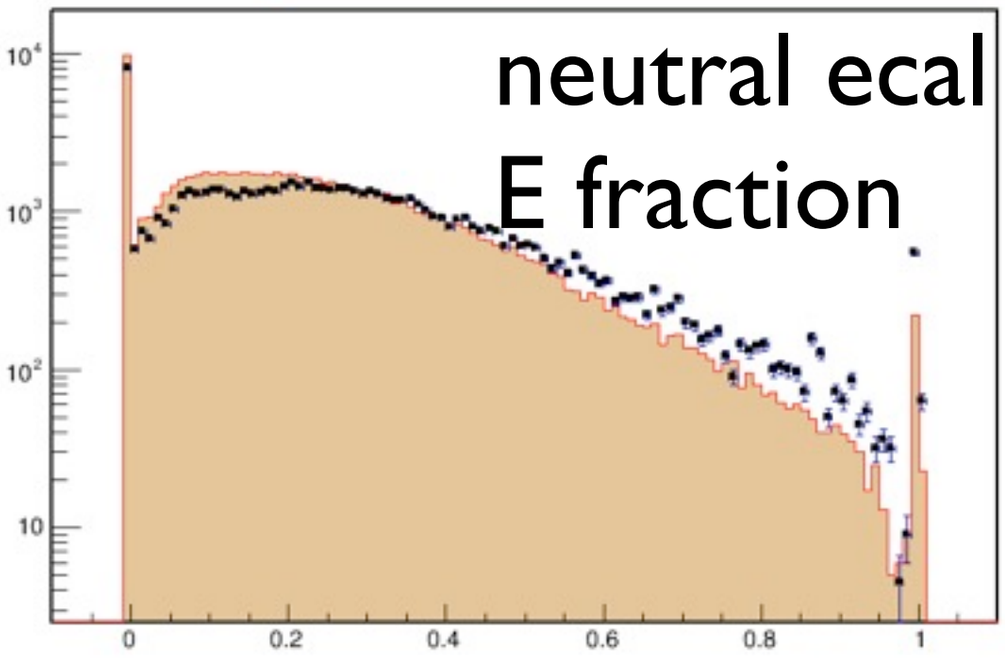
ttbar
- fast sim
- full sim

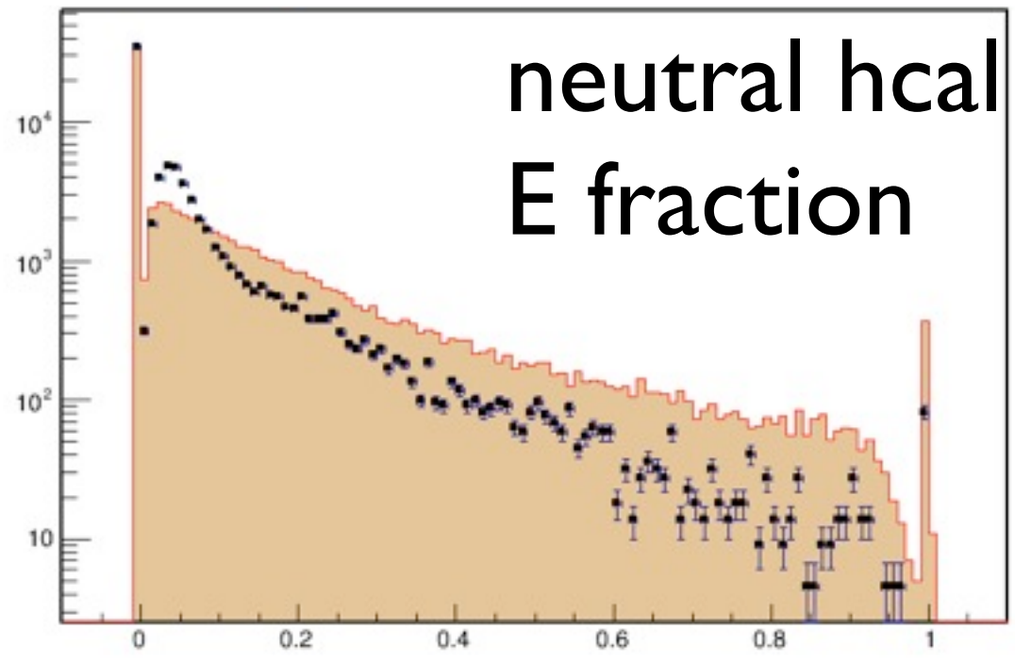


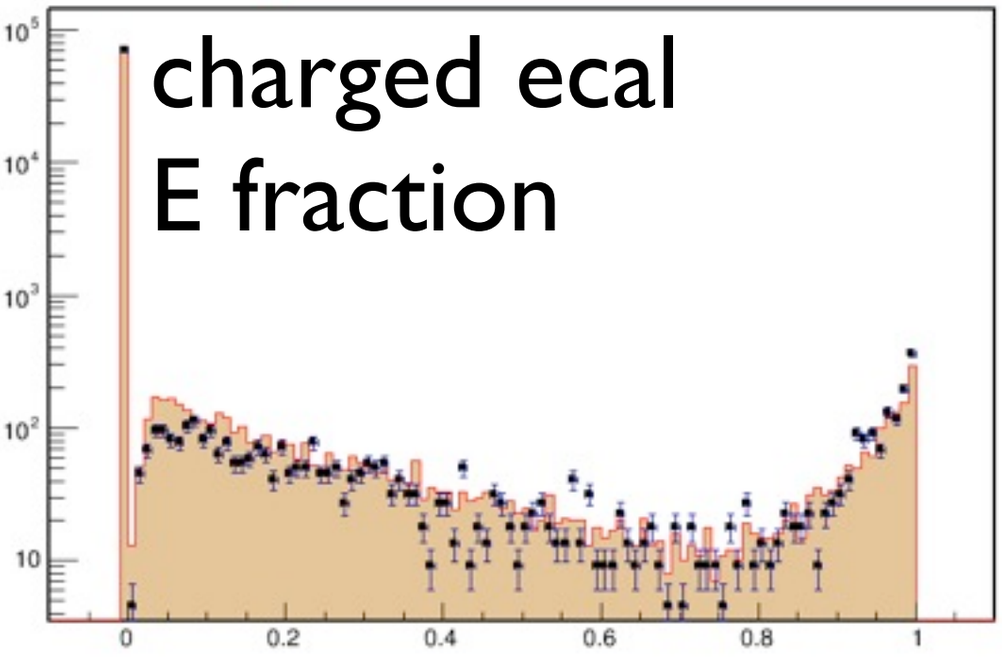


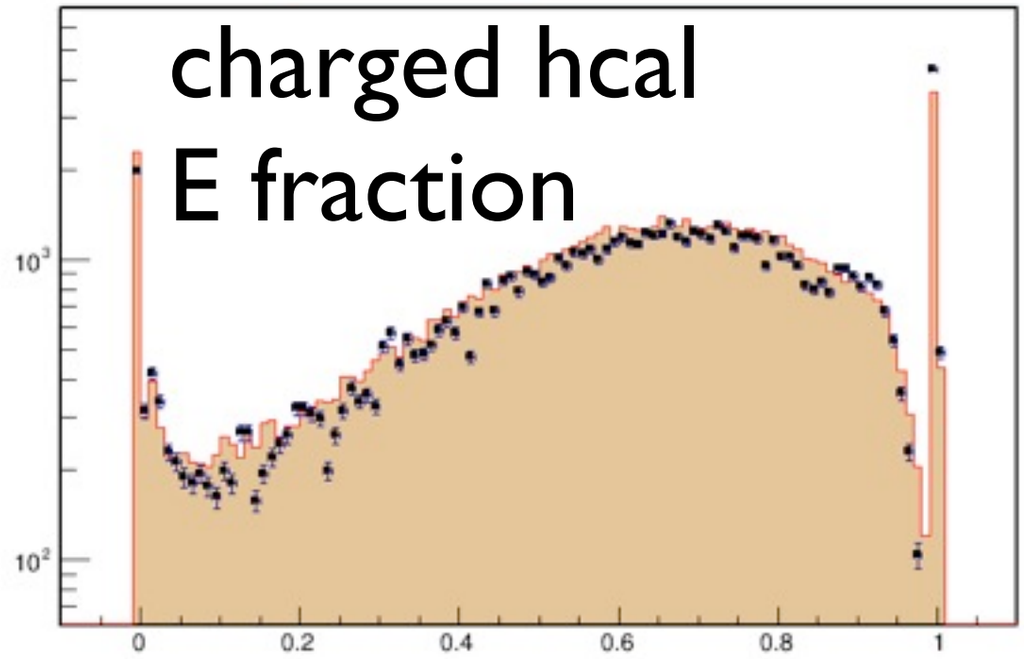
jets (PF): id

ttbar
- fast sim
- full sim

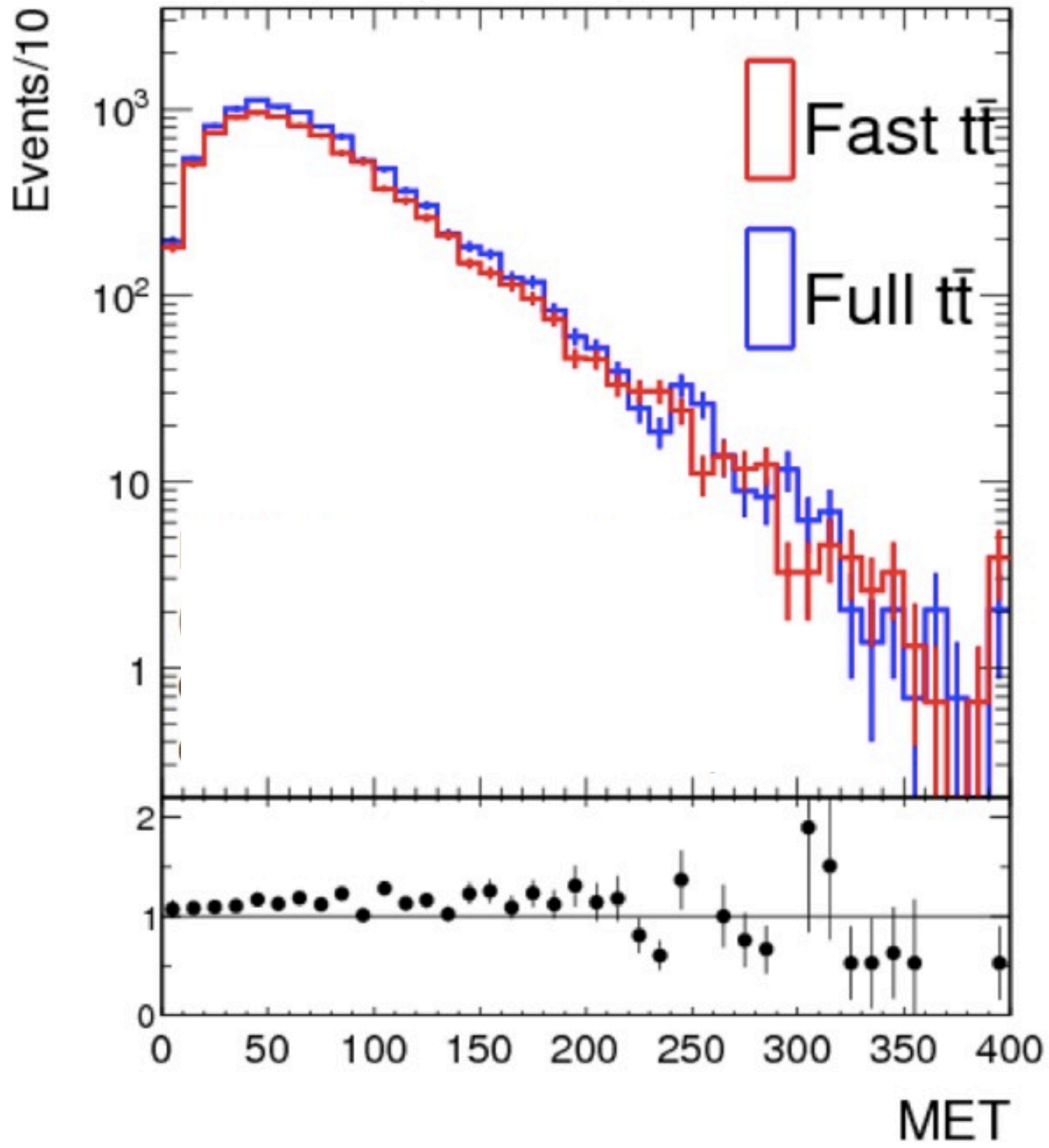








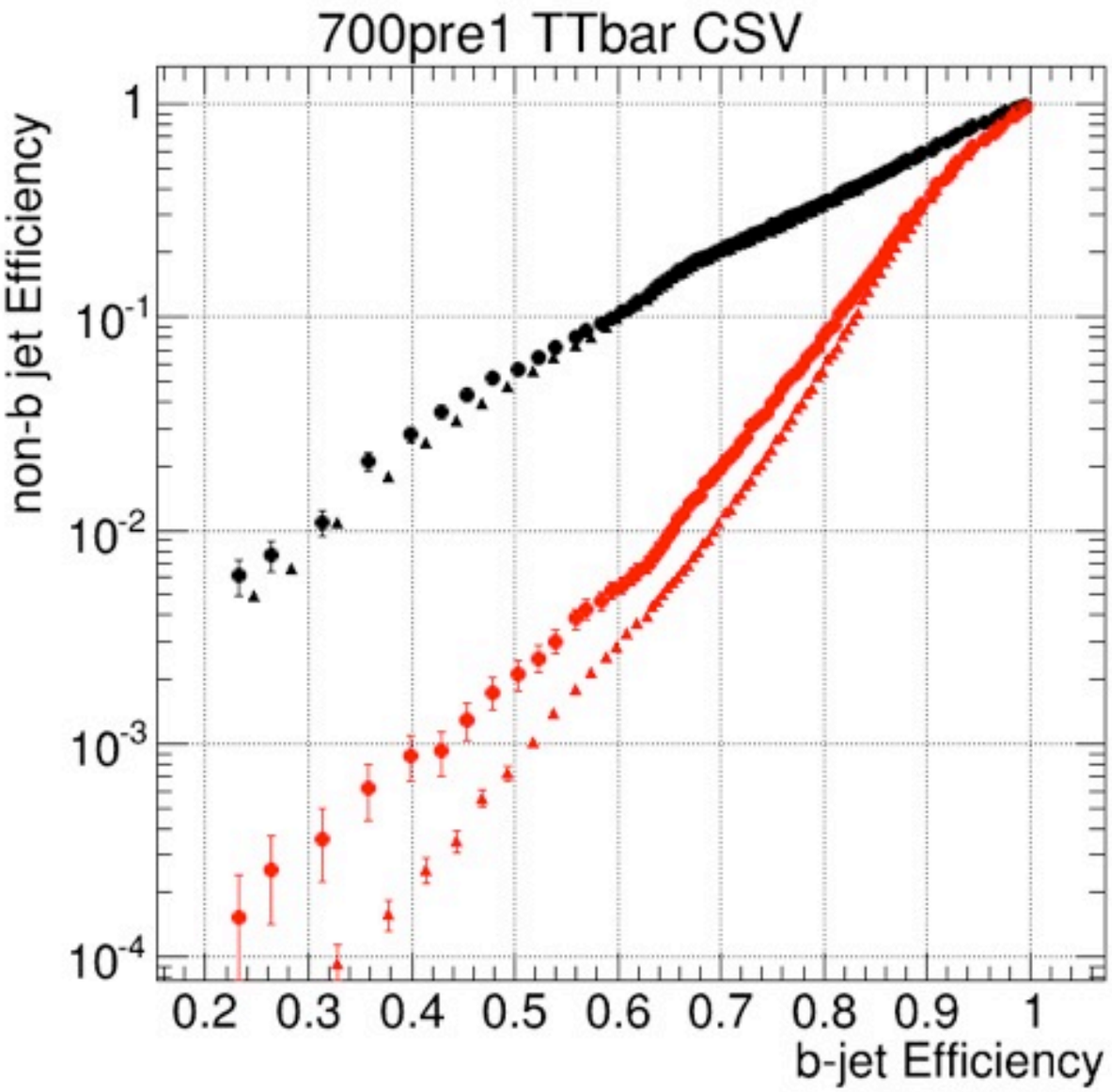
MET

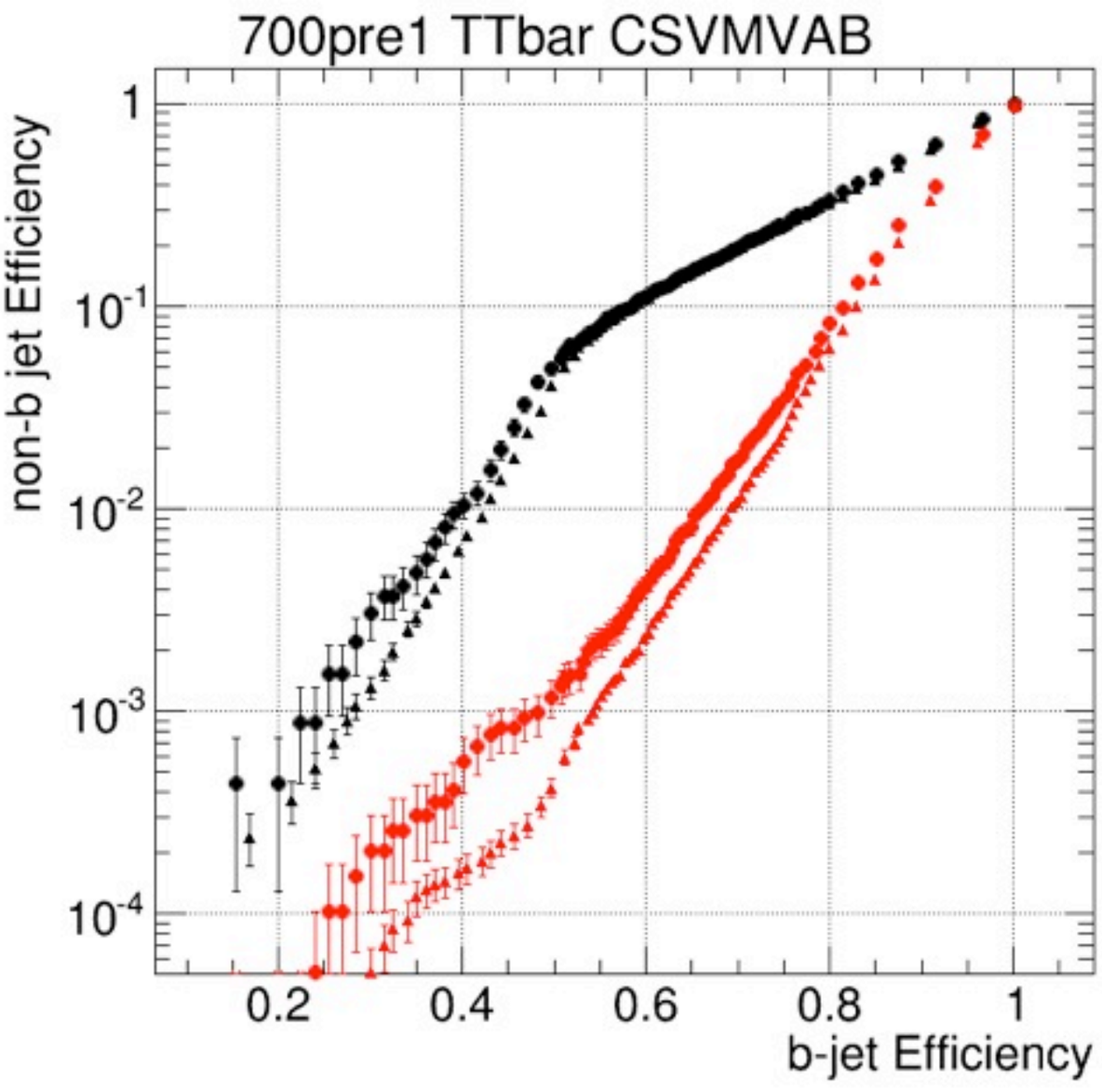

Events/10
10^3
10^2
10
1
Fast $t\bar{t}$
Full $t\bar{t}$
2
1
0
0
50
100
150
200
250
300
350
400
MET

b-tagging

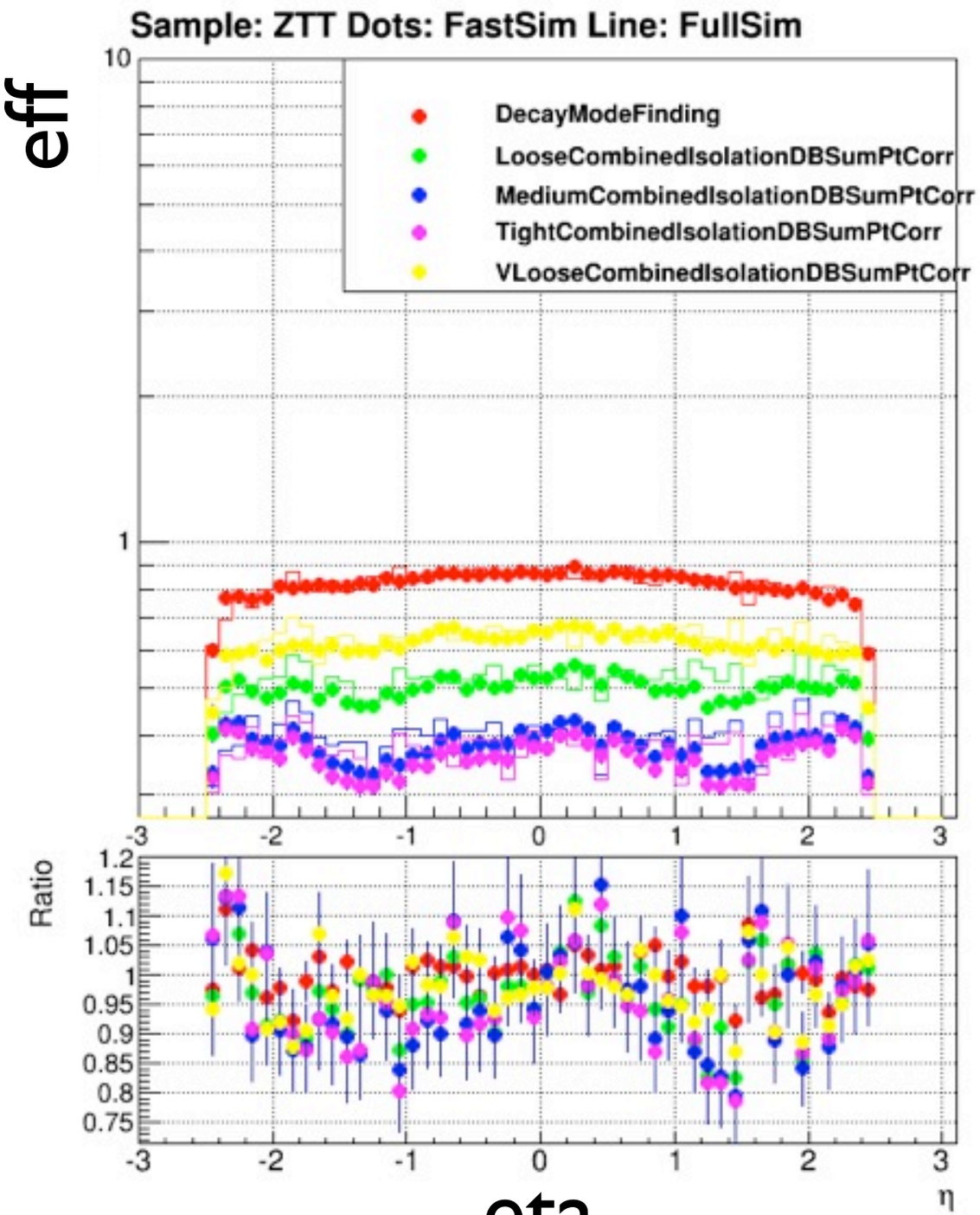
ttbar

- ▲ fast c-jets vs b-jets
- • full c-jets vs b-jets
- ▲ fast uds vs b-jets
- • full uds-jets vs b-jets

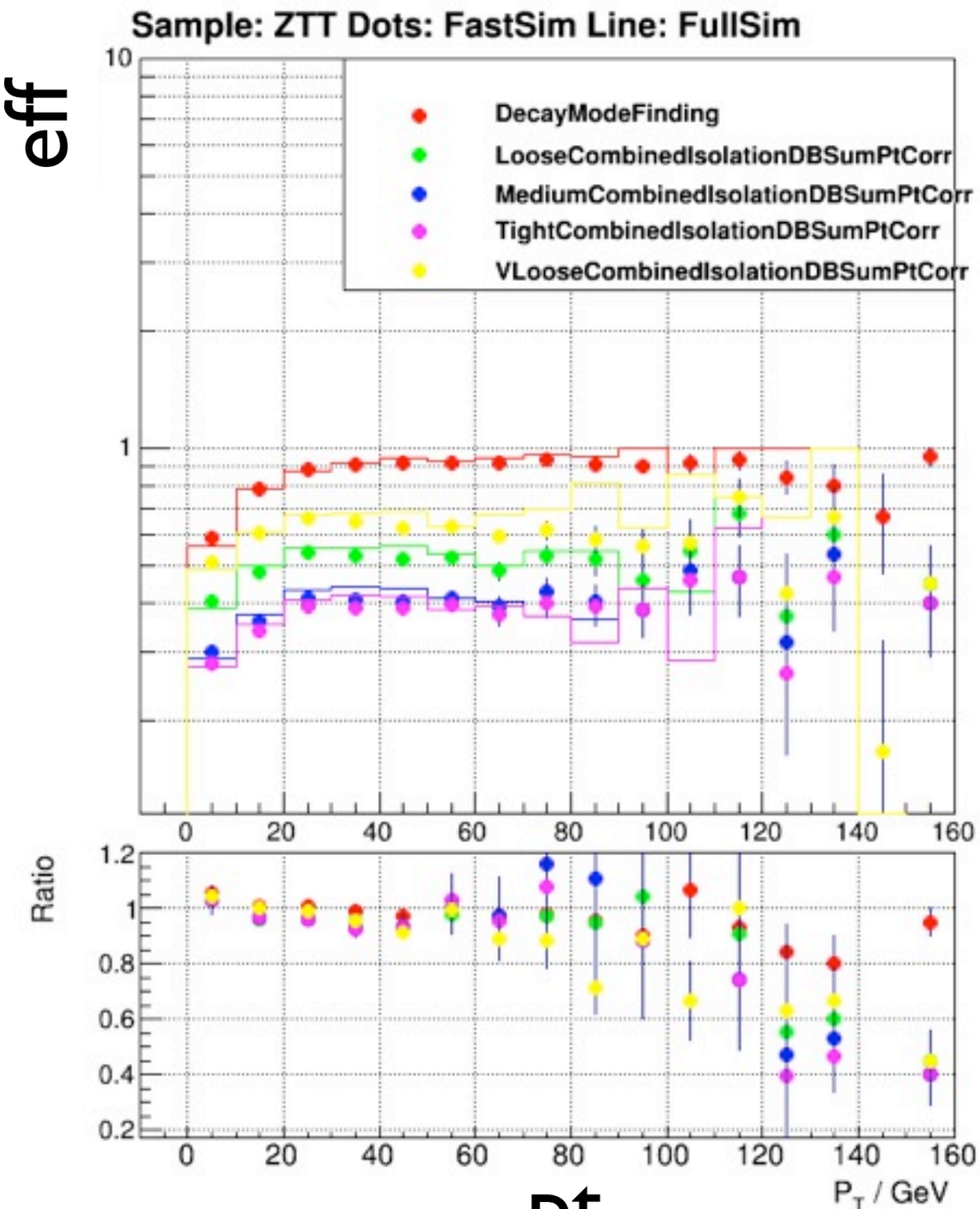




tau eff

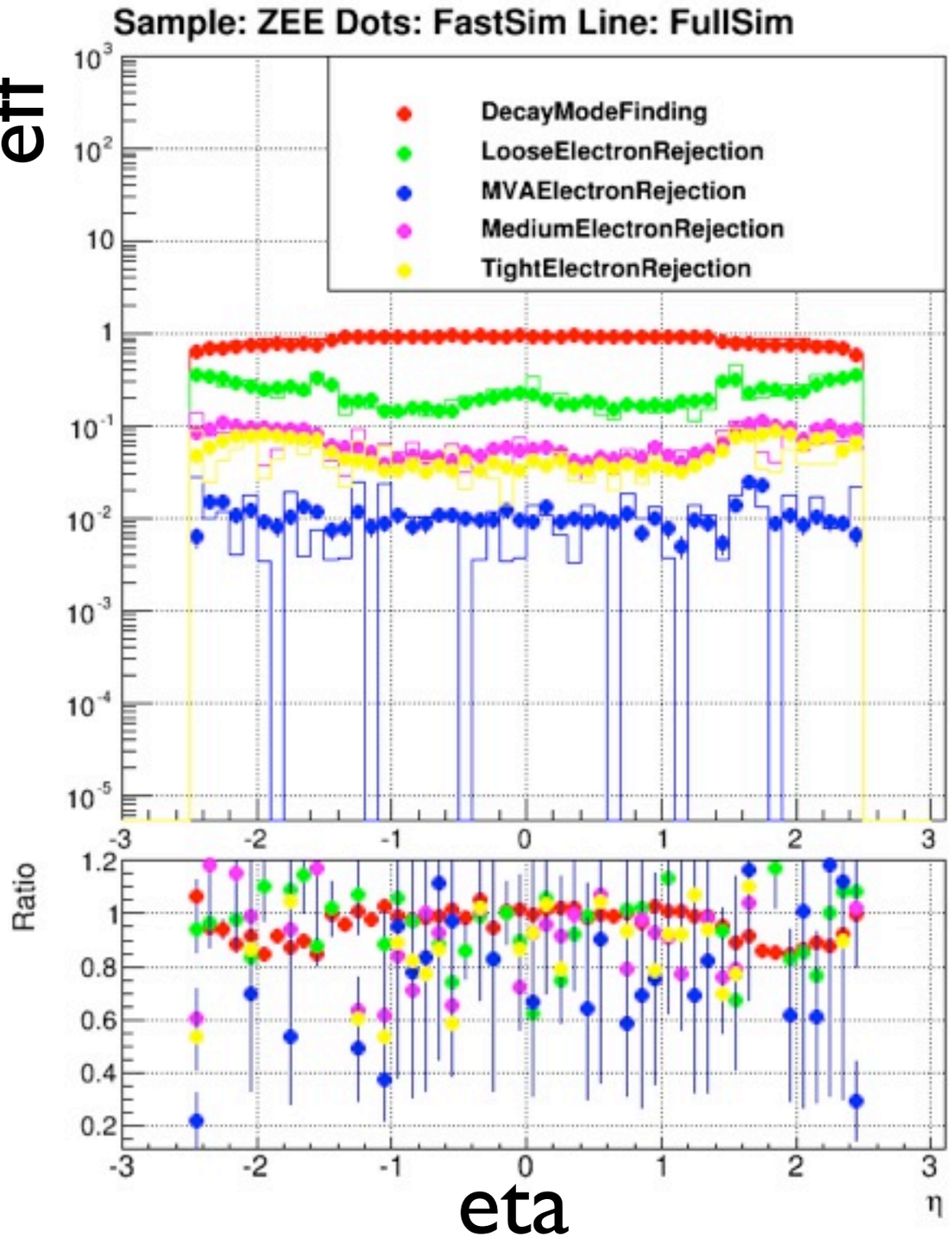


eta

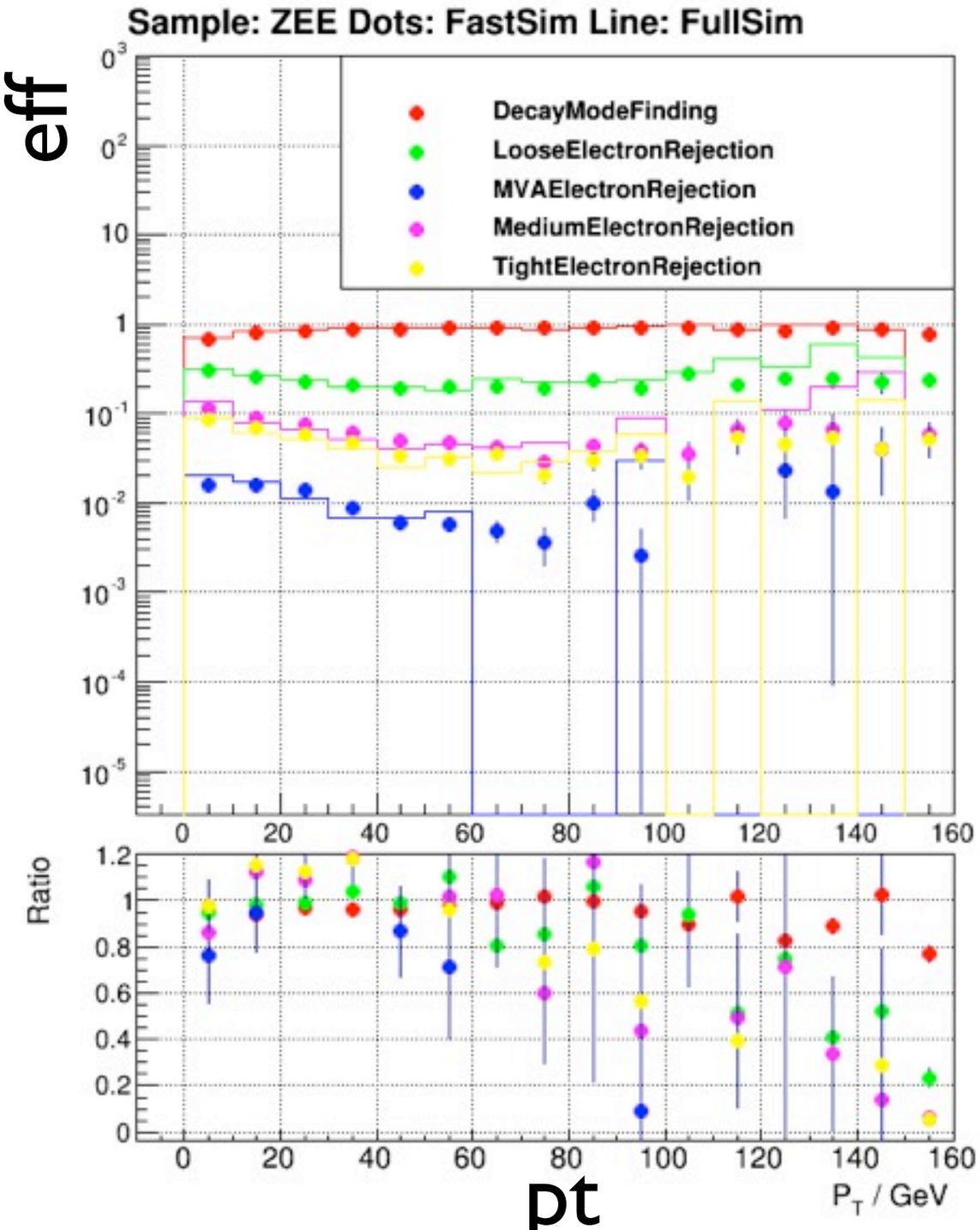


pt

tau fake rate



eta



pt

Last minute note

- How do physics analysis groups deal with fastsim inaccuracies?
  - they don't use fastsim
  - they use fastsim and introduce fastsim/fullsim or fastsim/data scaling factors and/or introduce fastsim related systematic uncertainties
- a good awareness of this could help setting priorities for development towards improved fastsim accuracy

Summary

- Validation of fastsim mostly based on fullsim
- fullsim vs fastsim comparison plots are produced and studied for most relevant variables for most relevant physics objects
- with some exceptions, fastsim performance ranges between very reasonable and excellent: for basic kinematic variables AND identification variables