Electron efficiency in 2010 data: some observations and thoughts

Kalanand Mishra
Fermilab

Egamma subgroup meeting
(February 18, 2011)

Motivation and outline

✦Efficiency is an important − although often overlooked − part of physics analysis

- For precision EWK analyses $e^{\pm}$ efficiency is the most significant uncertainty.
  - see my presentation in Egamma forum on Nov 18 for detailed numbers

http://indico.cern.ch/subContributionDisplay.py?subContId=3&contribId=3&confId=113521

- For other CMS measurements/searches (e.g., top, Exotica, SUSU, Higgs) electron efficiency still constitutes a significant source of systematic uncertainty.

✦I think this sub-group should take over the centralized task of providing efficiency estimations for most common e/γ selections being used in CMS. More on this at the end of this presentation.

Outline of this talk:

- Tag & probe tool for precise efficiency measurement
- Methodology to apply data-driven efficiency in analysis
- A quick look at the efficiency and scale factors in 2010 data
- Plans and logistics

Tag&probe machinery: PhysicsTools/TagAndProbe

An important tool for precision efficiency measurements

- *PhysicsTools/TagAndProbe* is a lightweight ROOT-based framework that allows one to compute any kind of electron/muon efficiency
- Highly flexible design: user can easily configure
  - *tag, probe, passing probe* selections
  - which variables to store in the Tree
- User can compute efficiency as a function of any of the probe variables
  - in any number of dimensions, e.g., E_T, η, ϕ, nJets, MET, Ht,...
- Backend fitter allows user to specify how to estimate signal yields for numerator and denominator.
- User has full control over efficiency computation. No black boxes.
  - comes with official CMSSW releases, with periodic updates.
  - wide user base; adopted for the first EWK, Quarkonia, top papers.

Details at: https://twiki.cern.ch/twiki/bin/view/CMS/TagAndProbe
https://twiki.cern.ch/twiki/bin/view/CMS/ElectronTagAndProbe

Hypernews: *Physics Analysis Tools:* hn-cms-physTools@cern.ch

Savannah: https://savannah.cern.ch/projects/physicstool/

An instantiation of tag&probe efficiency sequence

Used in inclusive W,Z cross section and ratio measurements, EWK-10/002, 005

Tag Selection

- GsfElectrons.
- Super cluster within |η| acceptance
- E_T > 20 GeV
- Isolation and Id cuts as in WP80
- Matched to lowest unprescaled electron trigger

Probe Selection

- E_T > 25 GeV, |η| in acceptance
- Fit the tag-probe invariant mass to get the number of signal events.

Obtain factorized efficiencies for passing probes:

SuperCluster → GsfElectron → WP80/WP95/EleSel → HLT

offline electron reconstruction efficiency with respect to acceptance

trigger efficiency w.r.t. offline selection

Estimated efficiencies:

ε_{REC} : SuperCluster → GSF electron

ε_{WP80} : GSF electron → electron WP80

ε_{TRG} : electron WP80 → Trigger

Methodology to apply data-driven efficiency

✦Efficiencies are determined from MC. corrected with data

$$\epsilon_X = \epsilon_{\text{MC-X}} \times \rho_{\text{eff-X}}, \quad \rho_{\text{eff-X}} = \frac{\epsilon_{\text{TNP-X}}(\text{data})}{\epsilon_{\text{TNP-X}}(\text{MC})}$$

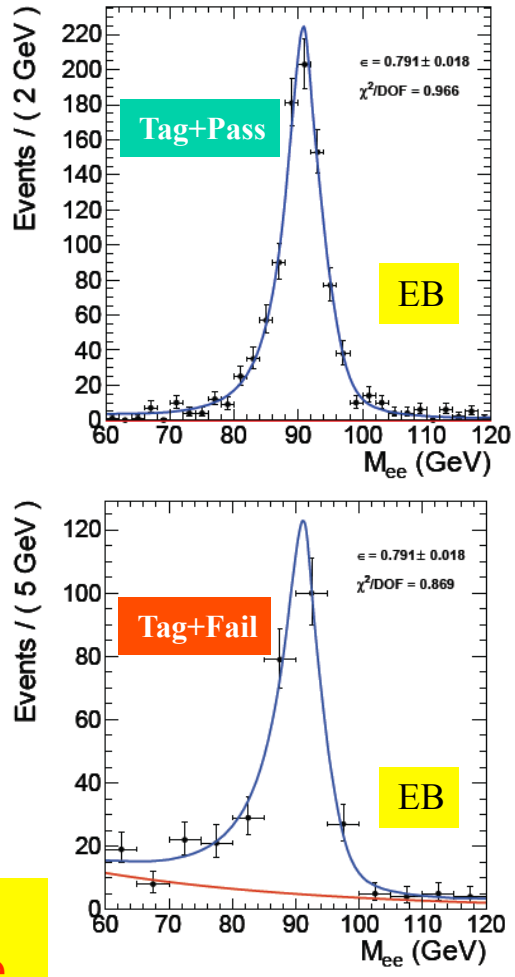
Efficiencies are determined in bins of kin. variable, e.g., E_T, η.

✦Data/MC correction factors (ρ) are estimated with Tag & Probe

- Tag : WP80 electron matched to trigger object
  Probe: passing candidate from previous stage
- M_{ee} required in window around Z mass, 60–120 GeV

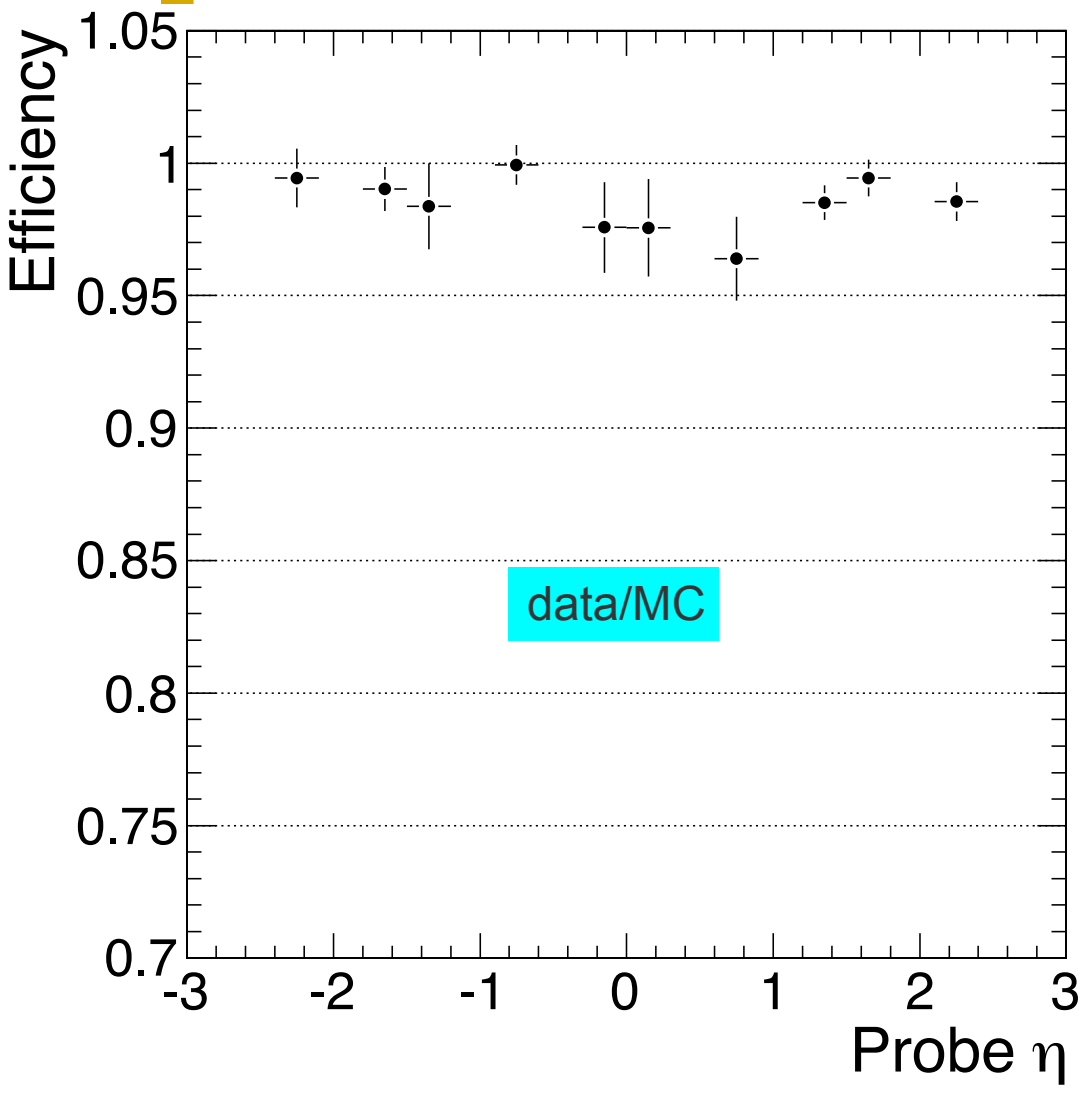
✦Simultaneous unbinned likelihood fit of M_{ee} in Tag+Pass and Tag+Fail categories:

-signal shape templates: (NLO lineshape ⊗ modified Crystal Ball) or (NLO lineshape ⊗ CMS simulation ⊗ Gaussian smearing)
- exponential or polynomial model for background



Absolute normalization of efficiency to data exposes you to kinematic biases, and should be avoided as much as possible.

Scale Factors in 2010 data: SC→Reco



Good:
- Reco scale factor is close to 1
- Is somewhat flat

Bad:
- Variation doesn't follow any regular pattern

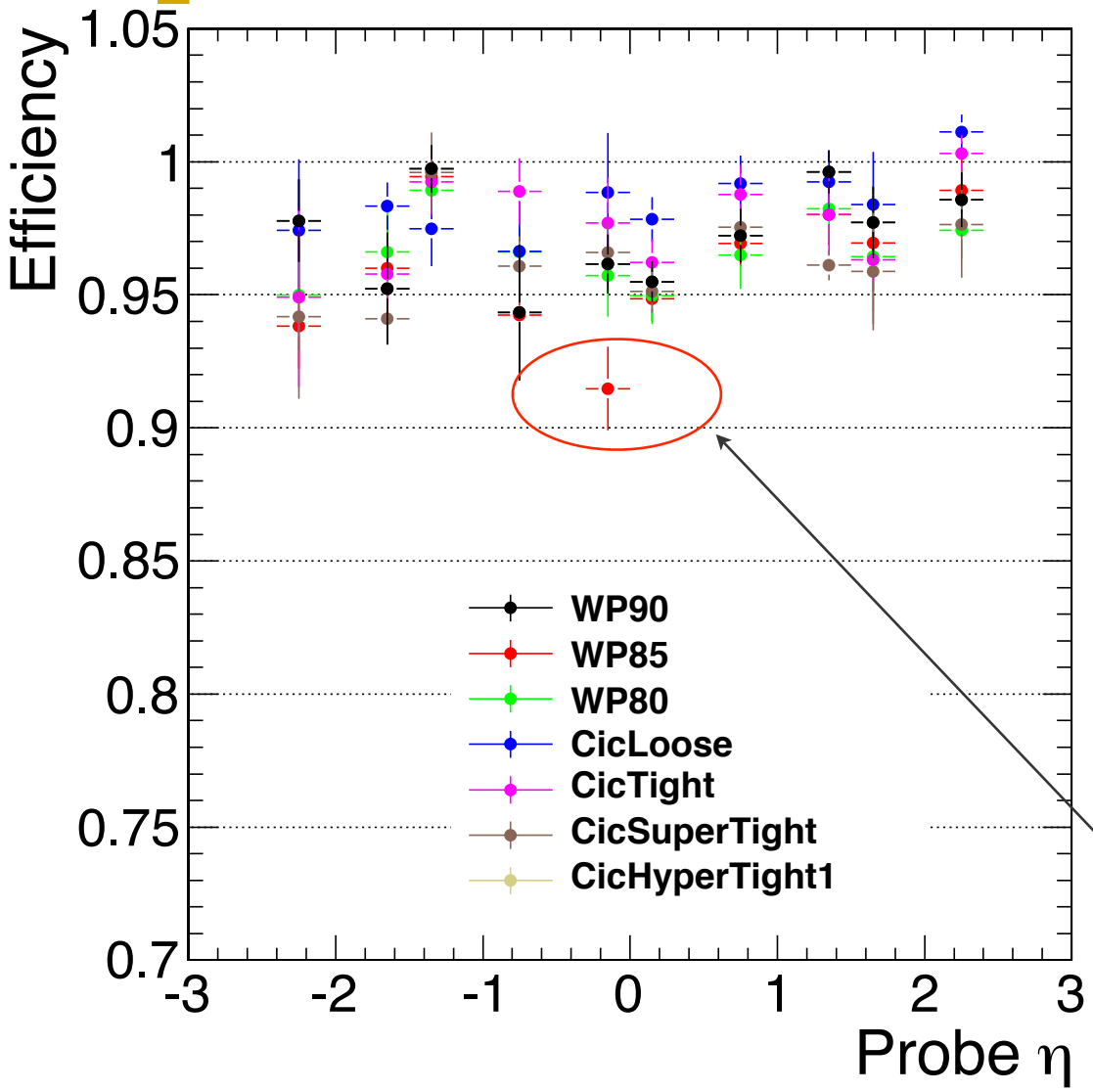
You need to sample the variation using fine granularity. Use as many bins as allowed by Z statistics.

Complication

Failing sample has large bkgd
⇒ large systematics

3.9.X data and MC

Scale Factors in 2010 data: Reco→Id

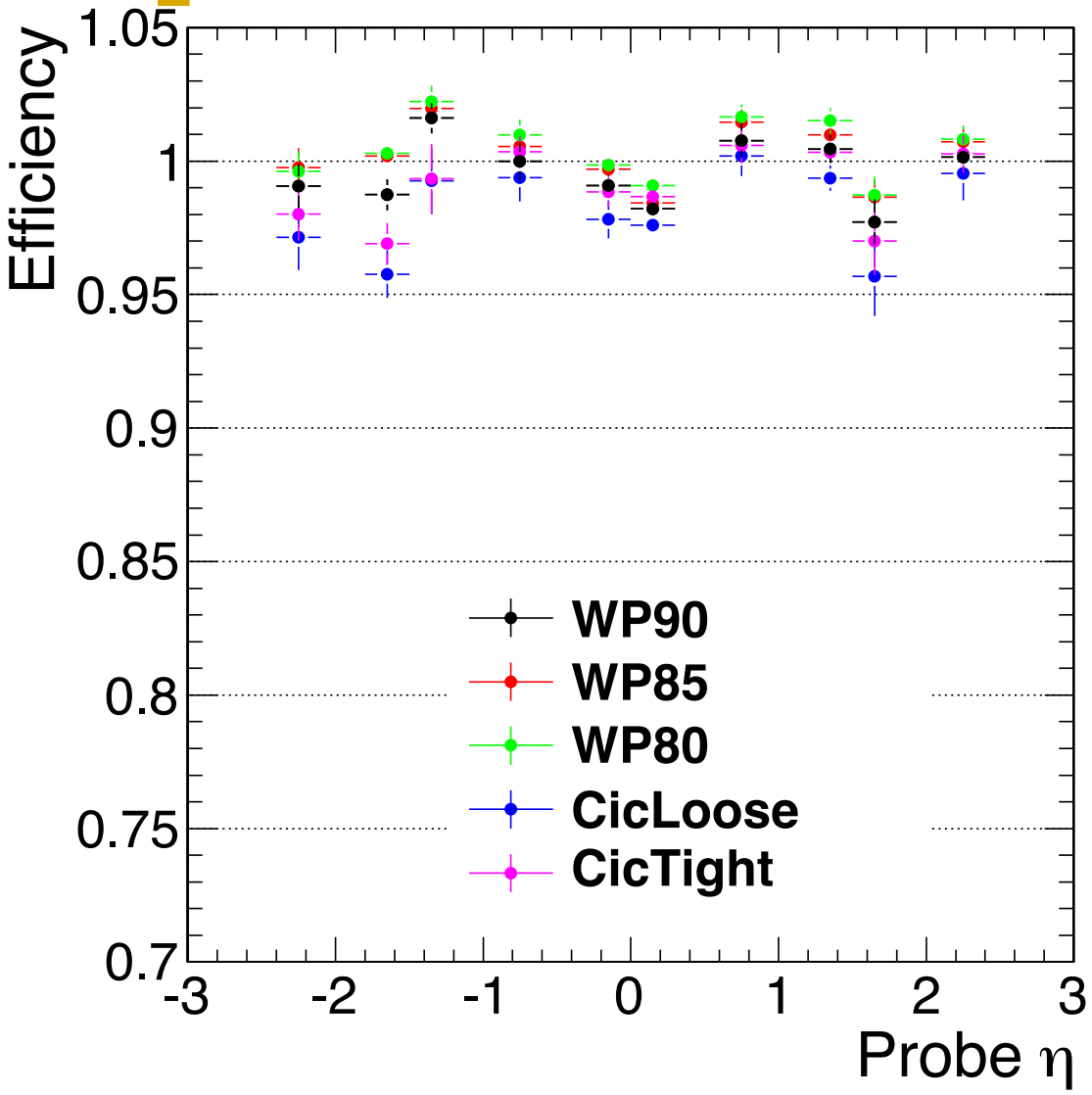


- Again, the variation doesn't follow any regular pattern

Need to sample the variation using fine granularity.

Fit failure ➜ under investigation

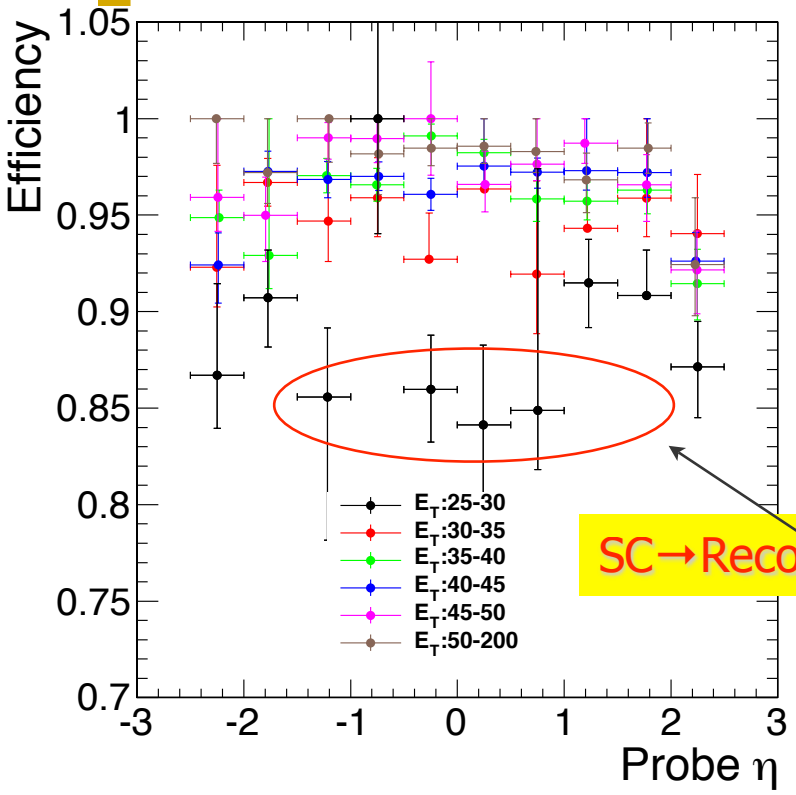
Scale Factors in 2010 data: Id→HLT



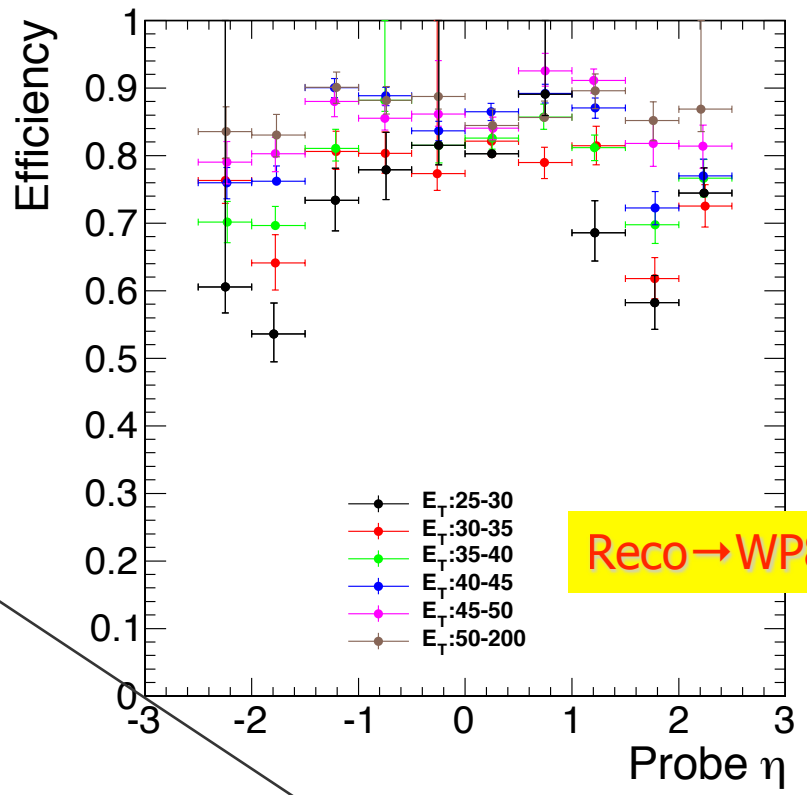
- Consistent with 1 within uncertainty

Note: The HLT path used in MC:
`"HLT_Ele17_SW_TighterEleIdIsol_L1R_v3"`
is tighter than the one in data

Detailed efficiency in data



SC→Reco



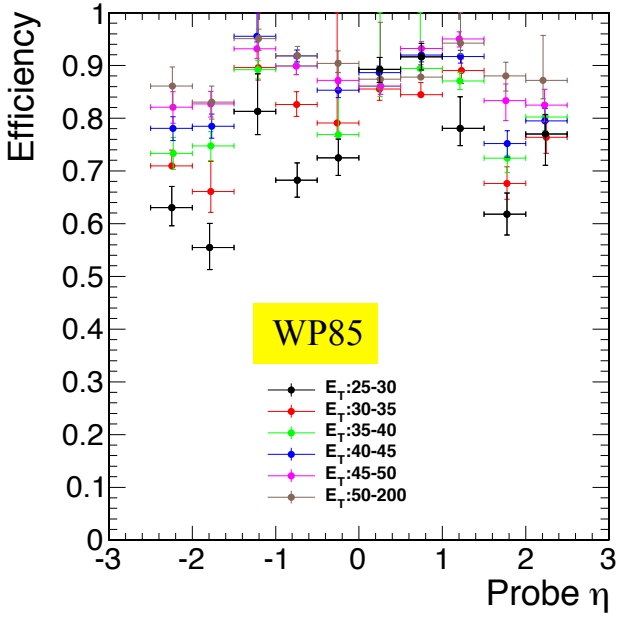
Reco→WP80

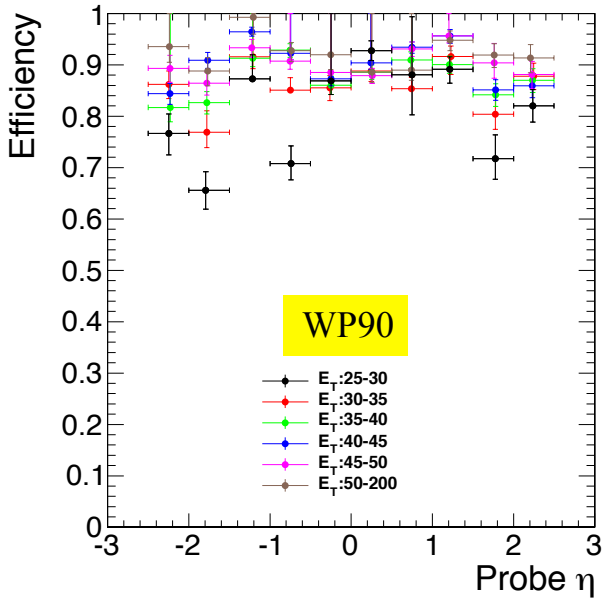
When performing thousands of fits in batch mode, a few are likely to fail. This plot shows an example. Someone needs to look into the fit projection plots, and re-do the fit in some cases.

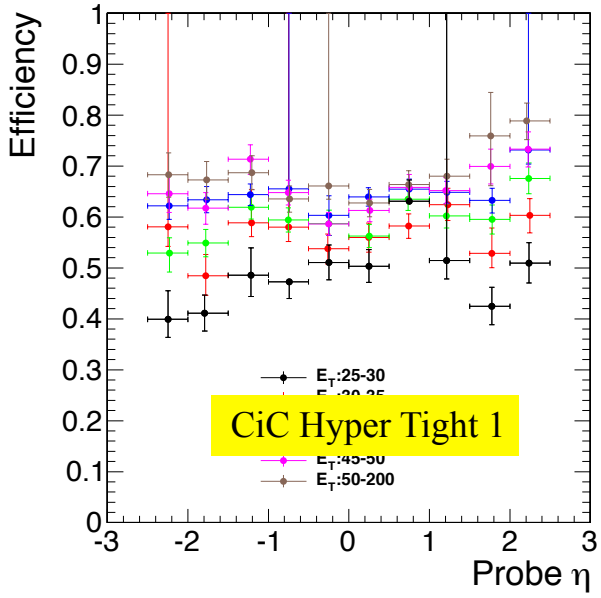
Fit failure ➜ manually repeat the fit

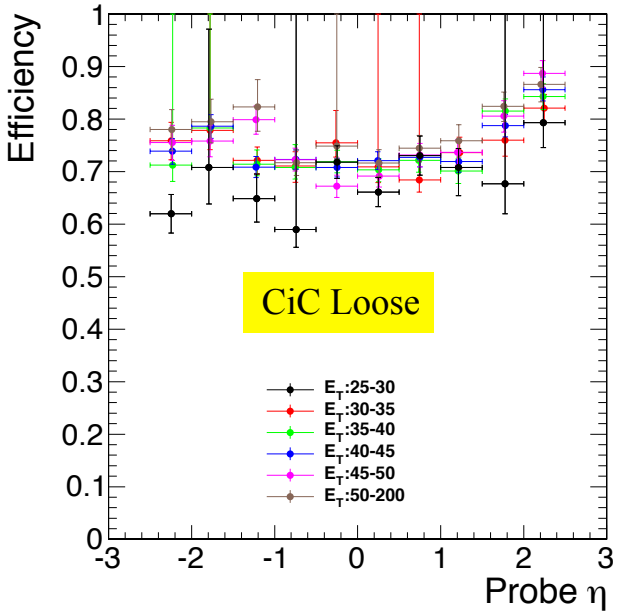
December 22 3.9.X re-Reco

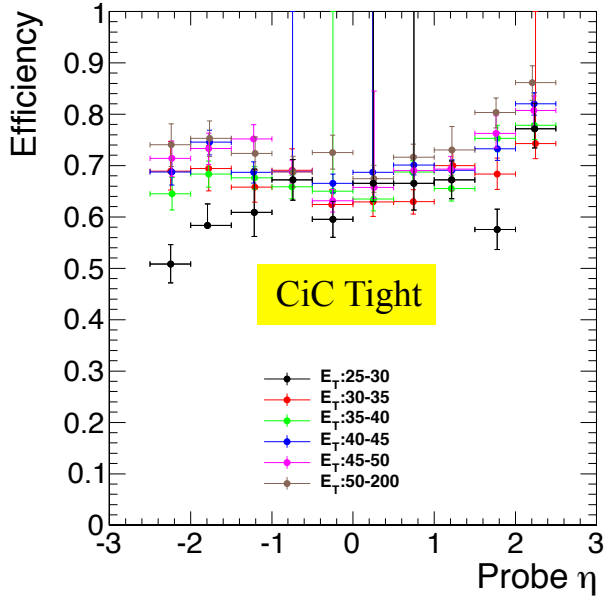
Detailed efficiency in data: Reco→Id

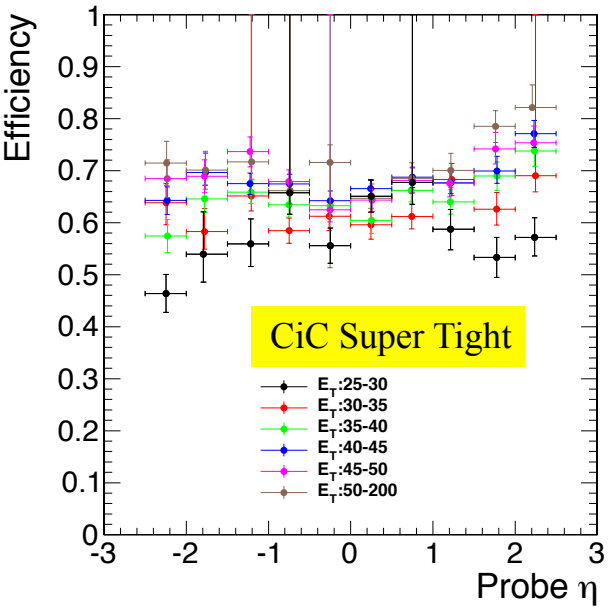




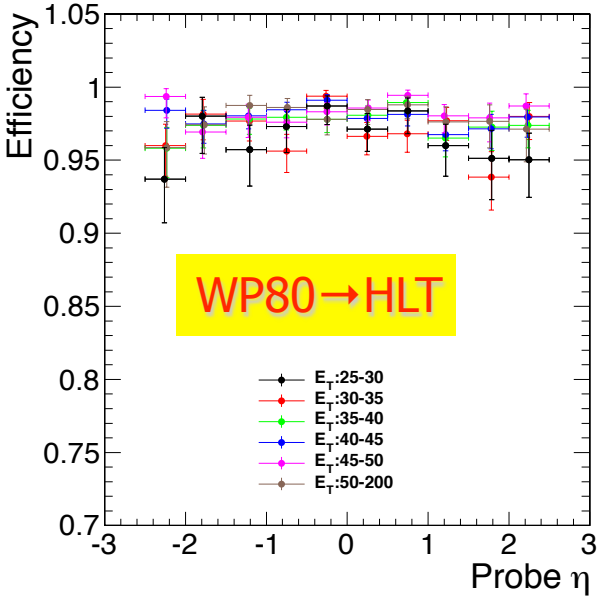


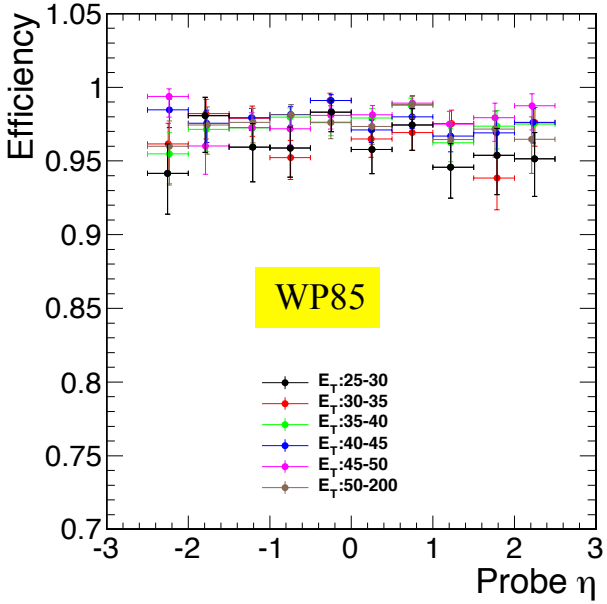


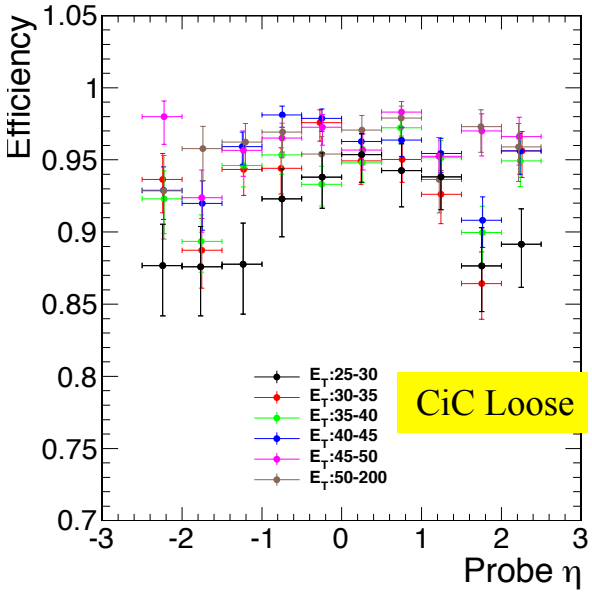


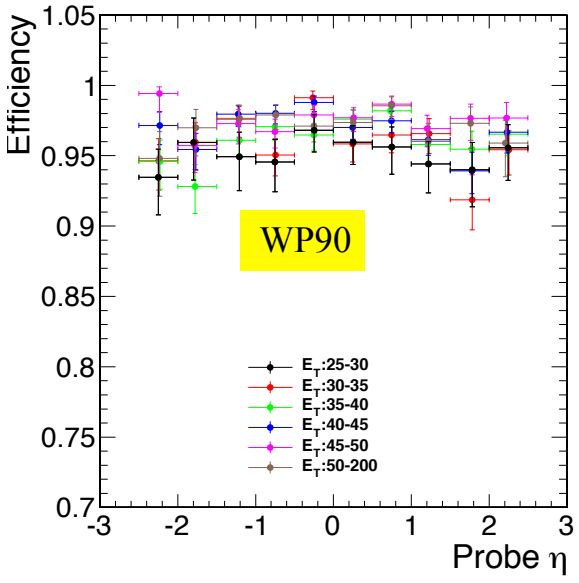


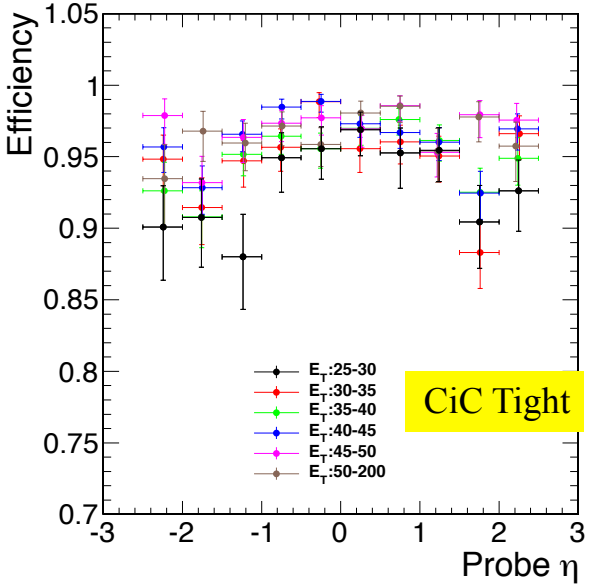
Detailed efficiency in data: Id→HLT





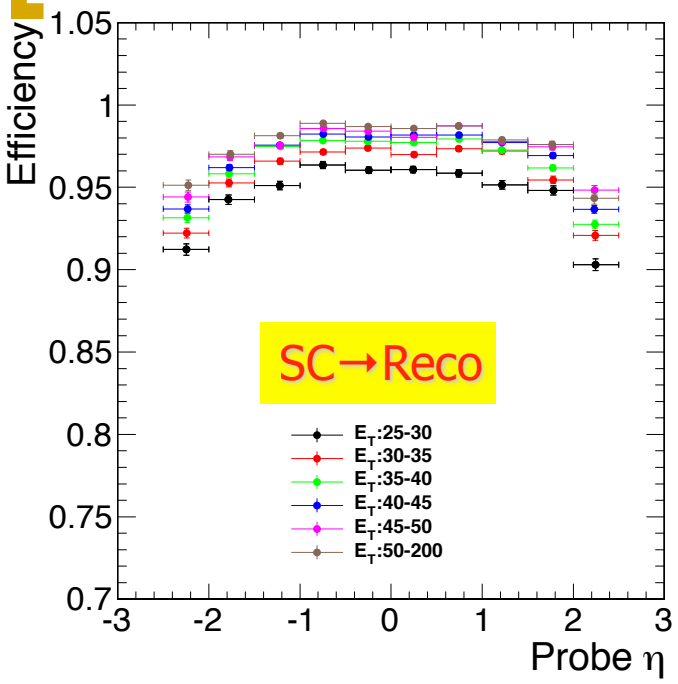


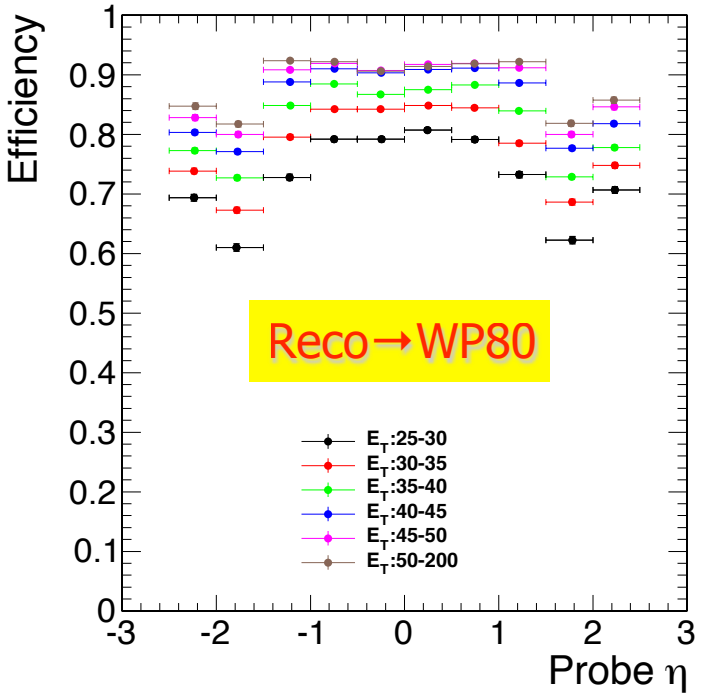


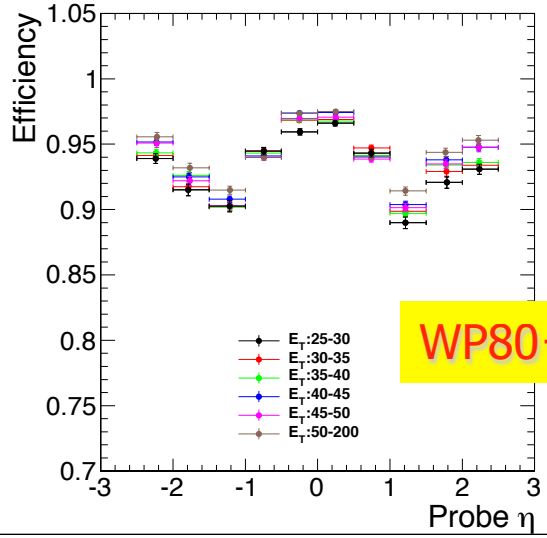


Tighter Id selection has higher HLT efficiency. This is as expected.

Detailed efficiency in MC





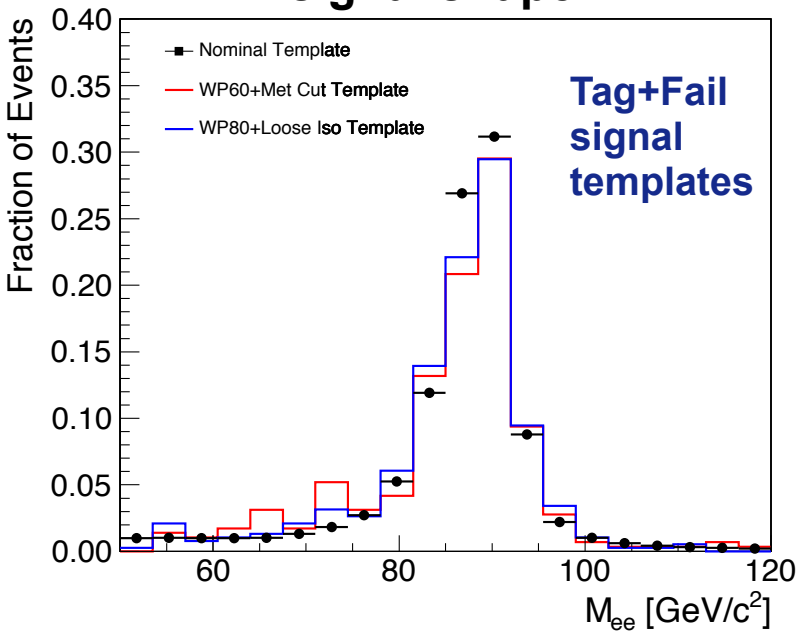


In inclusive W,Z analysis we are currently trying to understand and pin down the percent level effects in efficiency estimation.

Winter10 3.9.X MC

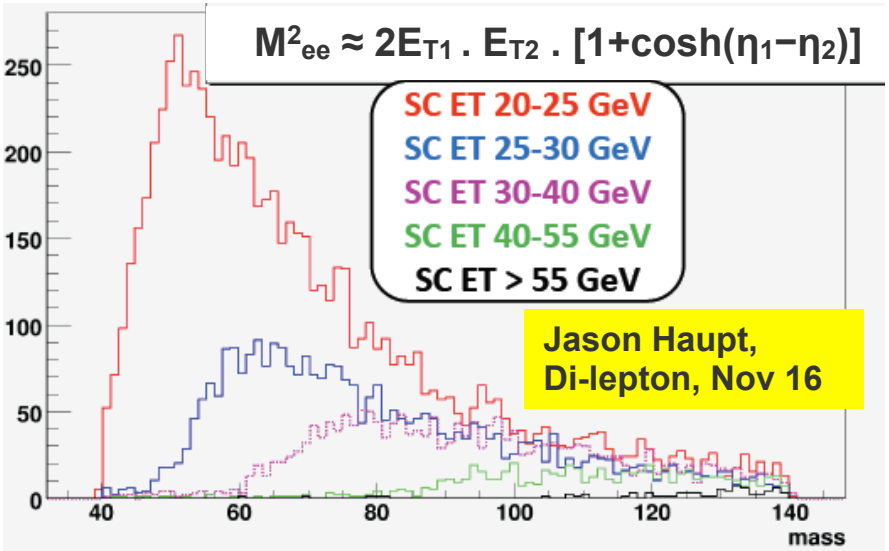
Sources of systematic uncertainty

signal shape



introduce nuisance parameters to account for the resolution, energy scale, and low mass tail

background shape



model the kinematic sculpting introduced when computing efficiency as a function of E_T using an analytic form: exponential fall at high mass and a turn over due to threshold effect at low mass. This behavior is modeled by 3 nuisance parameters.

Likelihood optimizes over all these nuisance parameters. They are allowed to float independently in each bin.

Make sure to verify the coverage !

Plans and logistics

✦*Egamma* POG − in particular this sub-group − is the appropriate forum to provide efficiency tables for most common e/γ selections being used in CMS

✦We should strive to become a one-stop shop for e/γ efficiency and deliver up-to-date ingredients for physics analysis
- -However, supporting a number of selections is a tedious task
- -Needs long term planning and commitment
- -Need to attract people from PAGs, provide them incentives to help
- -Ultimately, everyone will benefit from this effort and the productivity of the CMS collaboration will increase

✦Lot of automation tools were developed during the first EWK/onia/top analyses
- -to compute efficiency tables, plots, and scale factors in factory mode
- -plugin to get efficiency in CMSSW jobs, averaging, & MC-weighting tools
- -already in cvs: PhysicsTools/TagAndProbe

✦I can provide guidance
- -being a developer of the *TagAndProbe* package can help/guide maintenance
- -people from this group need to take over the regular day-to-day production and monitoring jobs