



Noise & jet Id studies in 900 GeV data

R Harris[†], C Jeong^{*}, K Mishra[†], S Ozturk^{**}, S Sharma[†]

[†] *Fermilab*, ^{*}*Texas Tech University*, ^{**}*Cukurova University*

Jet Algorithm Meeting, January 14, 2010

Event selection



◆ Basic cut

- BX * BSC triggers * physics declared * no beam halo * a primary vertex
- $|PVz| < 20$ cm
- both jets corrected $p_T > 10$ GeV, and $|\eta| < 2.6$
- $|\Delta\phi - \pi| < 0.5, 1, 2$ (Try three different purities) → For dijet events only
- N90hits > 1

◆ Cut to check

- fHPD < 0.99 & fRBX < 0.99
- Emf > 0.01
- nTrkCalo > 0

◆ Two jets with highest corrected p_T are studied

Some definitions for this presentation



Definitions

- **Cut efficiency in data** = #events passing the cut / #events
- **Cut efficiency in MC** = #MC events passing the cut / #MC reco'd events
- **Rejection rate** = efficiency in MC – efficiency in data

Remarks

- ◆ Efficiencies defined this way are observed efficiency of the cuts in the sample.
- ◆ When the data efficiency is a lot less than the MC efficiency then we are potentially rejecting noise (*i.e.*, things that are not present in the Monte Carlo simulation of the 900 GeV MinimumBias events).

HCAL "noise" removal cut efficiency & rejection power



◆ Fraction of jet energy in Ecal cut

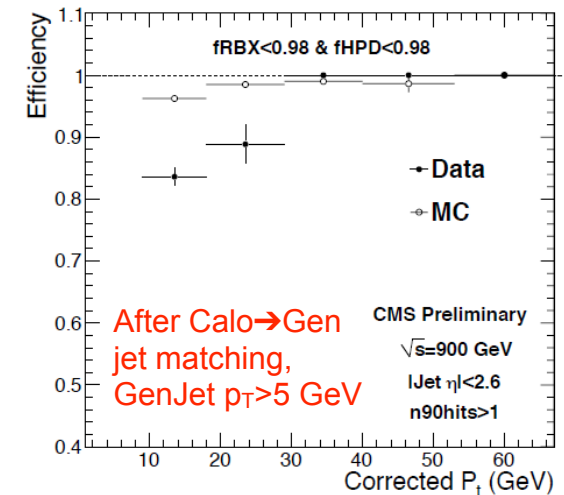
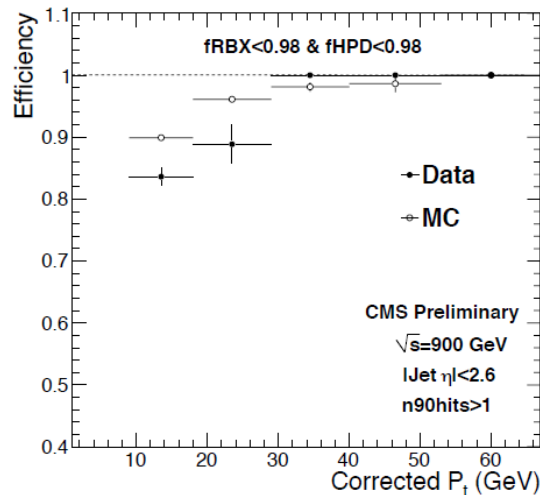
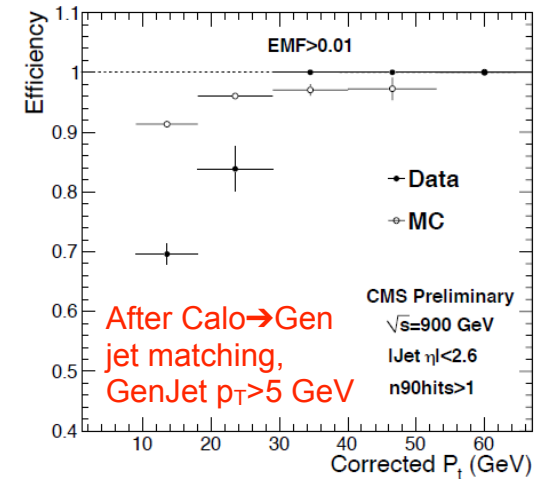
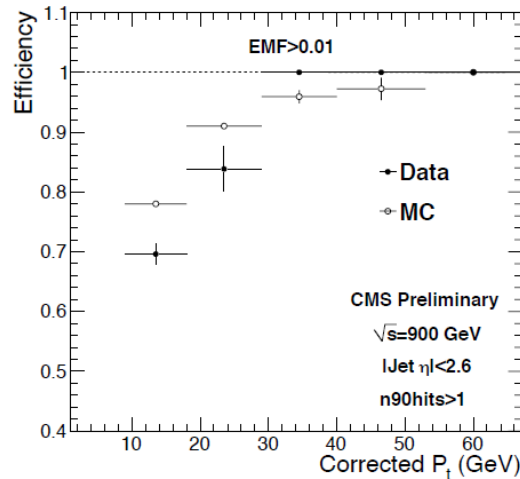
- EMF > 0.01 cut is removing a sizable fraction of events from both data and MC.
- But is not fully efficient at low p_T → it removes some real/good events

◆ Fraction of jet energy in single RBX or HPD cut

- fRBX & fHPD < 0.98 cut removes less from data and also has higher MC efficiency.

→ more on this later

Dijet sample

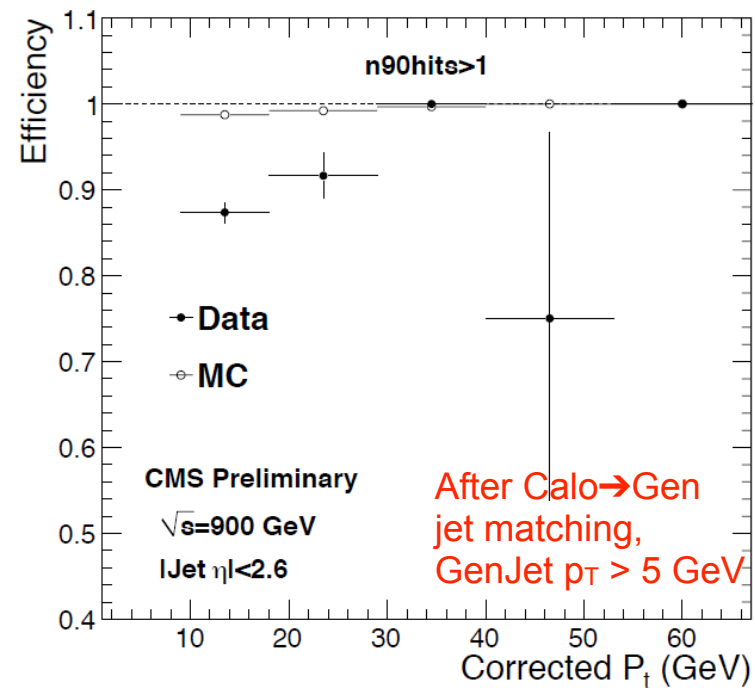
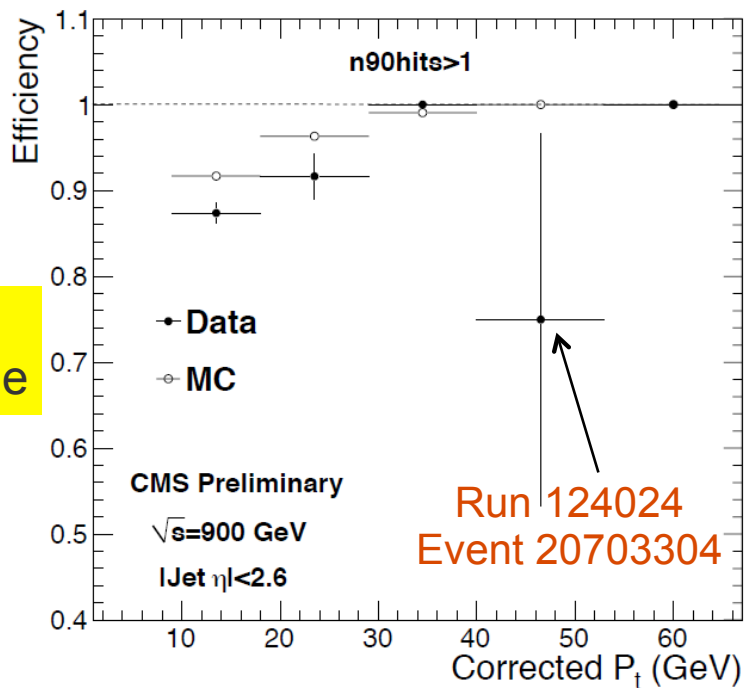


ECAL "noise" removal cut efficiency & rejection power



- ◆ **N90hits > 1 cut rejects Ecal spikes in data and has high MC efficiency**
 - ✓ Very close to 1 after requiring CaloJet matches a GenJet within $\Delta R=0.5$
 - ✓ But GenJets have $p_T > 5$ GeV \rightarrow large # unmatched CaloJets fluctuating high
- ◆ **However, it does have some inefficiency \rightarrow see next slide**

Dijet sample



N90hits cut can be improved !!!



◆ We suspect that a part of the inefficiency is due to single HCAL rechits at low p_T

- ✓ Inefficiency may be due to removal of jets made from single HCAL RecHit
- ✓ Some of these jets are real

◆ The R9 cut discussed by MET is perhaps a better cut, but is at lower level

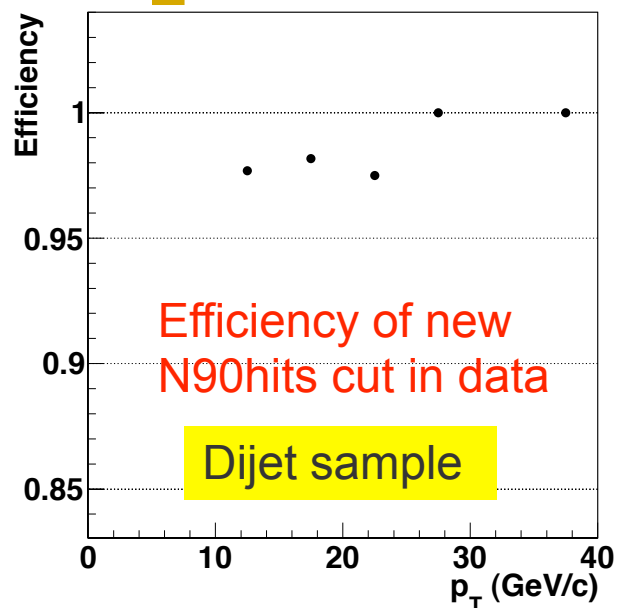
- ✓ we aim to find an appropriate cut at jet level

◆ We can improve the N90hits cut

- ✓ by combining with EMF information, e.g., if N90hits == 1 then reject only those events which have $EMF > 0.95$ → **see next slides**
- ✓ by combining with fHPD & fRBX information, we may be able to reduce HCAL noise and simultaneously improve N90hits efficiency → **need to investigate**



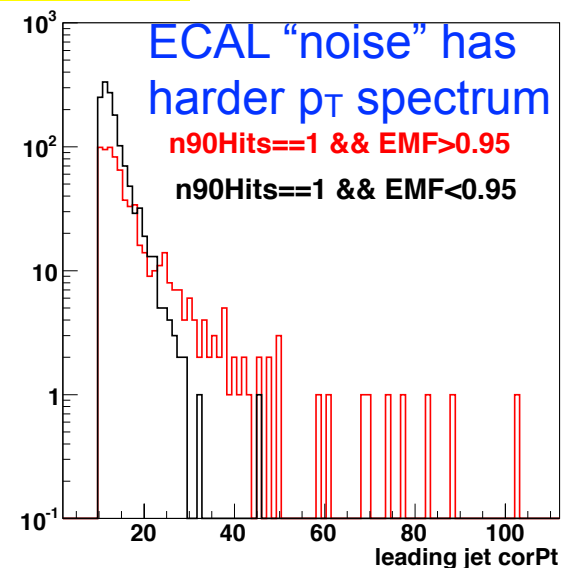
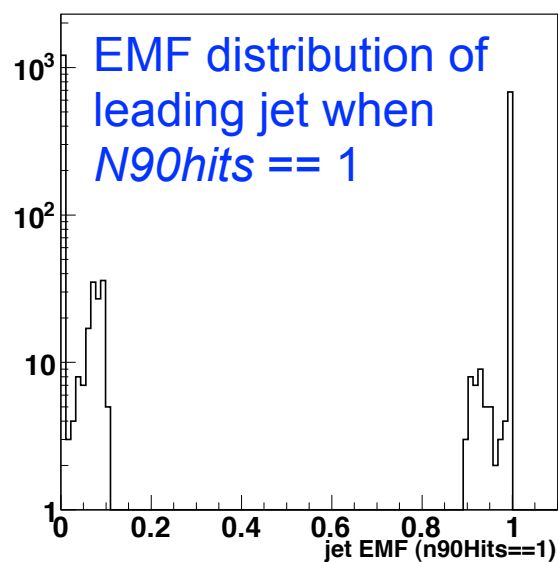
Improved N90hits cut



Efficiency for new n90hits cut

$N90hits > 1$ OR NOT($N90hits == 1$ AND $EMF > 0.95$)
in data is very close to 1 and is almost identical to the efficiency in MC !

Inclusive jet sample

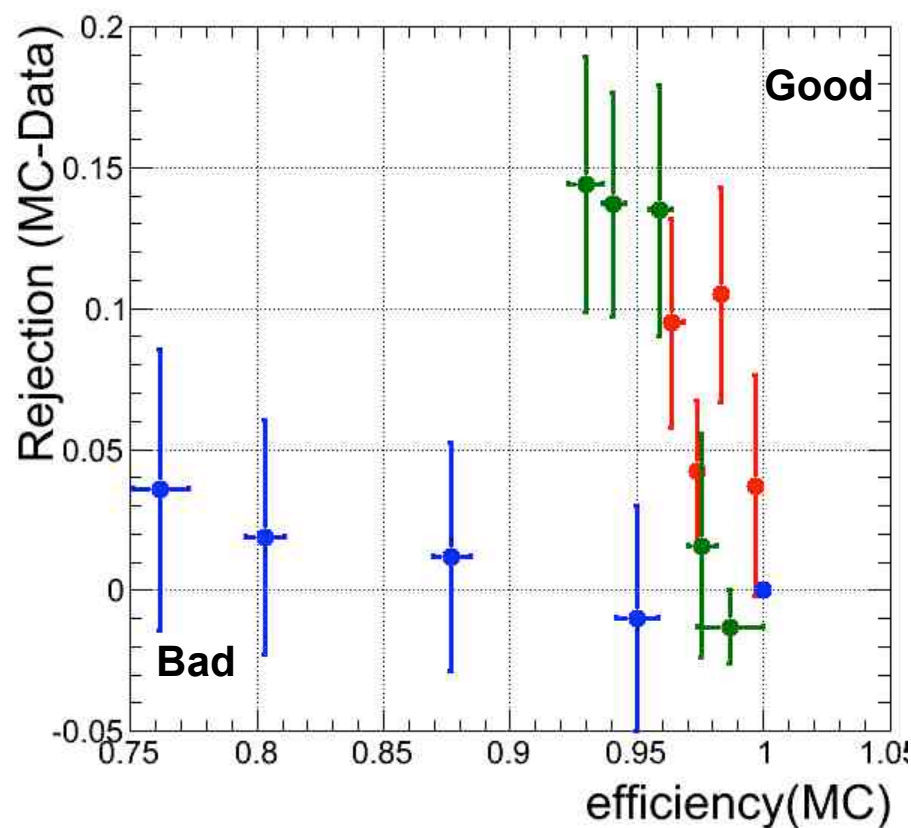
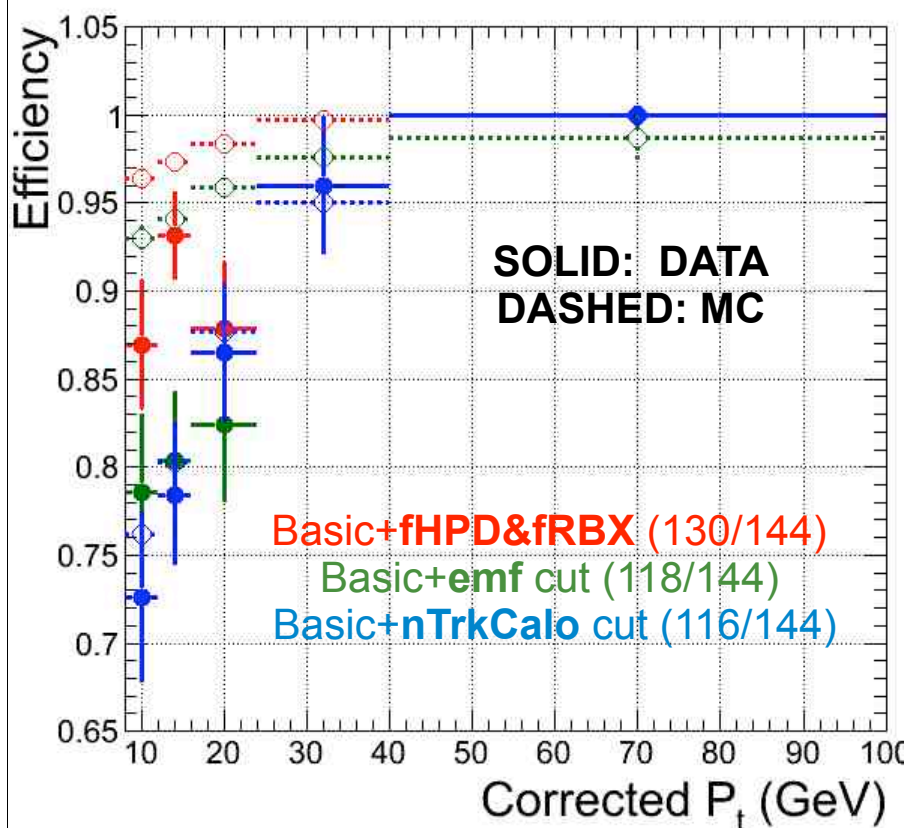




Cut efficiency in high purity sample: $|\Delta\phi - \pi| < 0.5$

Dijet
sample

Tracking cut doesn't have enough rejection power and has poor efficiency (\rightarrow presence of neutral hadron jets)

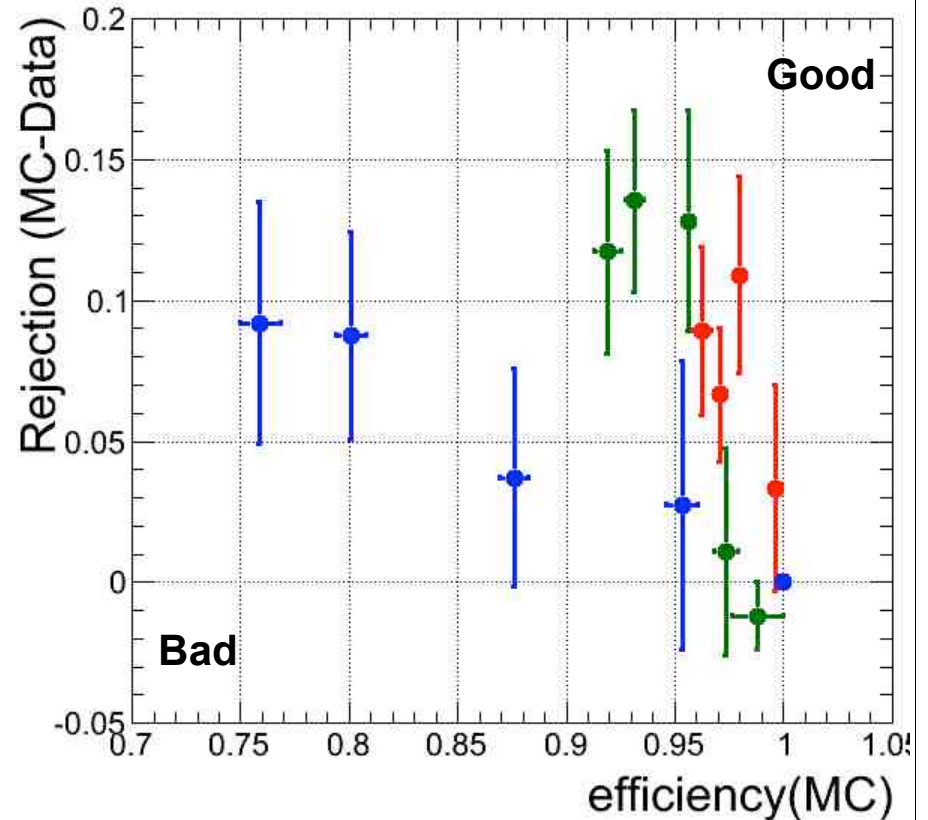
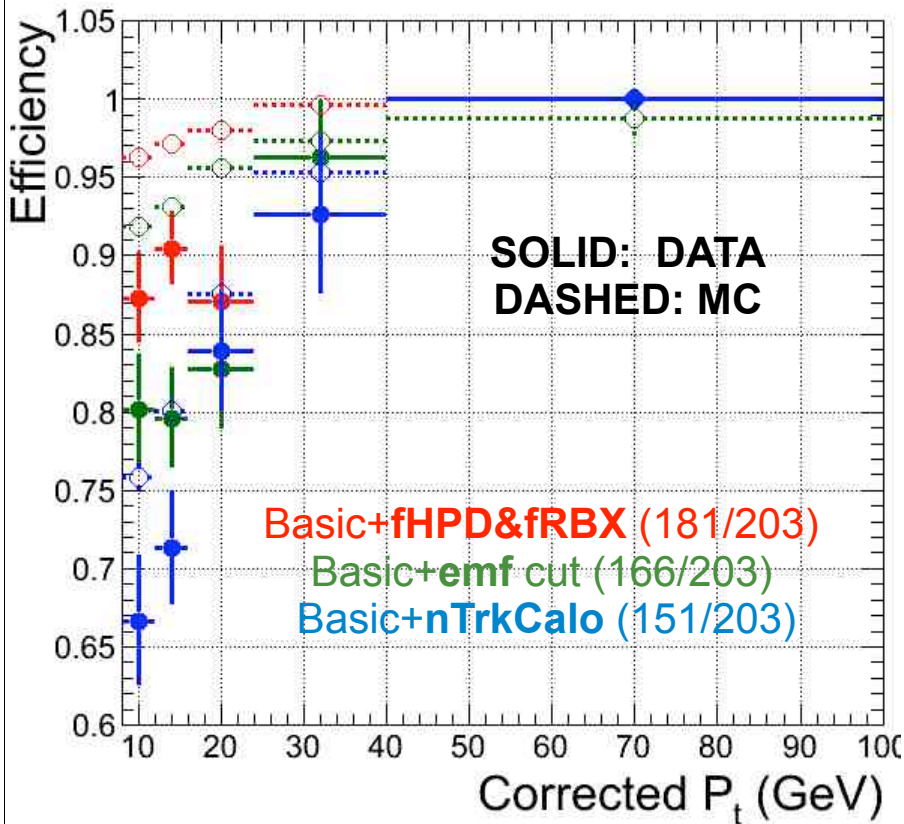




Cut efficiency in medium purity sample: $|\Delta\phi - \pi| < 1$

Basic and fHPD&fRBX cut rejection rate goes upto 10% for 98% efficiency. Basic and EMF cut can reject upto 13% for 95% efficiency.

Dijet sample

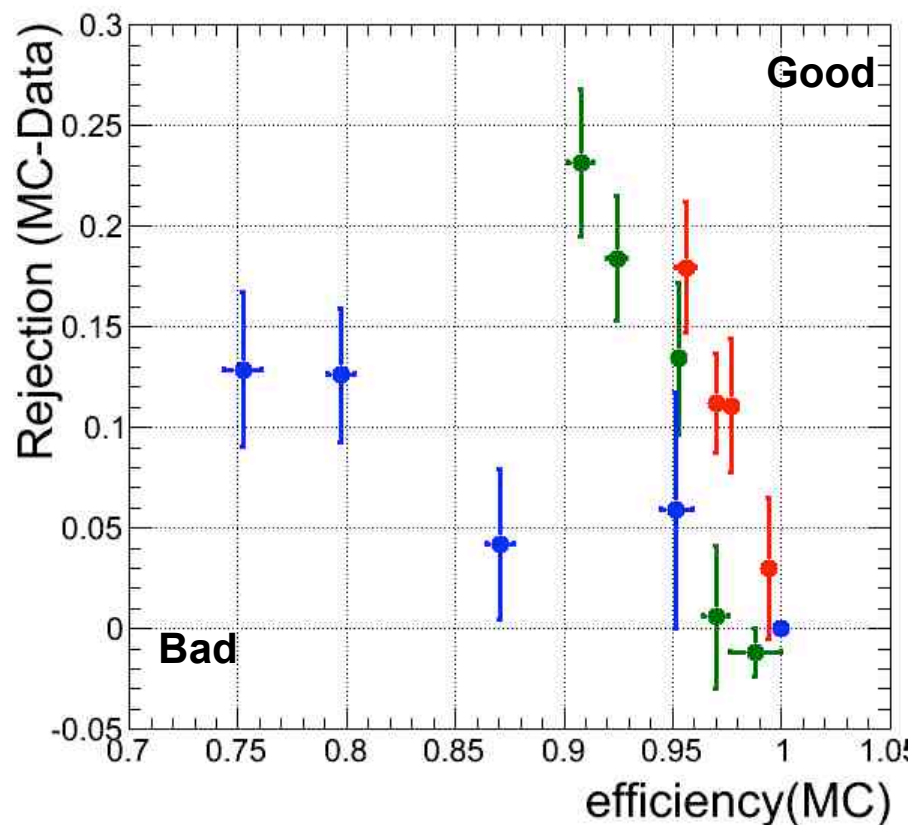
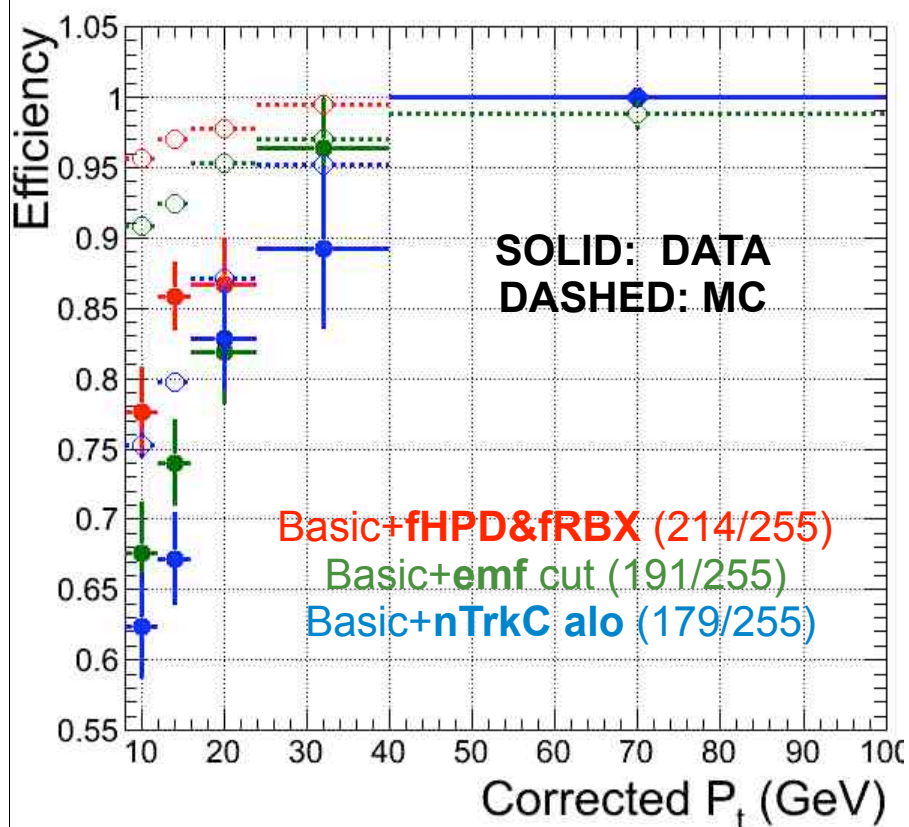




Cut efficiency in low purity sample: $|\Delta\phi - \pi| < 2$

Basic and fHPD&fRBX cut rejection rate goes upto 18% for 95% efficiency. Basic and EMF cut can reject upto 23% for 90% efficiency.

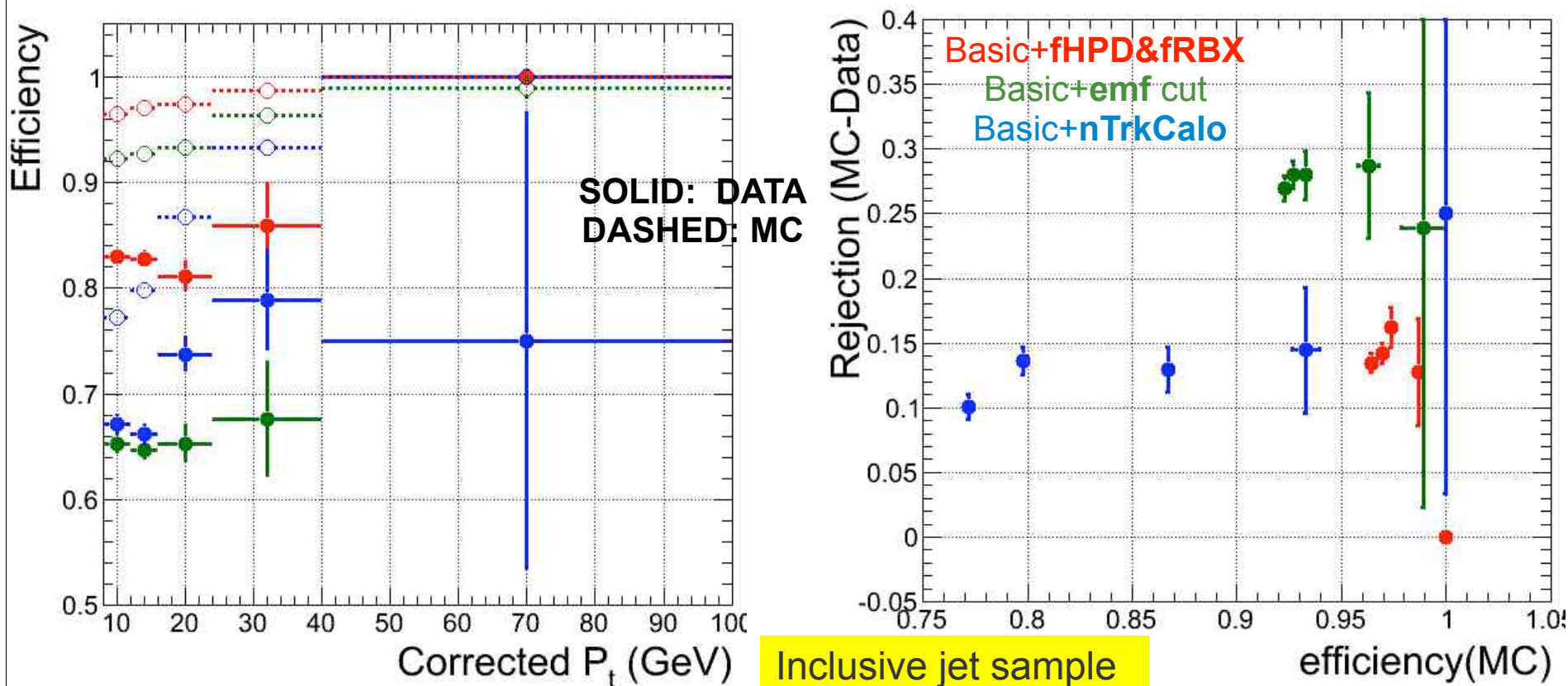
Dijet sample



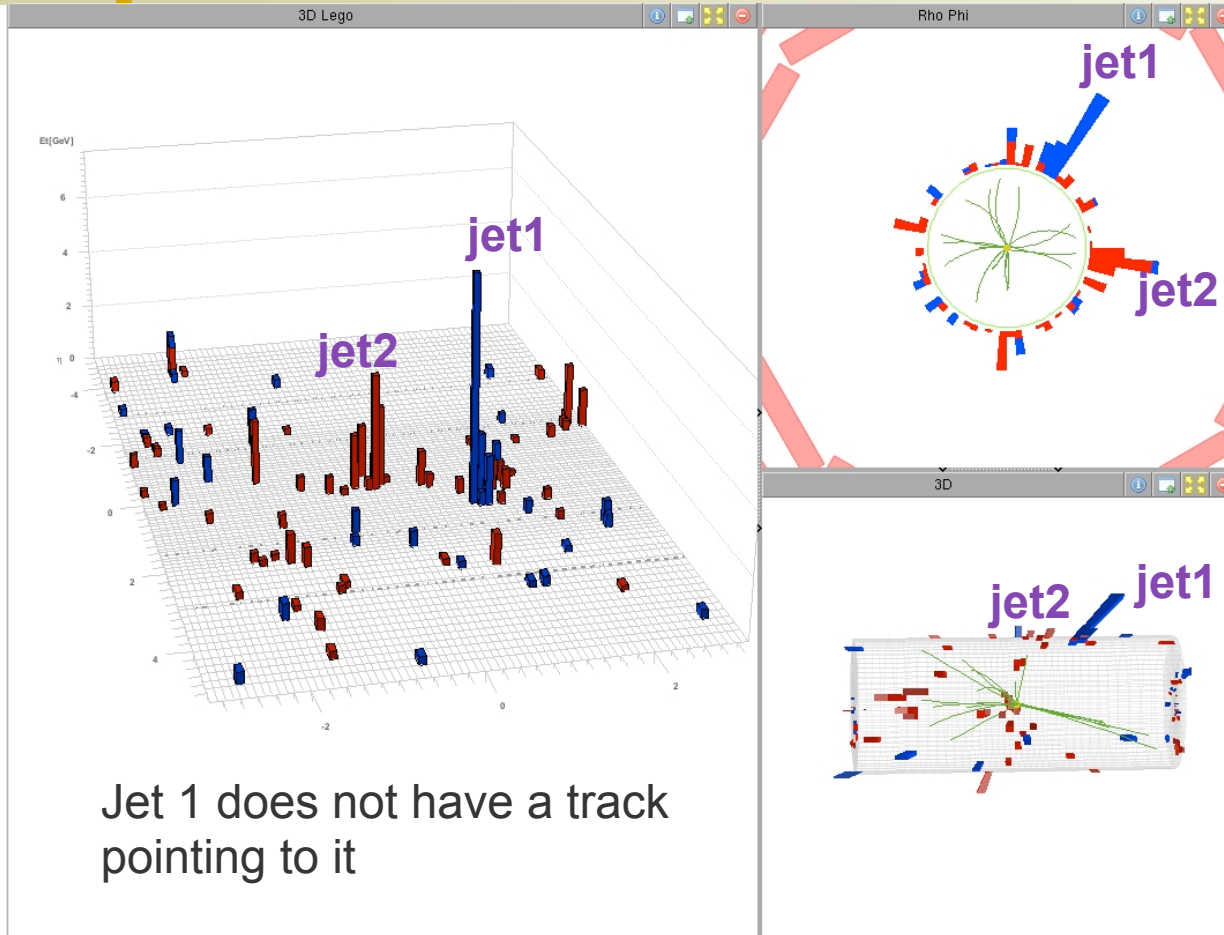


Cut performance in inclusive jet sample

Inclusive jet sample is less pure than the dijet sample. Basic and fHPD&fRBX cut rejection rate goes upto 15% for 97% efficiency. Basic and EMF cut can reject upto 28% for 95% efficiency.



Why nTrackCalo is not a good cut? (example)

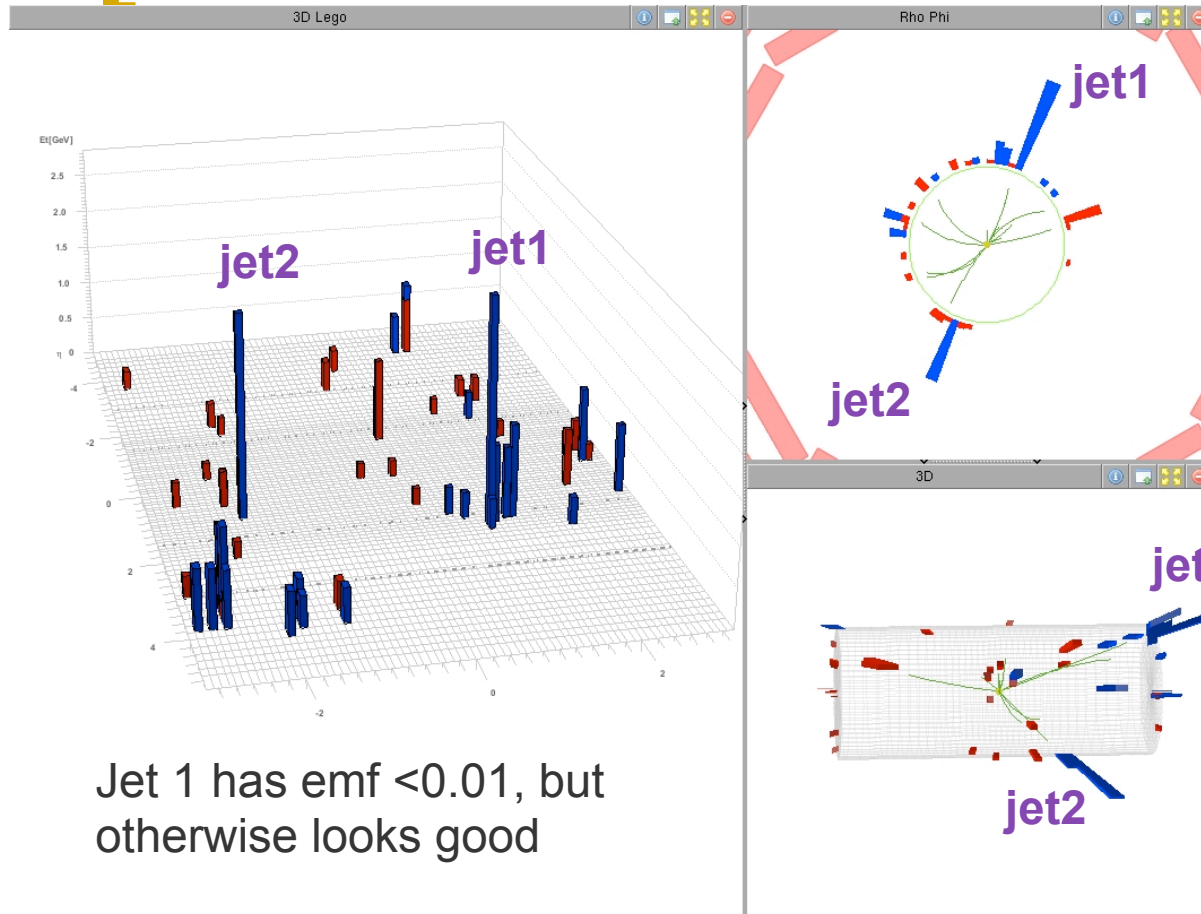


Run 124022
Event 26114551

rejected by
nTrkCalo cut

AK5CaloJets	Corr. P_T (GeV)	η	ϕ	EM Fraction	n90Hits	n90	Sigma_eta	Sigma_phi
Jet 1	30	0.8	1.1	0.14	6	5	0.138	0.117
Jet 2	24	0.0	-0.2	0.96	13	5	0.151	0.113

Why emf cut is so inefficient ? (an example)



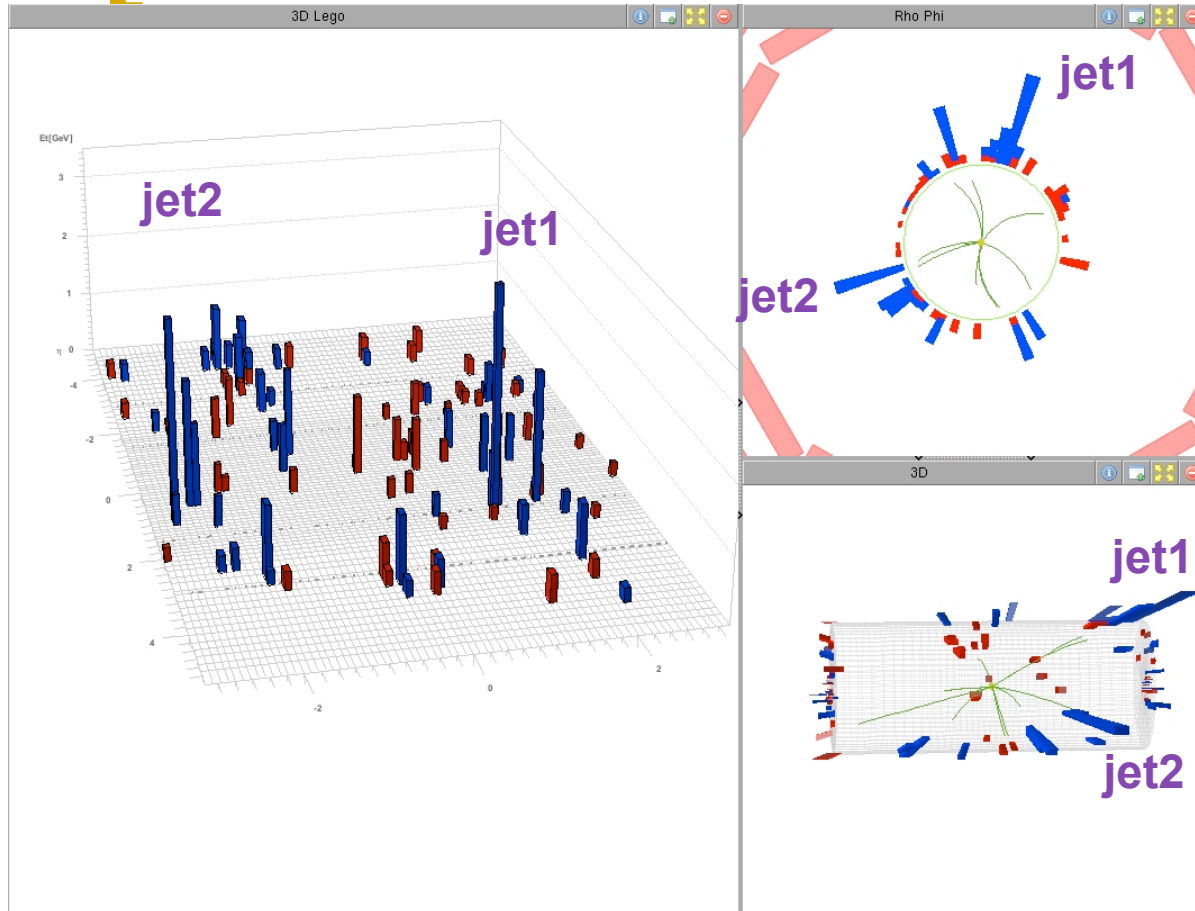
Jet 1 has emf < 0.01 , but otherwise looks good

Run 124022
Event 34026543

rejected by emf cut

AK5CaloJets	Corr. P_T (GeV)	η	Φ	EM Fraction	n90Hits	n90	Sigma_eta	Sigma_phi
Jet 1	20	1.7	1.2	0	6	5	0.096	0.144
Jet 2	11	0.8	-2.0	0.1	2	3	0.141	0.049

Cut variables can be correlated



Run 124020
Event 7328738

This noisy event can be rejected by either (fHPD & fRBX) or emf or nTrkCalo cut

AK5CaloJets	Corr. P_T (GeV)	η	ϕ	EM Fraction	n90Hits	n90	Sigma_eta	Sigma_phi
Jet 1	18	1.4	1.3	0	3	3	0.107	0.040
Jet 2	11	0.9	-2.8	0	2	2	0.030	0

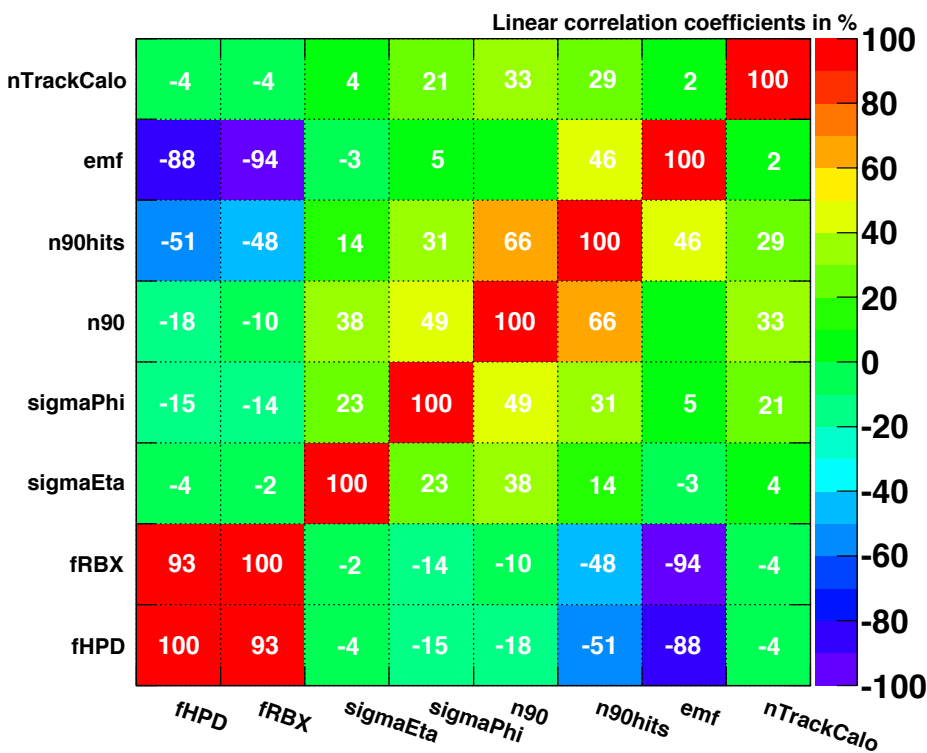
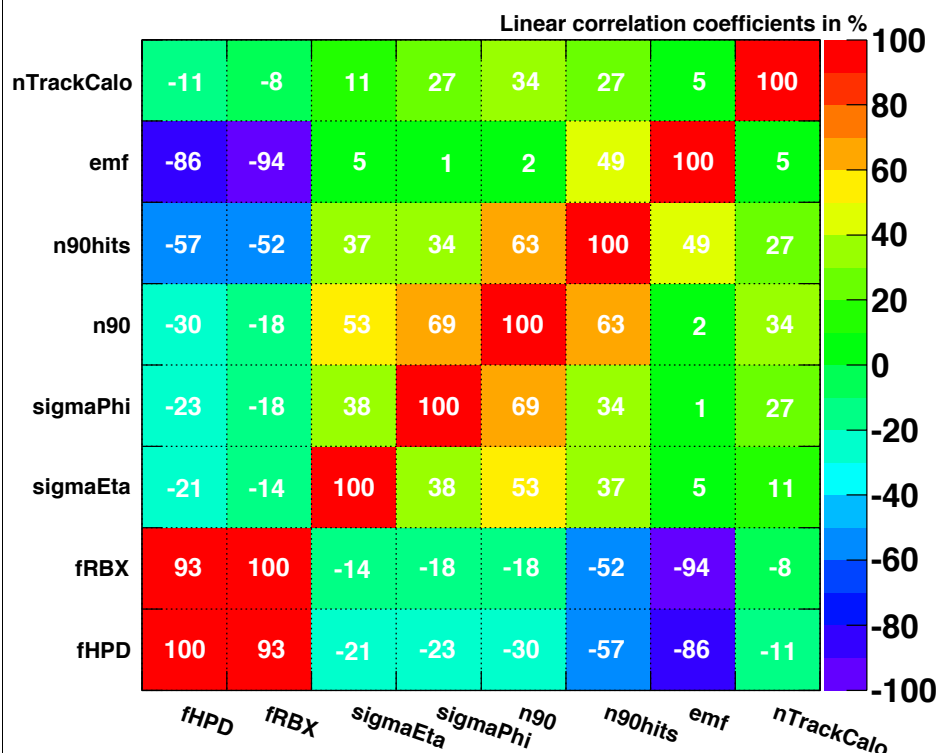
Correlation matrix for jet variables

Dijet sample



Data

MC (Calo → Gen matched)



- ◆ The correlation matrix in data and MC look similar
 - How a cut on one variable affects other variables will be similar in data and MC
- ◆ *N90hits* and *emf* are somewhat correlated
- ◆ *fRBX* and *fHPD* are very correlated

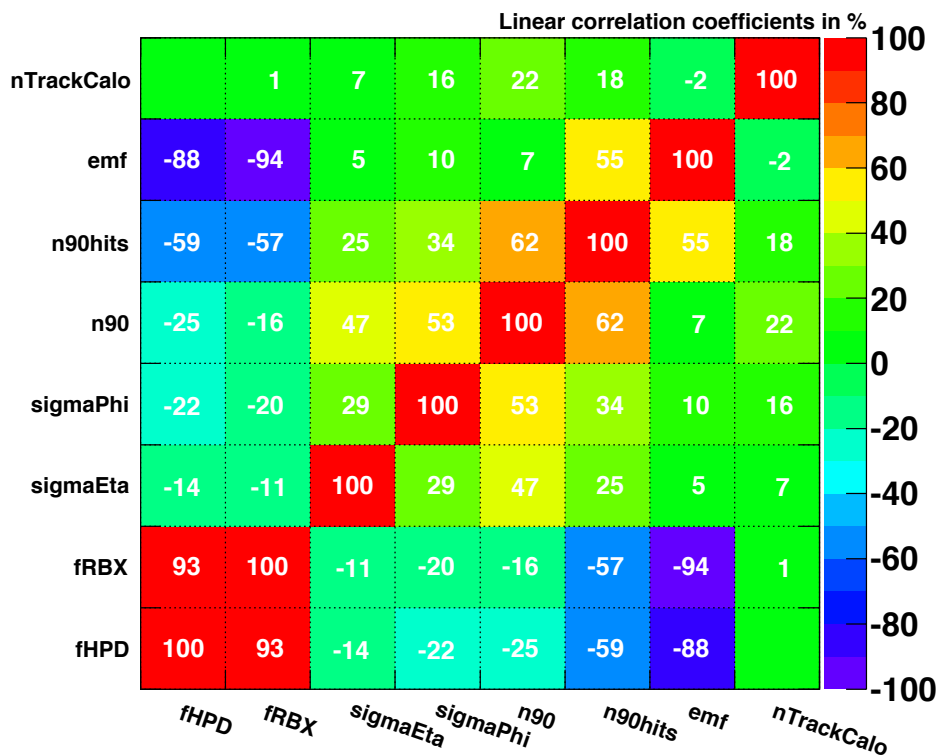
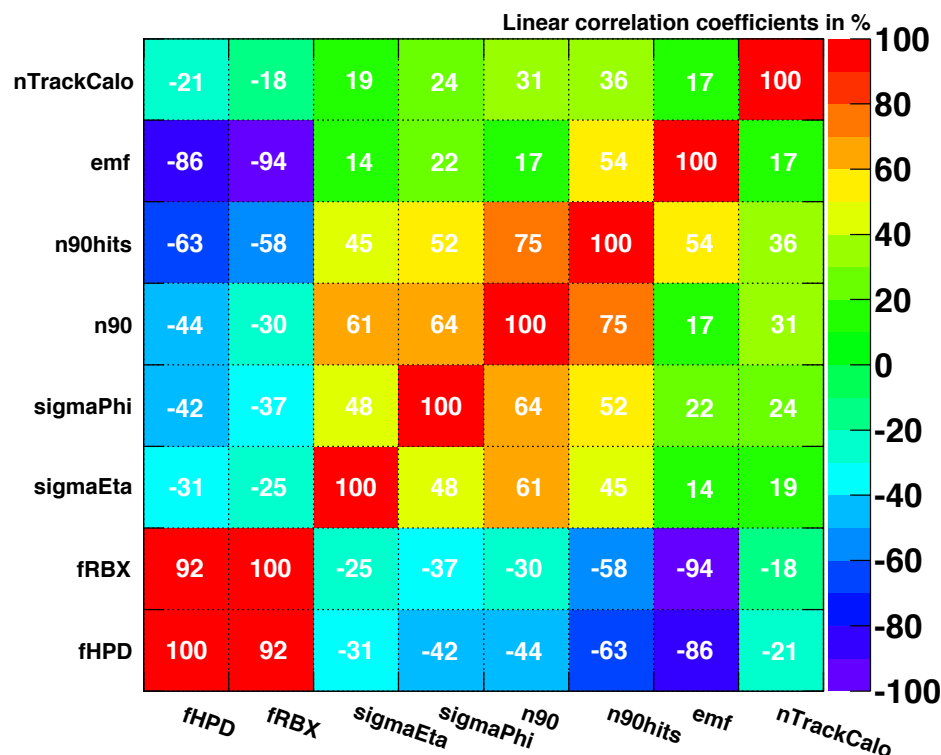
Correlation matrix for jet variables



Inclusive jet sample

Data

MC (Calo → Gen matched)



- ◆ The correlation matrix in data and MC again look similar
- but there are differences in some places
- perhaps because data is less pure

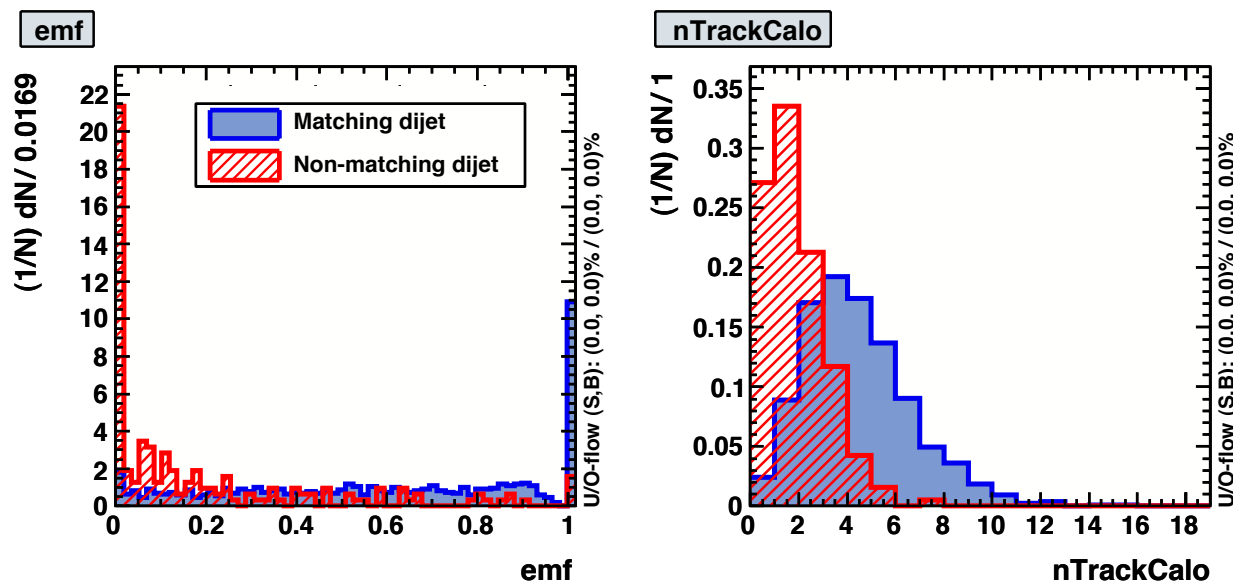


Normalized shape of jet Id variables in MC-I

Dijet sample

Two caveats

1. Monte Carlo “matched” and “unmatched” events do NOT correspond to “signal” and “noise” respectively
2. This is purely a shape analysis → relative abundance of the “matched” and “unmatched” events is not displayed in these plots, only relative shapes are shown

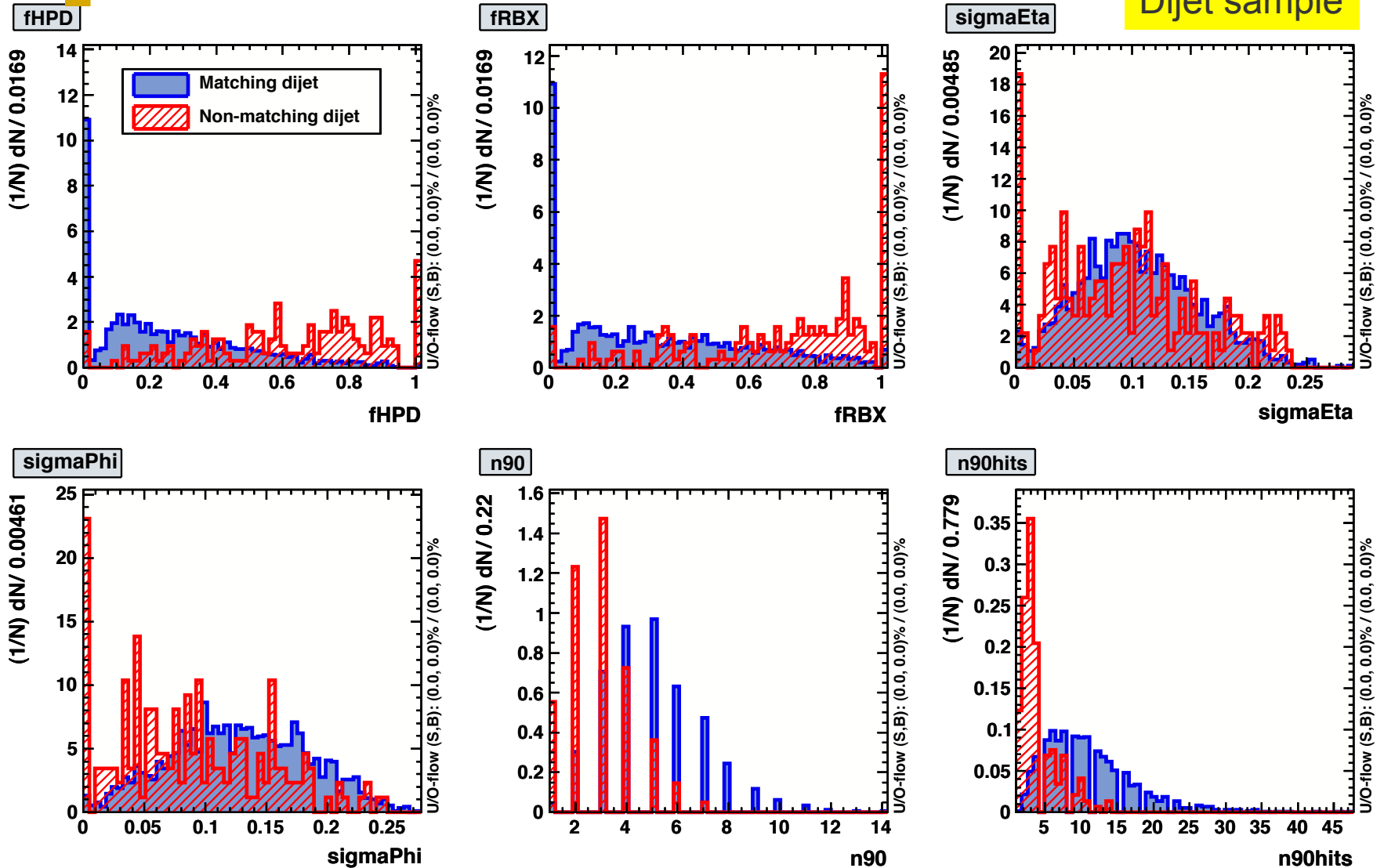


→ continued on the next slide

Normalized shape of jet Id variables in MC-II



Dijet sample

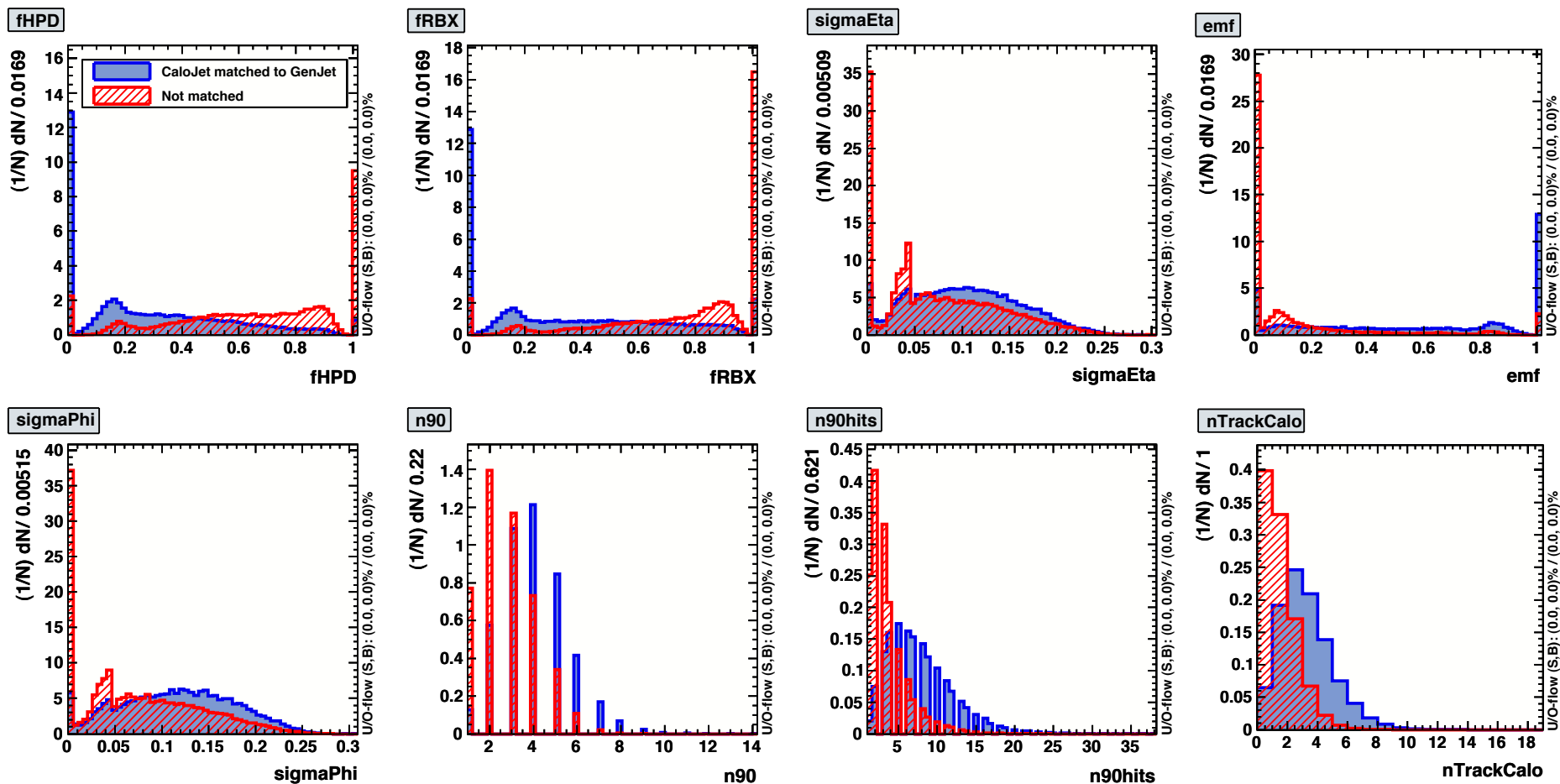




Normalized shape of jet Id variables in MC



Inclusive jet sample



Summary



We are using N90hits (>1) cut to reject ECAL spike

- this cut is also removing events with single HCAL rechits, some of these events are real and not noise
- can improve this cut
 - taking into account EMF information

Not that straightforward to reject HCAL noise

- emf, fHPD, fRBX are highly efficient, but not fully efficient at low p_T in MC
- we think that they are also rejecting good events
 - see Seema's talk
- In case of single HCAL rechit, we can use fHPD/fRBX information to discriminate noise from real events

Thank You !

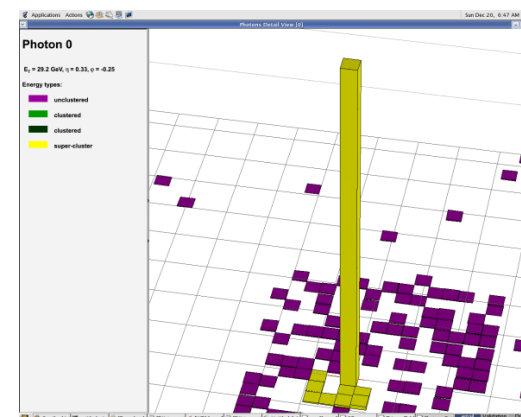
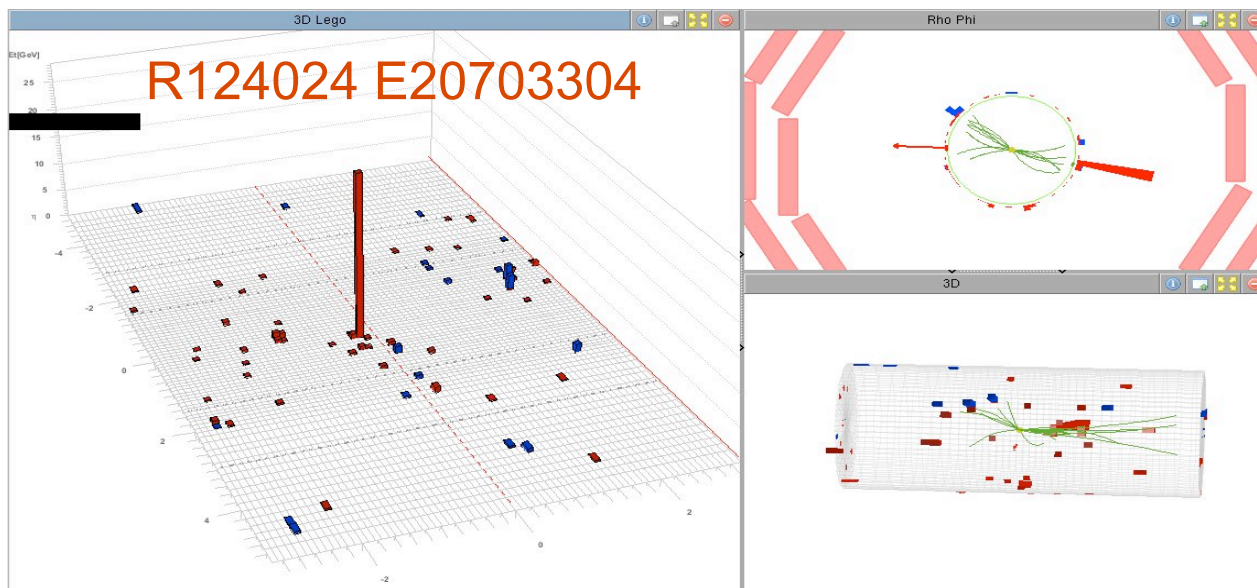
Backup slides

Rejection of ECAL spikes

◆ A cut on $N_{90hits} > 1$ effectively removes these events

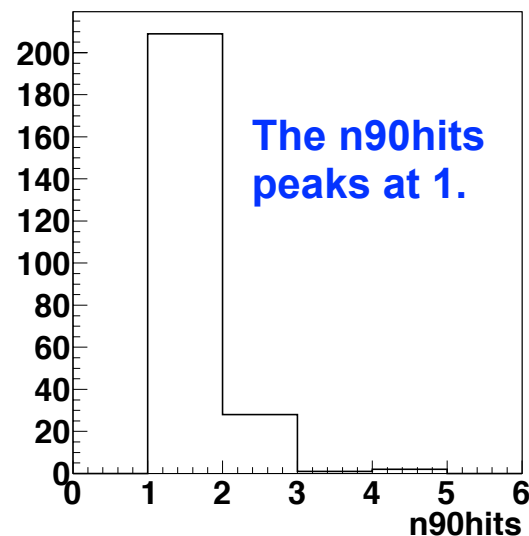
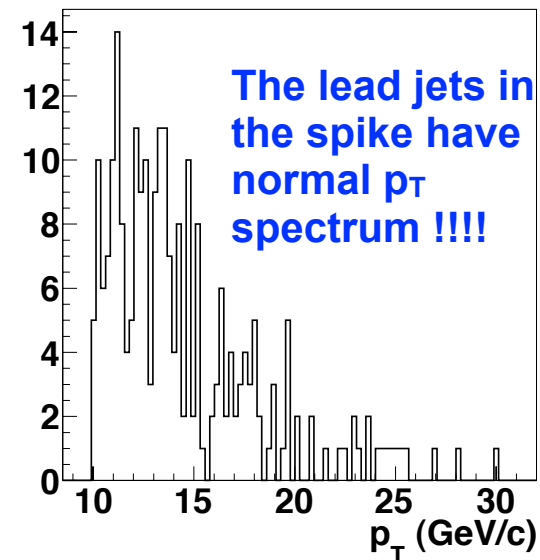
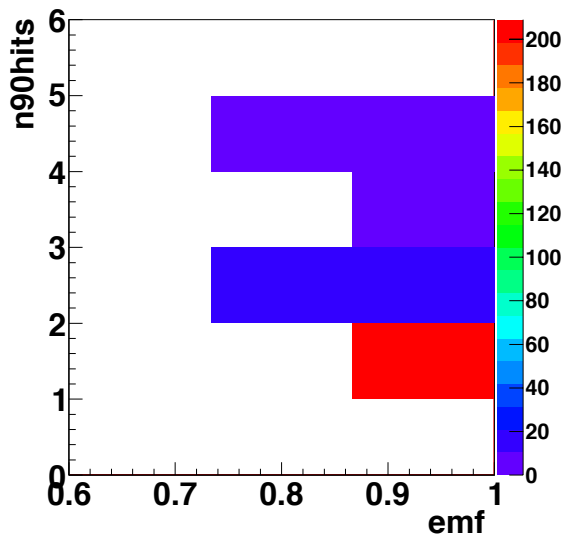
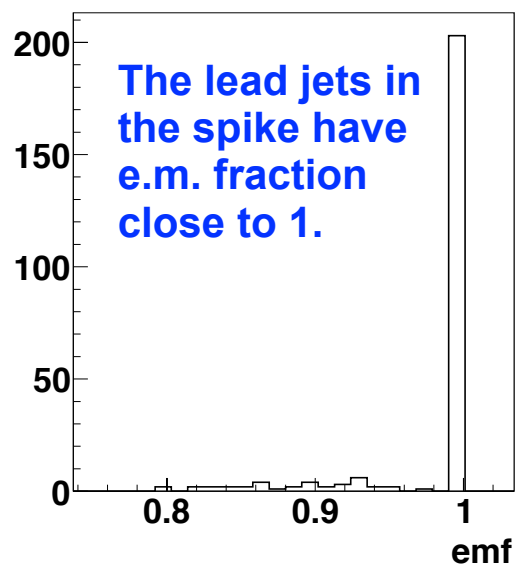
- N_{90} Hits is number of RecHits with $> 90\%$ of jet energy

Chiyong Jeong



AK5CaloJets	Corr. P_T (GeV)	η	ϕ	EM Fraction	n90Hits	n90	Sigma_eta	Sigma_phi
Jet 1	50	0.4	-0.3	1	1	1	0.076	0.032
Jet 2	19	-0.5	2.5	0.3	7	4	0.061	0.076

Further investigation of the spike (inclusive jet)



- ◆ Noise often has associated track at calo face !!!
- ◆ Can be removed by requiring $n_{90hits} > 1$ or by cutting on emf

- $n_{90hits} \equiv$ minimum number of rechits which contribute at least 90% of the jet energy

- ◆ The n_{90hits} cut has a better overall efficiency, so we currently use it in dijet analysis