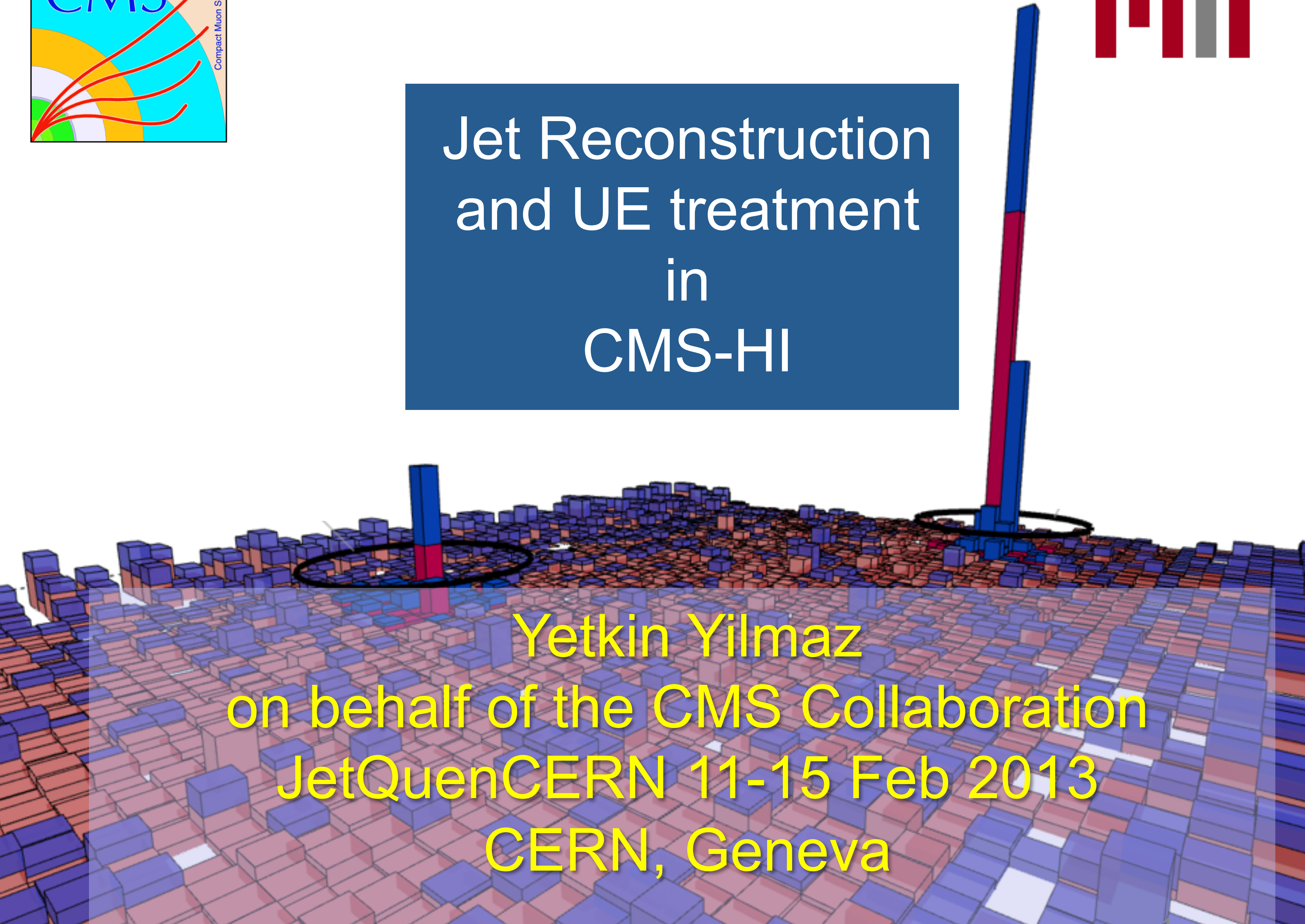
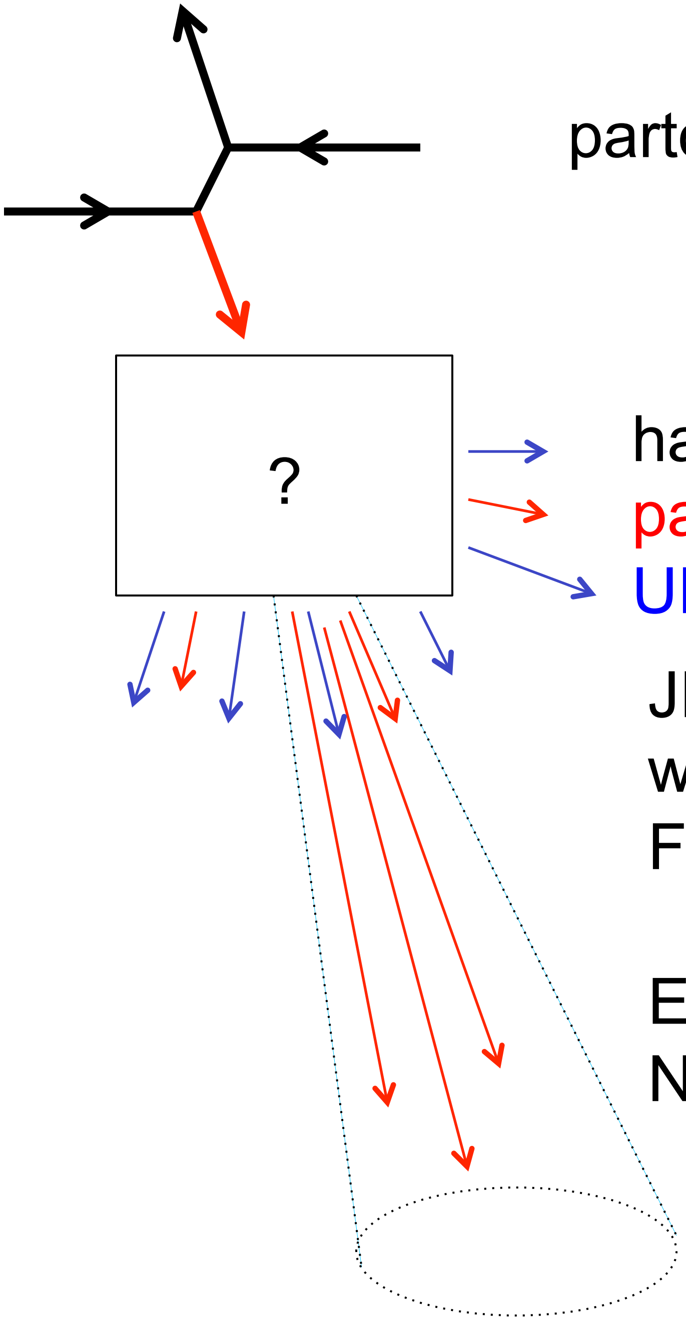


Jet Reconstruction and UE treatment in CMS-HI

Yetkin Yilmaz
on behalf of the CMS Collaboration
JetQuenCERN 11-15 Feb 2013
CERN, Geneva

Which jets?



partons+Underlying Event(UE)

hadrons:

parton-associated

UE-associated

JETS:

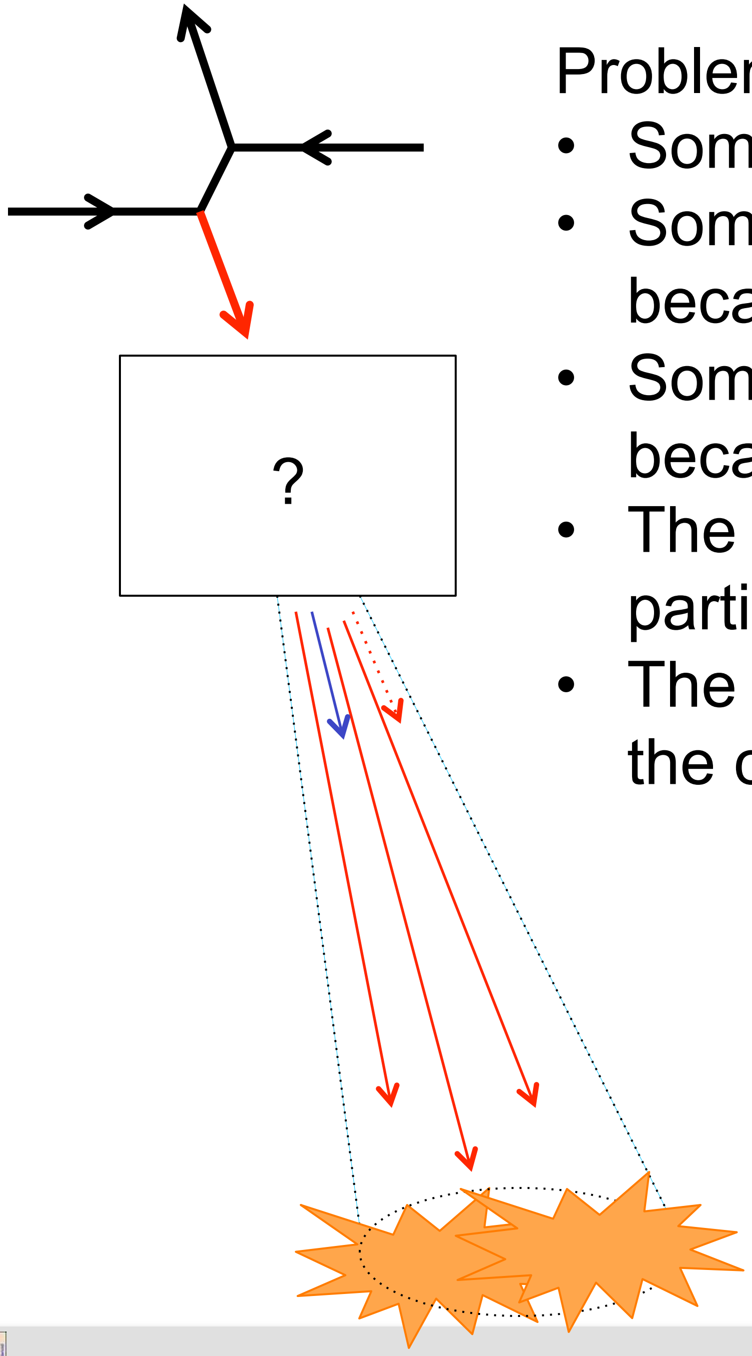
well defined by the clustering algorithm,

FastJet anti-k_T, R = 0.3

Energy-corrected to particle-level (PYTHIA) jets

NO constituent p_T threshold

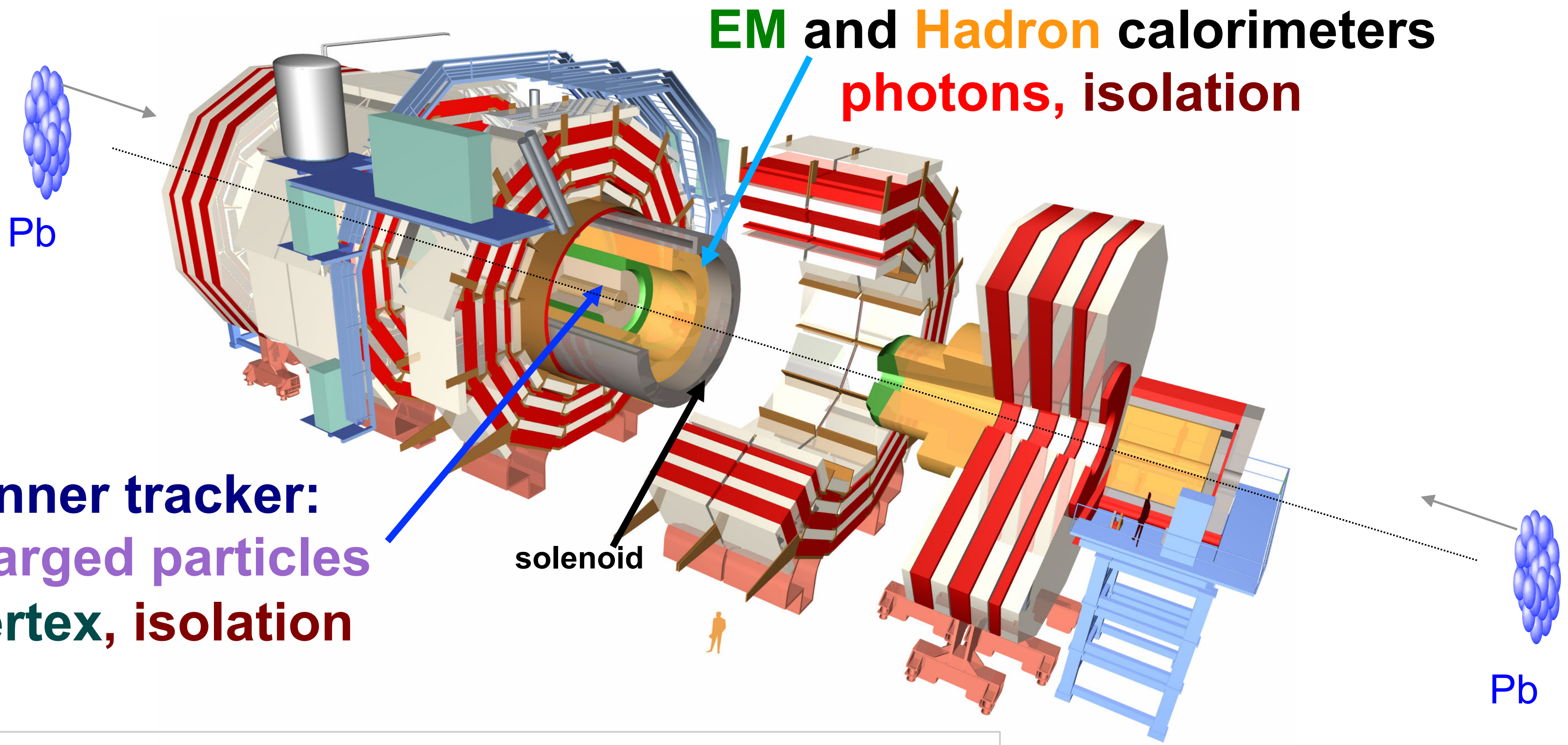
Which jets?



Problems:

- Some UE may still be there
- Some parton associated particles are lost because of reconstruction
- Some parton associated particles are lost because of bkg subtraction
- The calorimeter energy deposit of the final particles fluctuates
- The particle composition is different from what the corrections assume

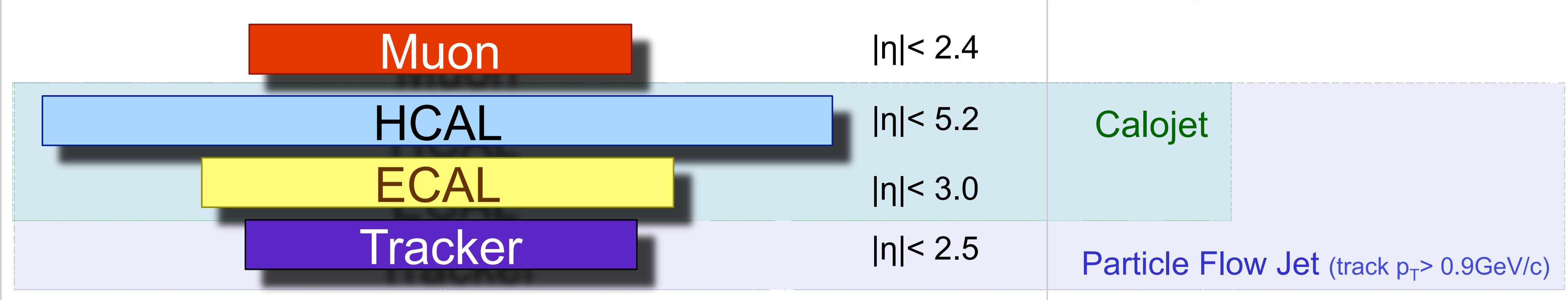
CMS detector



EM and Hadron calorimeters
photons, isolation

Inner tracker:
charged particles
vertex, isolation

solenoid



Muon |η|< 2.4

HCAL |η|< 5.2

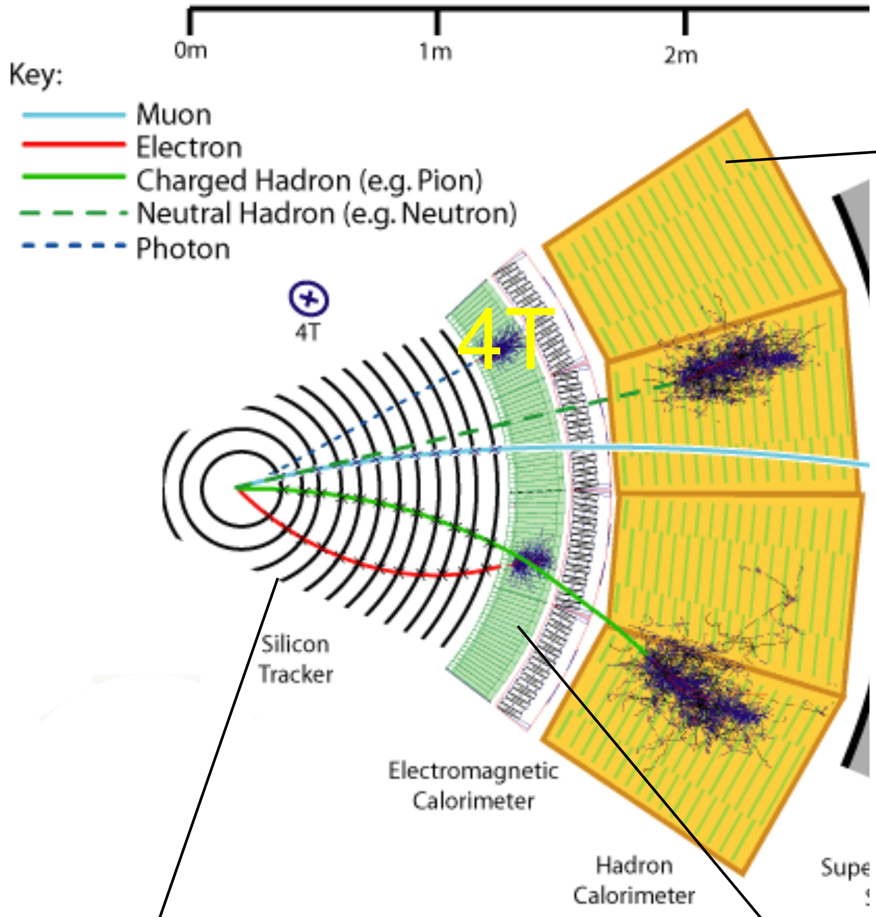
ECAL |η|< 3.0

Tracker |η|< 2.5

Calojet

Particle Flow Jet (track p_T> 0.9GeV/c)

CMS Detector



Hcal energy

- Neutral hadrons
- Capture charged hadrons that tracking missed
- Event-by-event shower fluctuations:
  - Non-linearity
  - Wide resolution
- Acceptance limited due to B-field
- Low granularity

TRACKS

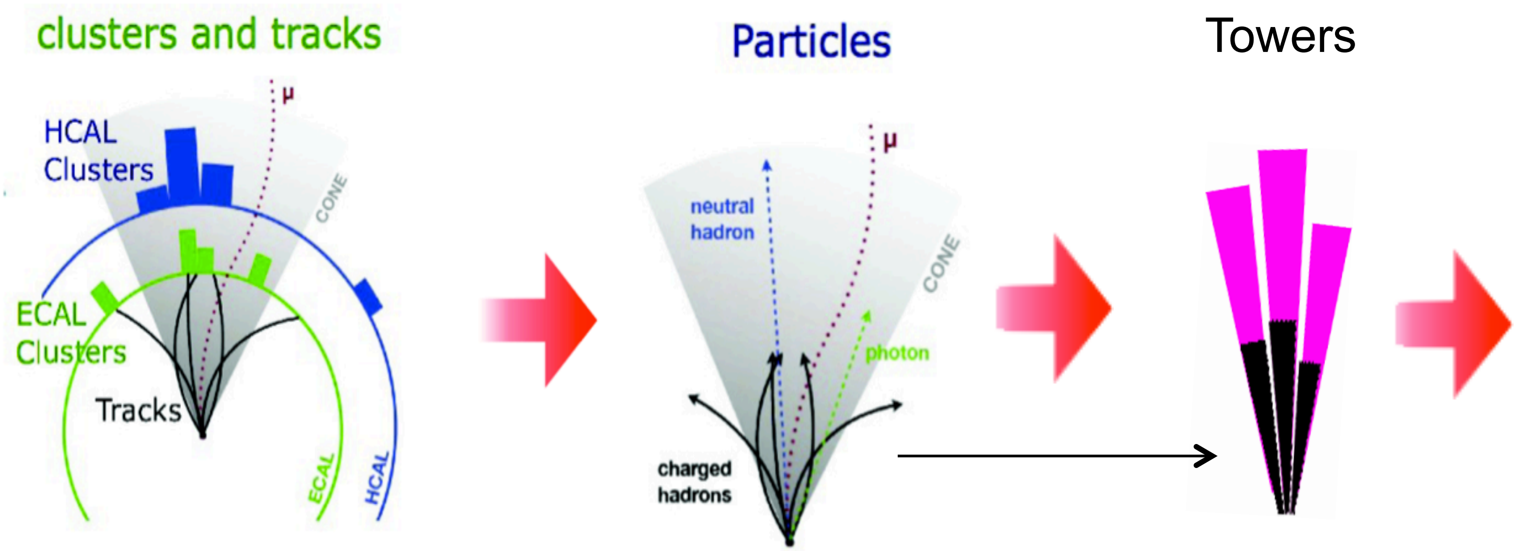
- Better resolution of p_T
- Blind to neutral energy
- Not 100% efficient
- Limited acceptance

EM candidates

- Photons

(Details: CMS-PAS-HIN-11-004)

ParticleFlow algorithm



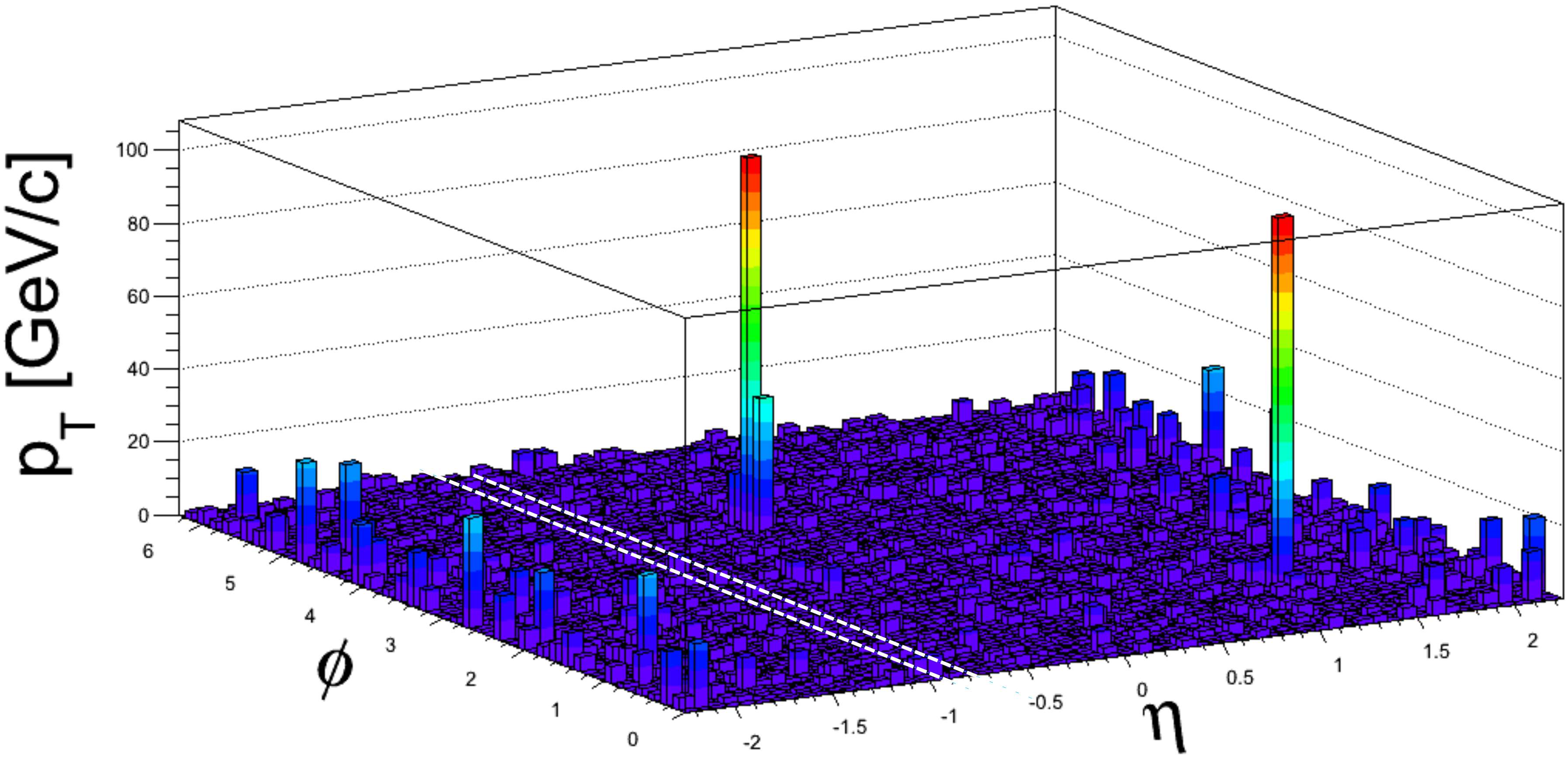
Background subtraction and jet clustering

(Tracking for only the primary vertex)

$\Delta\eta \times \Delta\phi$
0.076 x 0.076
in barrel

Calorimeter clusters and tracks are matched
(Details: CMS-PAS-HIN-11-004)
The candidates are merged into pseudo-towers in order to subtract background per segmentation

PU-subtraction algorithm



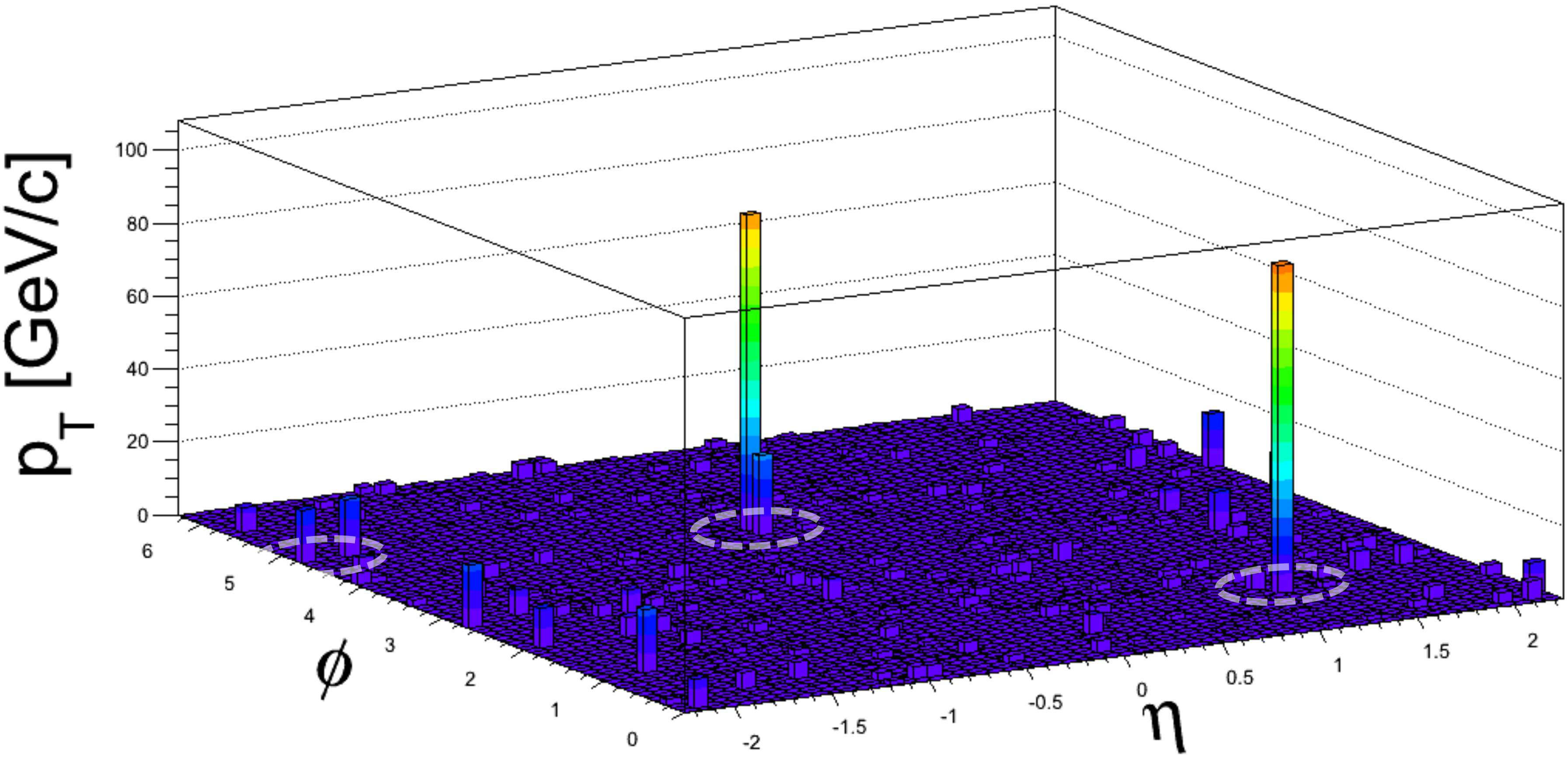
Estimate background
for each tower ring of constant η
estimated background = $\langle p_T \rangle + \sigma(p_T)$

- Captures dN/dη of background
- Misses ϕ modulation – to be improved

Tunable parameters:

- Coefficient of RMS

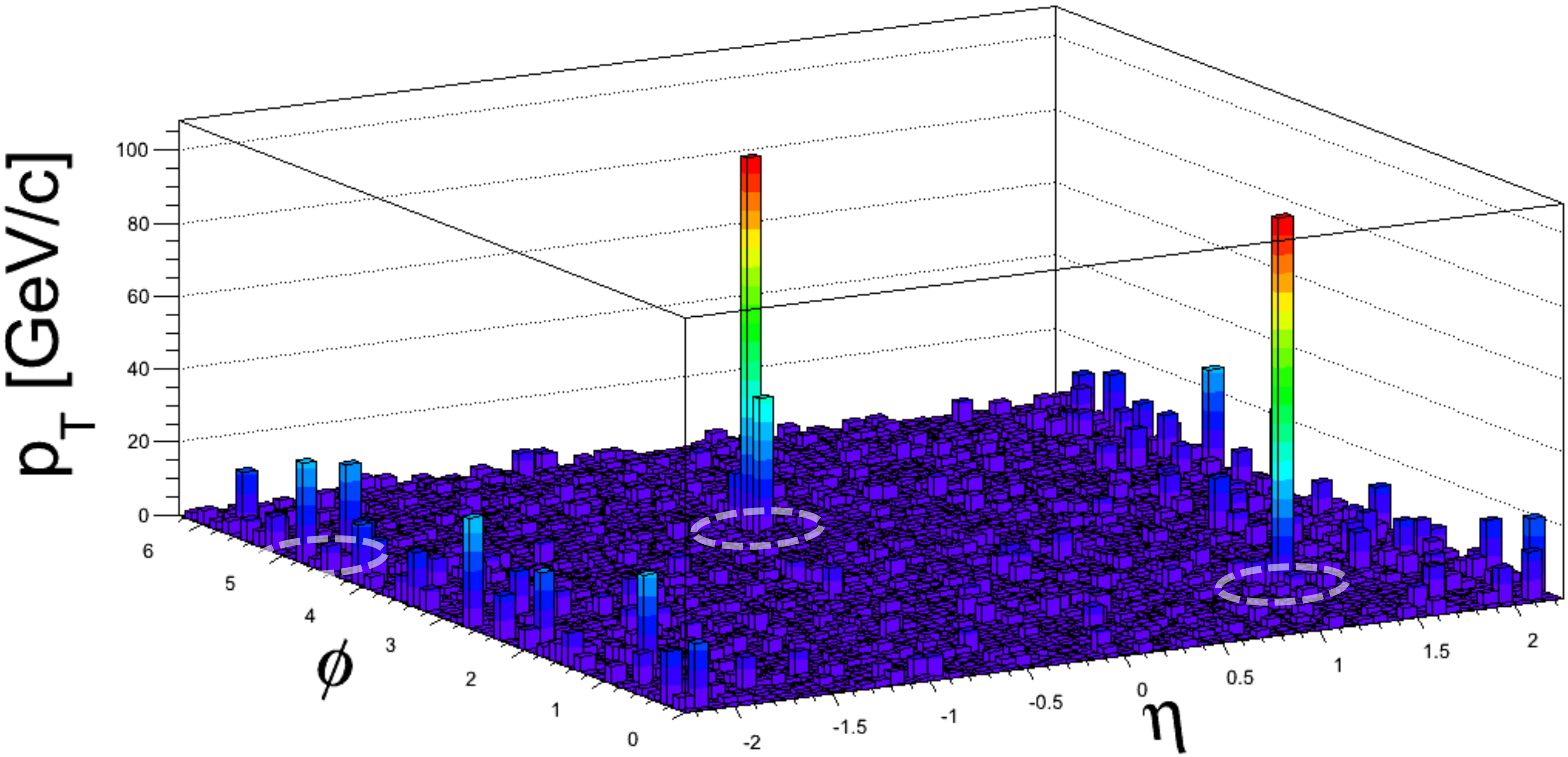
PU-subtraction algorithm



Subtract background from all towers

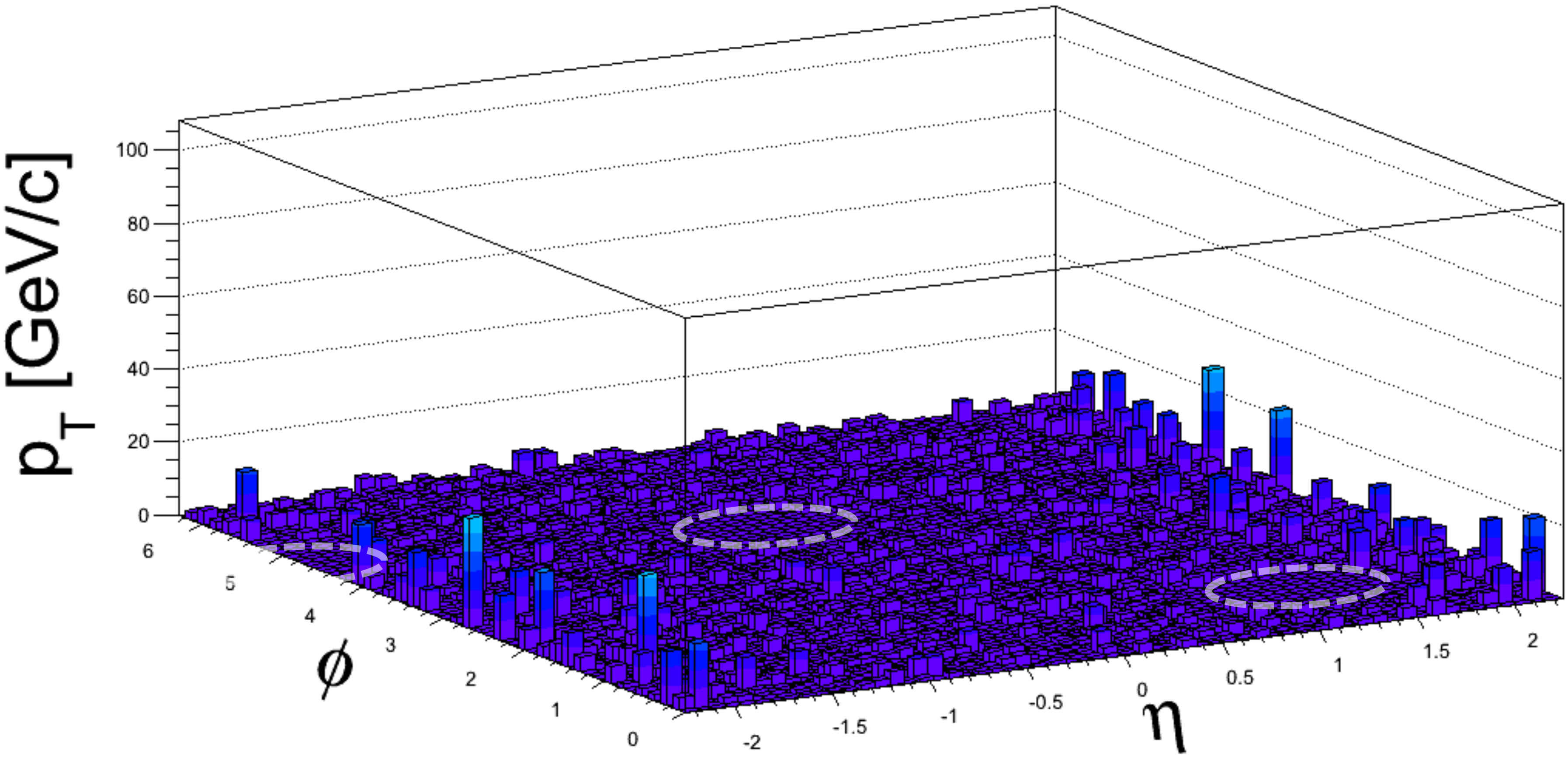
Run the clustering algorithm (anti-k_T)

PU-subtraction algorithm



Start over, knowing where the jets roughly are

PU-subtraction algorithm



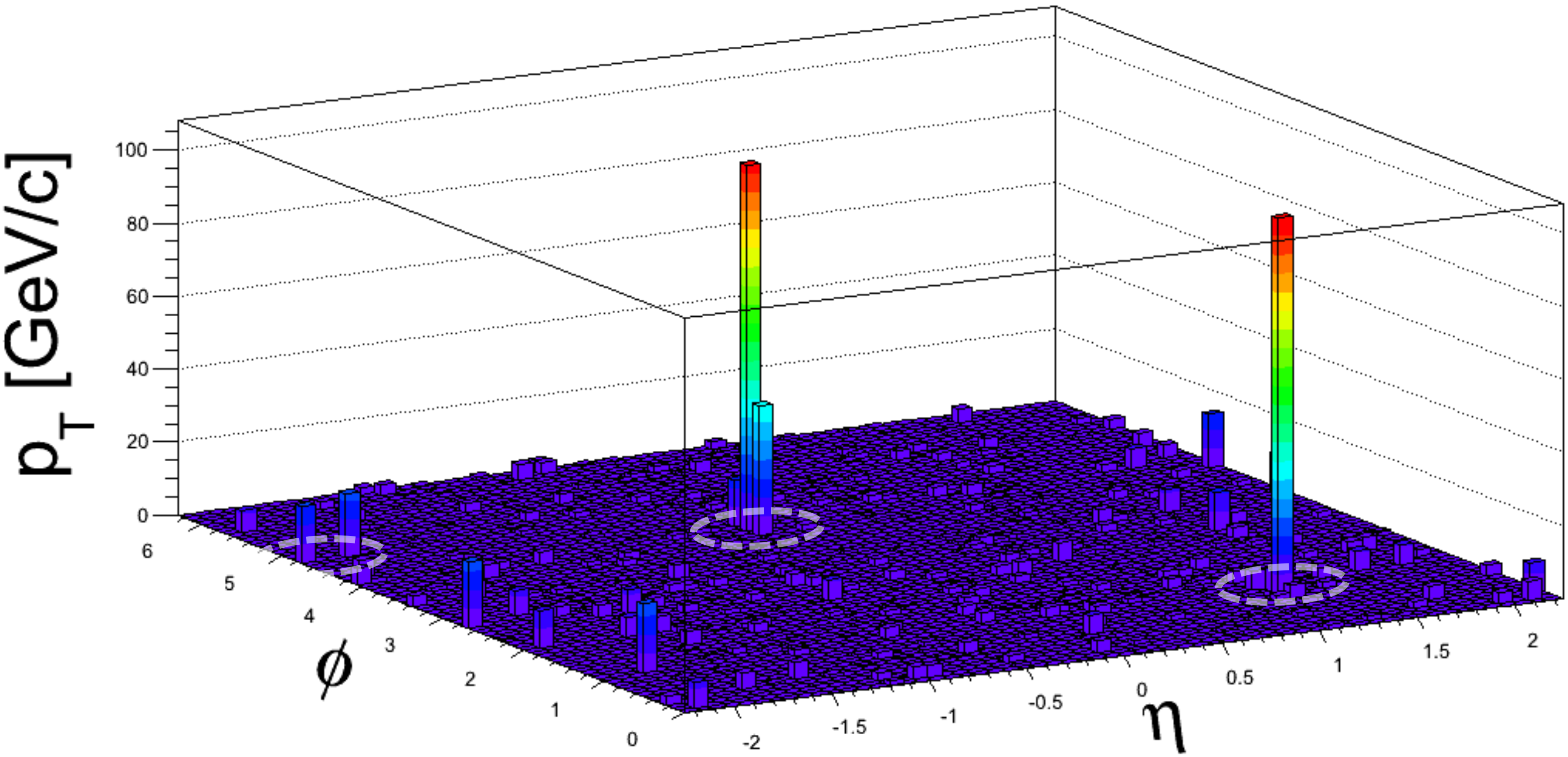
Start over, knowing where the jets roughly are
Exclude a certain area around the jets
Re-estimate the background for all towers

Tunable parameters:

- Coefficient of RMS
- Raw jet threshold
- Radius of exclusion

(not necessarily = R)

PU-subtraction algorithm



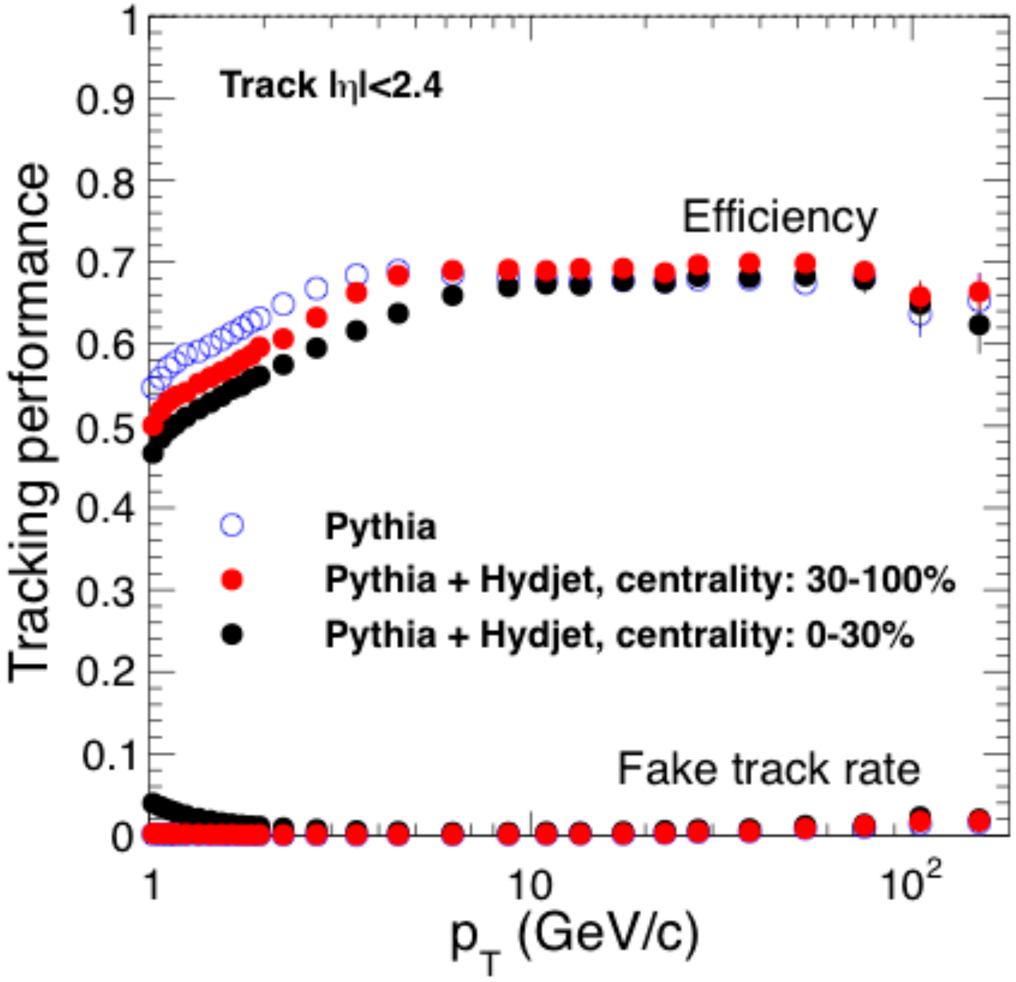
Start over, knowing where the jets roughly are
Exclude a certain area around the jets
Re-estimate the background for all towers
Subtract final background
Cluster jets

Tunable parameters:

- Coefficient of RMS
- Raw jet threshold
- Radius of exclusion

(not necessarily = R)

Tracking



CMS-PAS-HIN-12-013

Validate efficiency by:

- Analysis of hadron spectra in pp
- Track multiplicity distributions
- CaloTower-track matching
- CaloJet-PFJet matching

Important to understand in fragmentation analysis:

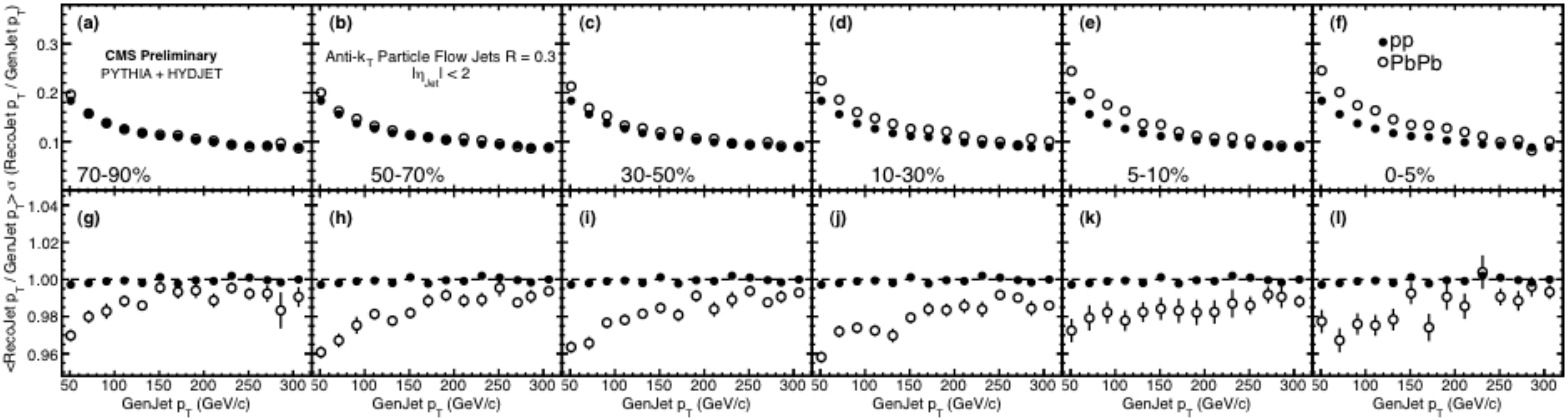
- A higher fluctuation in jet response may correspond to efficiently found tracks
  - Impact on tracking efficiency within “leading jet”
  - Impact on AJ dependence of the tracking efficiency

Systematic uncertainties in analyses

	Jet energy scale	Jet pointing resolution	Jet energy resolution	Fake jets / Noise	Track UE background subtraction	Tracking efficiency
Dijet (γ-Jet) pT Imbalance	*	N/A	**	X	N/A	X
Dijet (γ-jet) Azimuthal Correlation	*	**	**	X	N/A	X
Dijet Missing p_T	*	*	*	X	N/A	**
Inclusive Jet Fragmentation	**	N/A	**	X	**	**
Inclusive Jet Shape	*	*	*	X	**	**
Inclusive Jet R_{AA}	*	N/A	**	*	N/A	X
Inclusive Jet Spectra	**	N/A	**	*	N/A	X
b-jet Fraction	**	N/A	**	*	N/A	**

X → negligible effect, * → important systematics, ** → dominant systematics

Jet energy response



- Corrections derived from PYTHIA (pp) simulations
- Resolution increases (N term – see later slide) by ~ 5 GeV with centrality
- Energy scale shifted by ~2-1% due to subtraction of low-p_T

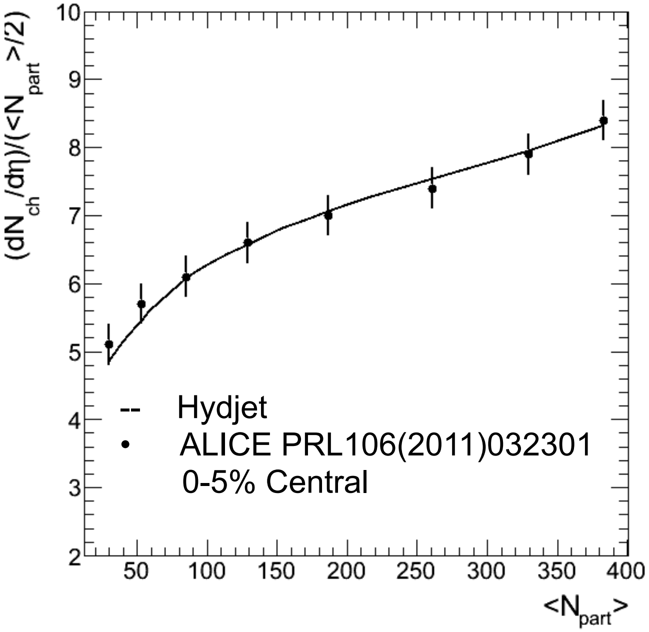
- Dependence on other properties (parton-type, fragmentation) are examined to evaluate systematic uncertainties

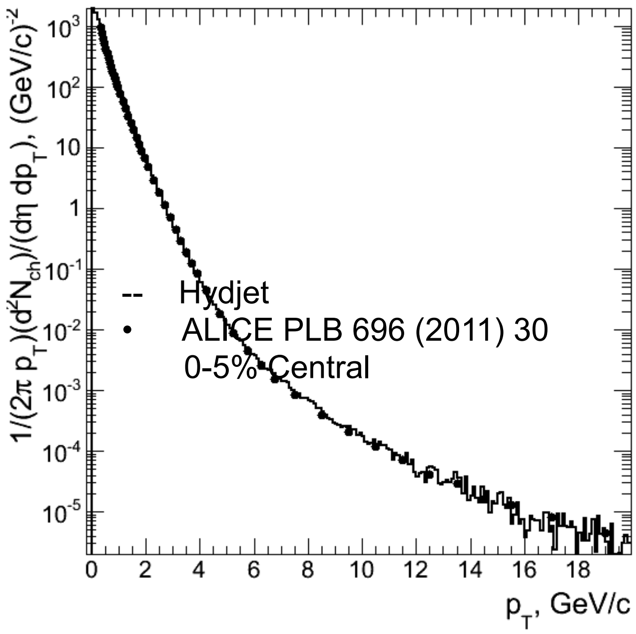
See: CMS-PAS-HIN-12-004

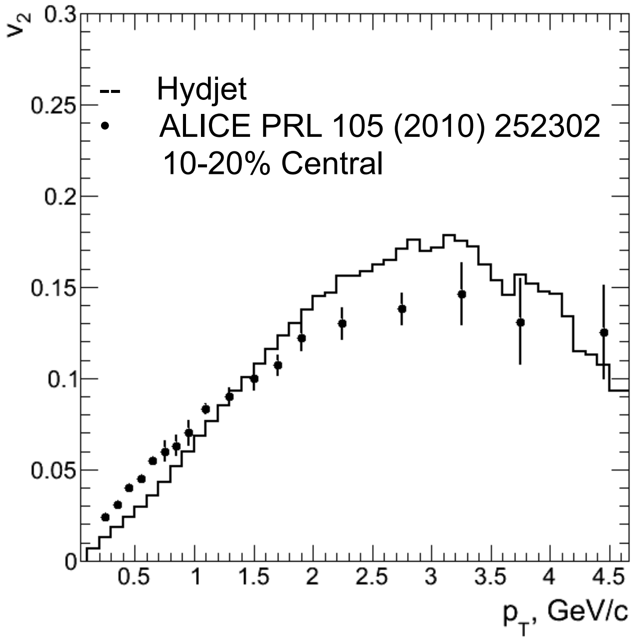
Underlying event effects

Effects in the reconstruction:

- Tracks : Efficiency, fakes
- Jets : Energy scale, resolution, fake-jets
- Hydjet 1.8 default tune successfully reproduces:
  - Charged hadron multiplicity
  - Charged hadron p_T spectrum
  - Azimuthal asymmetry of low-p_T particles (Elliptic Flow)
- Pythia dijets embedded into Hydjet and fully simulated

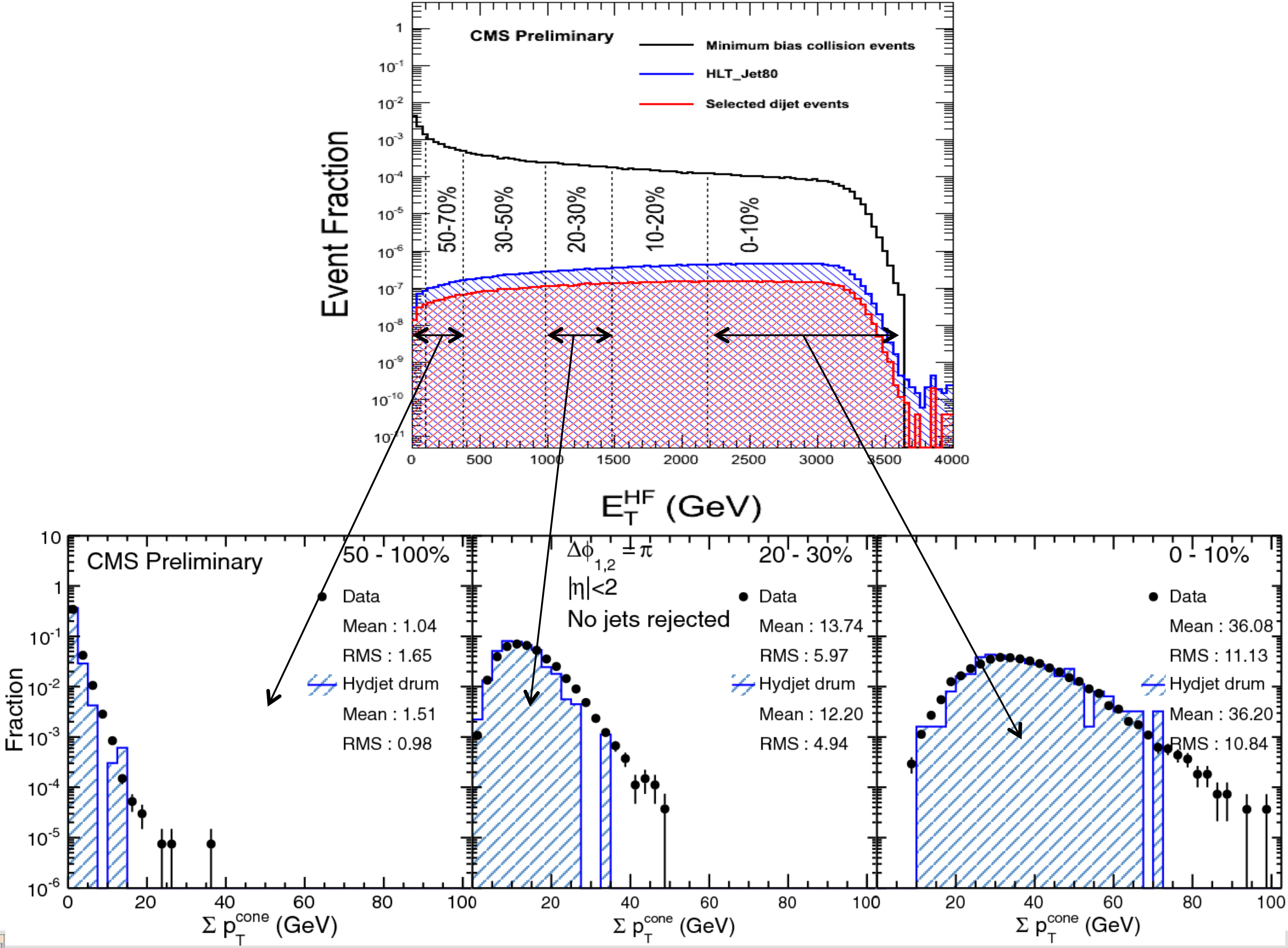




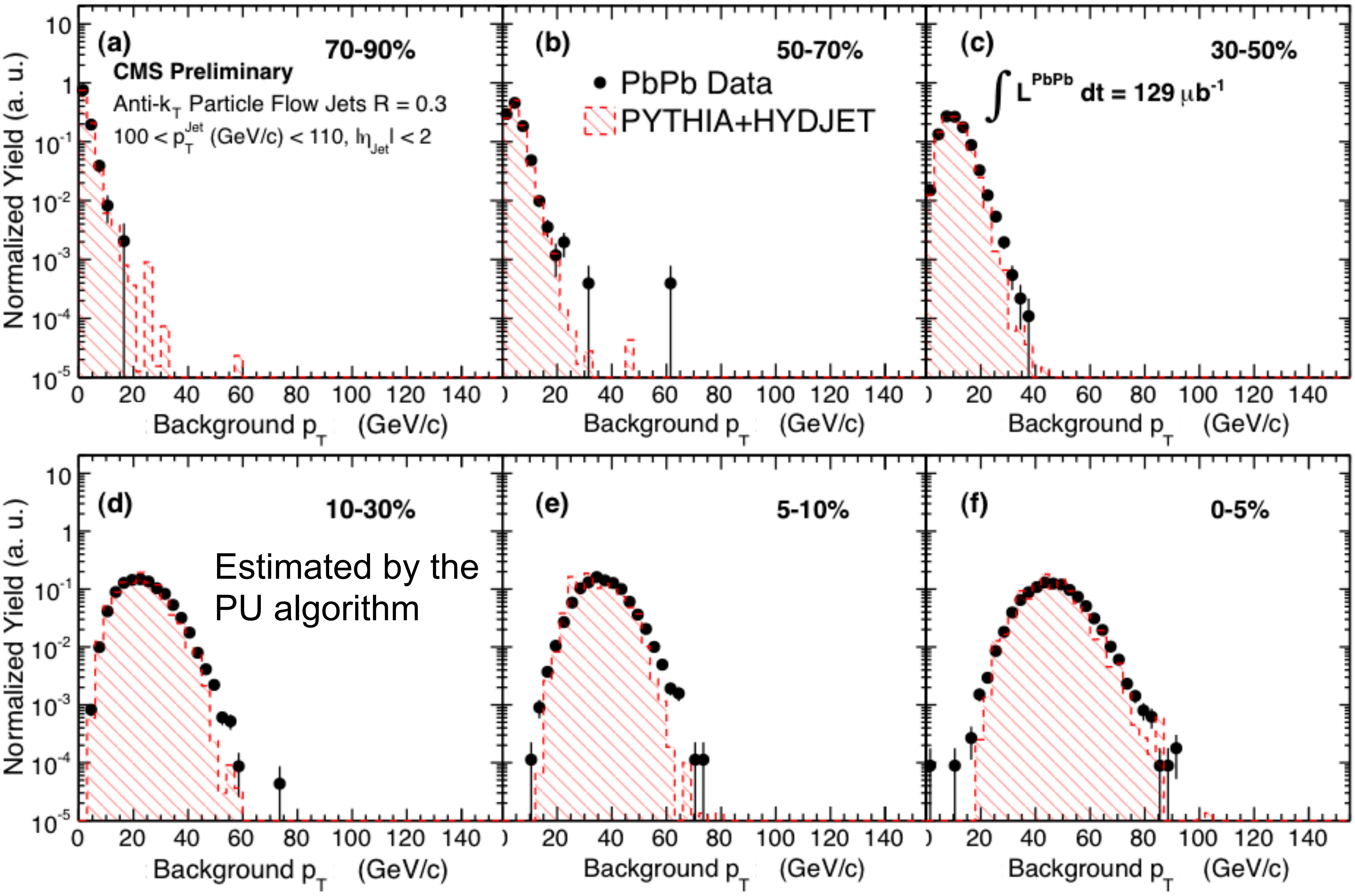


http://lokhtin.web.cern.ch/lokhtin/hydro/plots

PbPb event simulations with Hydjet 1.8

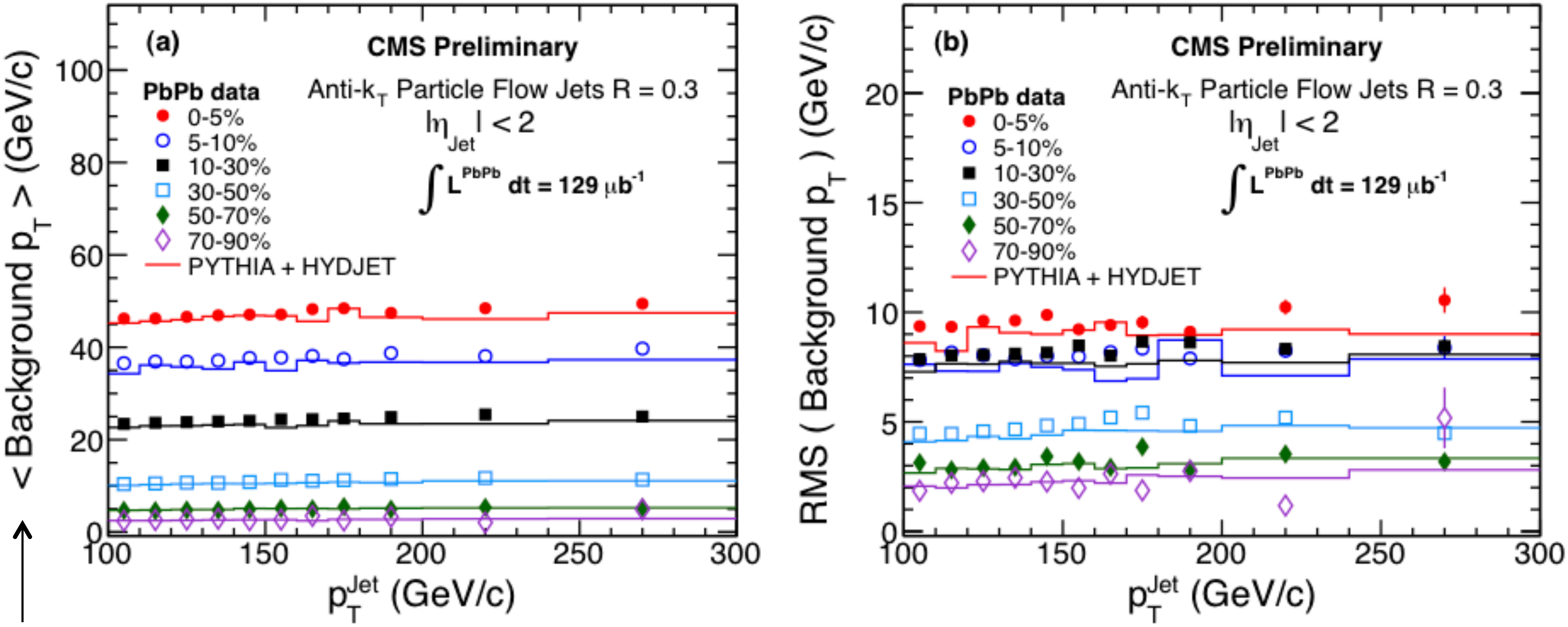


Validation of background



Estimated by the PU algorithm

Validation of background

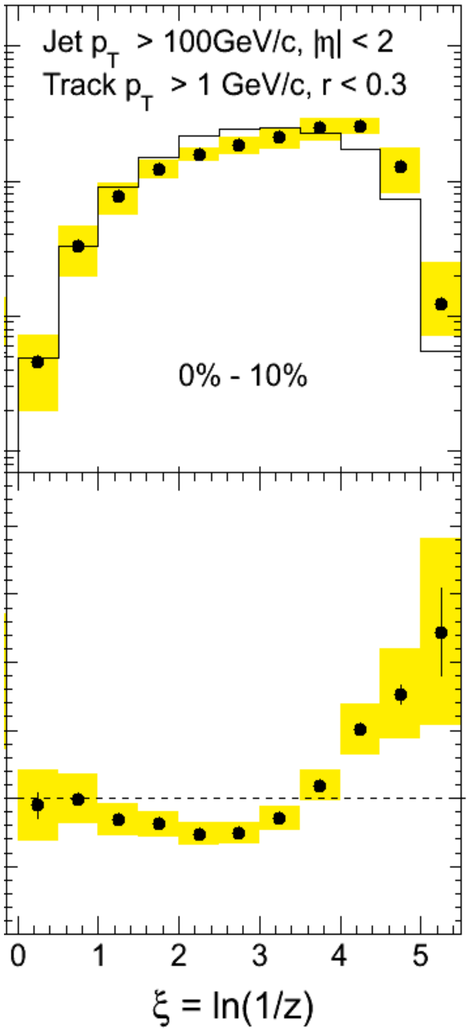


Background under each jet, as estimated by the PU algorithm

Very slight dependence on jet p_T

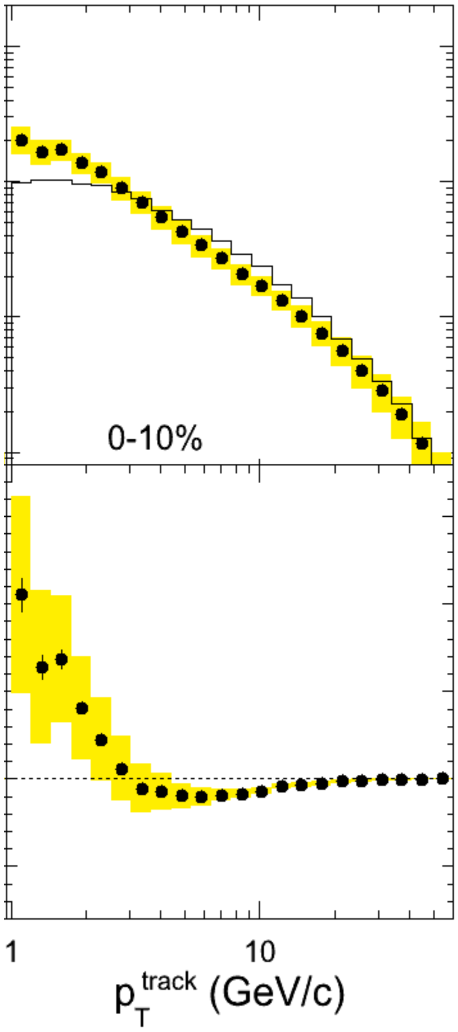
Jet composition, and tower occupancy may change

Fragmentation effects on jets



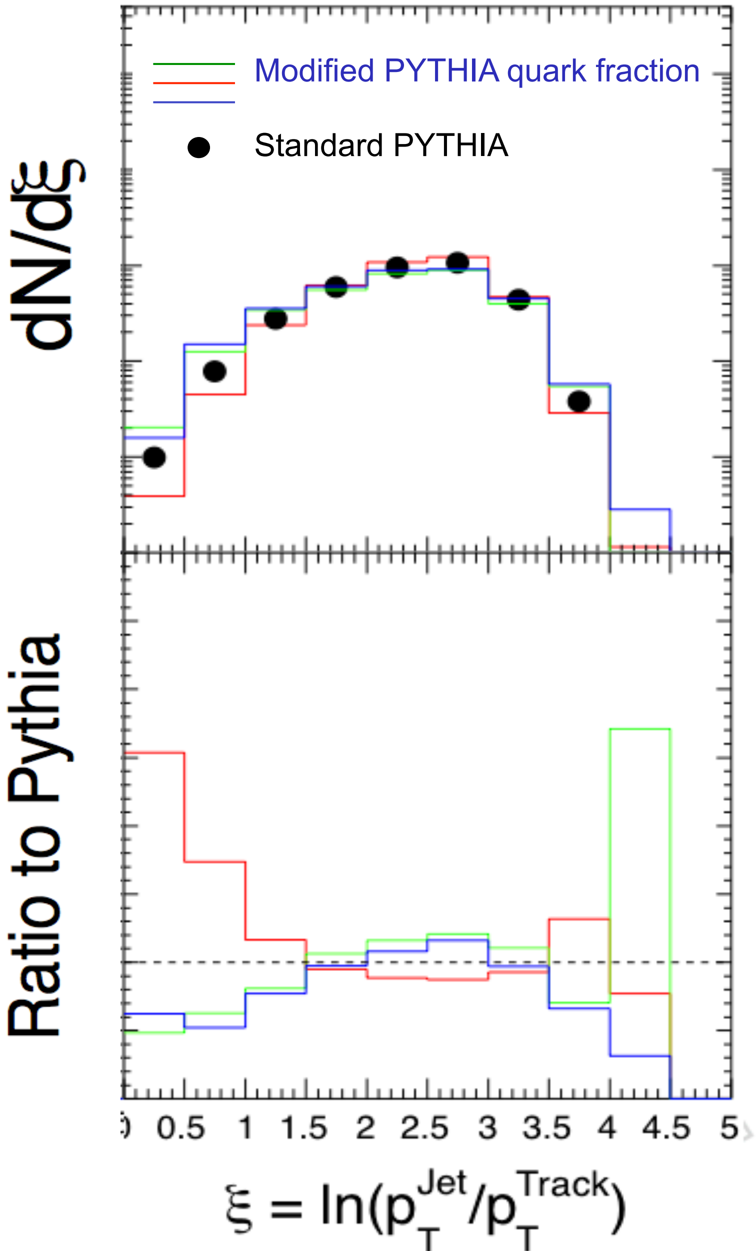
- The hard part of the fragmentation is slightly modified
  - May affect calorimeter-related resolution
- Effects can be estimated by
  - Modified Pythia parton content
  - Various Pyquen tunes

- There appears to be an enhanced soft component
  - May interfere with PU subtraction to affect energy scale
- Effects can be estimated by
  - Embedding tracks into jets
  - Various Pyquen tunes



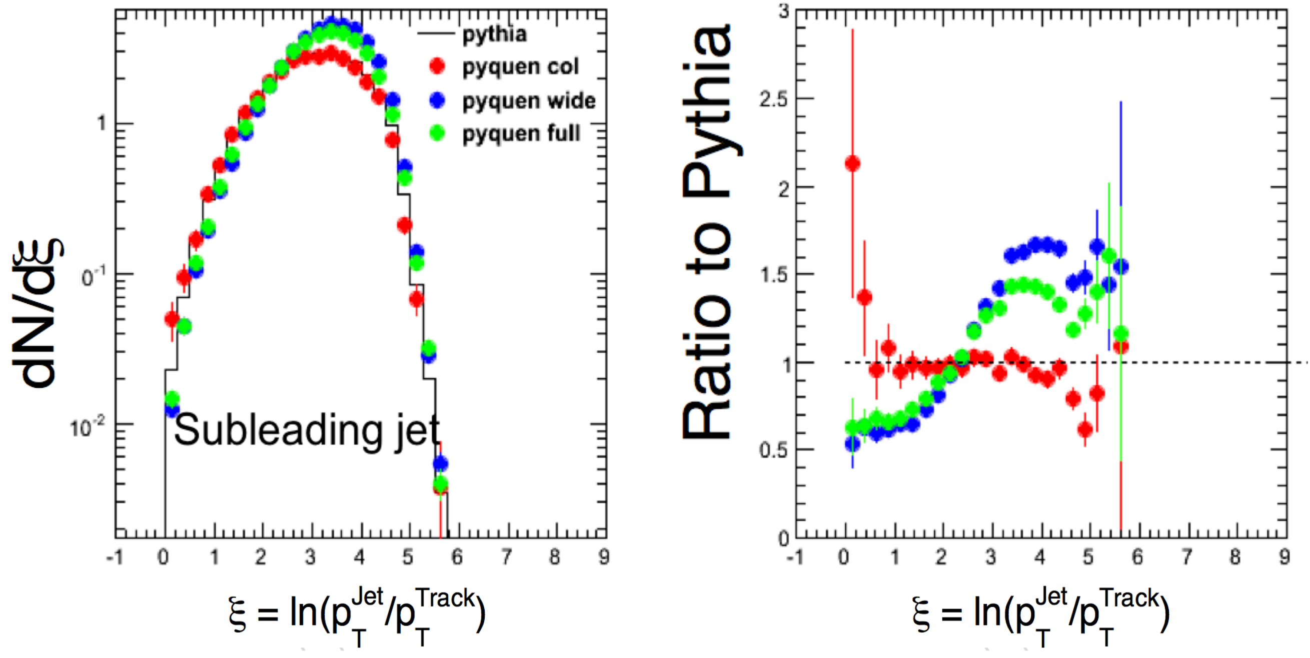
CMS-PAS-HIN-12-013

Fragmentation effects on jets



MODIFIED PYTHIA:
Parton fractions are selected/reweighted to bias the sample for a specific type of hard fragmentation

PYQUEN:
Different tunes with Radiational/Collisional energy loss



Fully simulated events are studied for determination of systematics

Jet performance and Data-Theory comparison

$$\sigma\left(\frac{p_{\mathrm{T}}^{\mathrm{Reco}}}{p_{\mathrm{T}}^{\mathrm{Gen}}}\right) = C \oplus \frac{S}{\sqrt{p_{\mathrm{T}}^{\mathrm{Gen}}}} \oplus \frac{N}{p_{\mathrm{T}}^{\mathrm{Gen}}},$$

C	S	N (pp)	N (50–100%)	N (30–50%)	N (10–30%)	N (0–10%)
0.0246	1.213	0.001	0.001	3.88	5.10	5.23

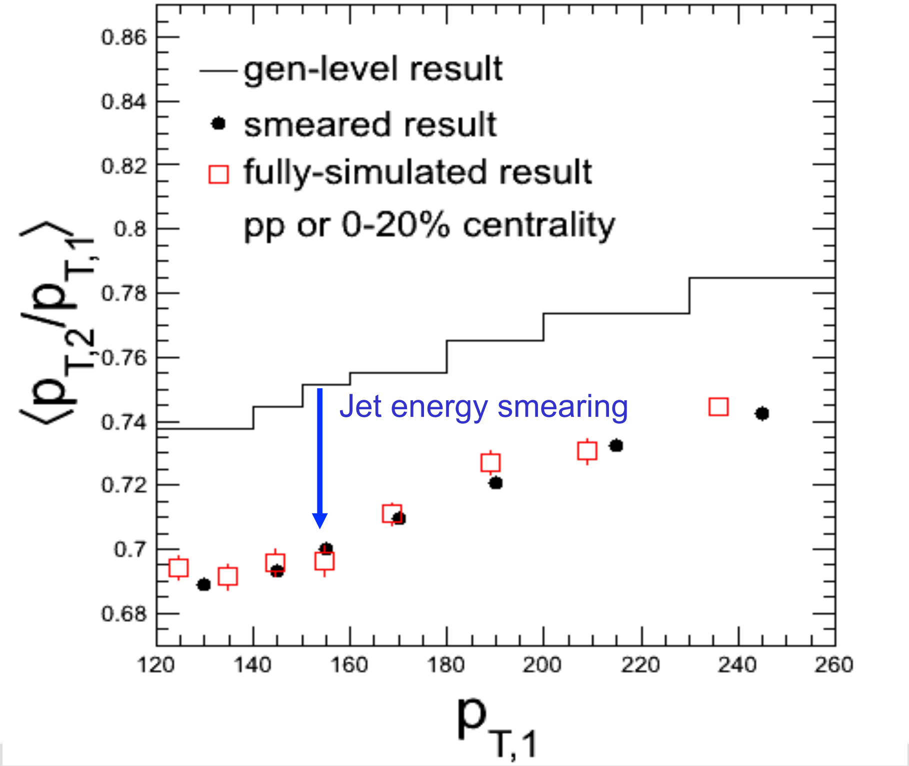
For 30 GeV jets : C component ~ 0.7 GeV, S ~ 6.5 GeV N ~ 5 GeV

For 120 GeV jets : C component ~ 3 GeV, S ~ 13 GeV, N ~ 5 GeV

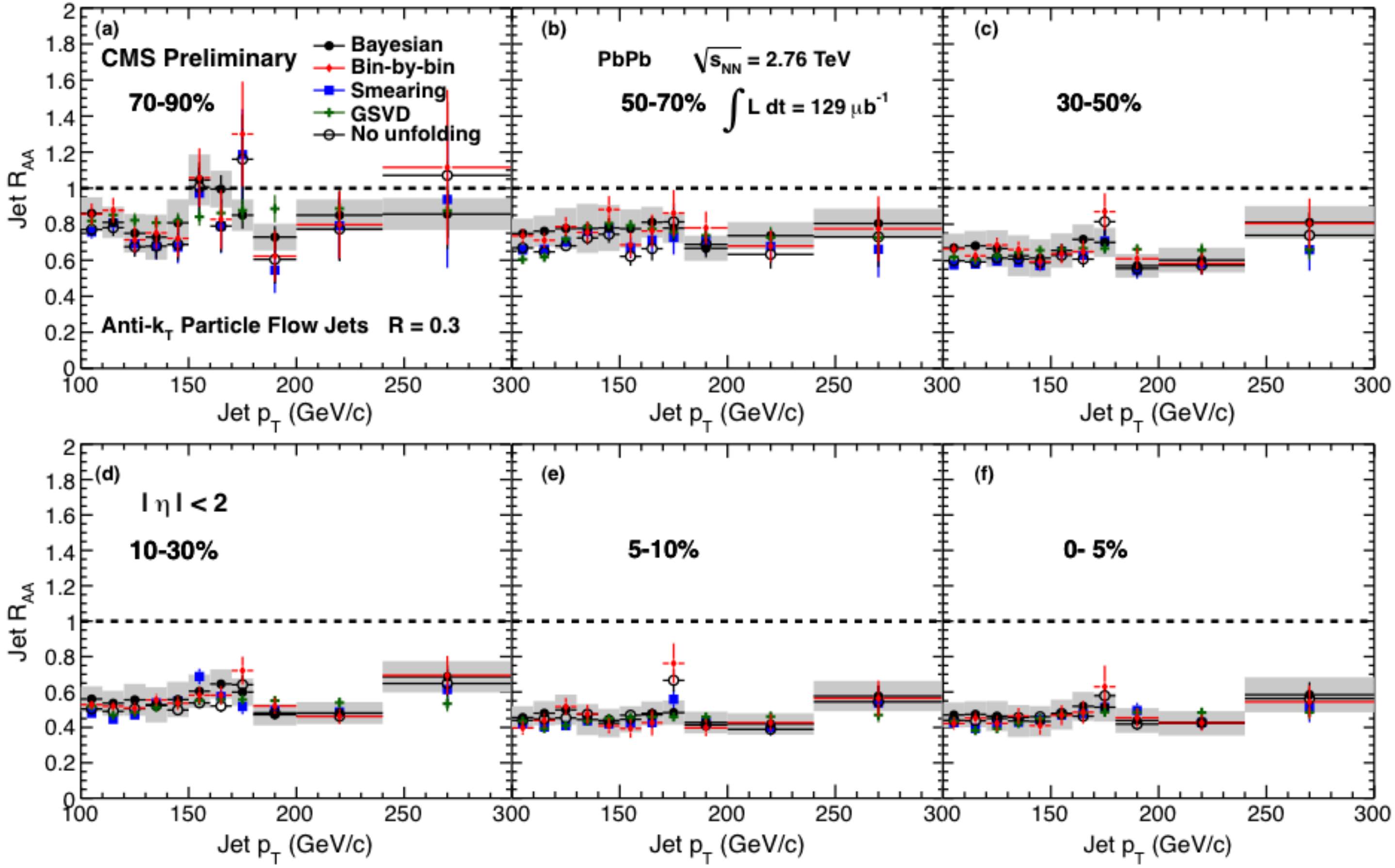
From PLB 718 (2013) 773 : Photon events

Not exactly inclusive jet resolution

To be updated



Unfolding



CMS-PAS-HIN-12-004

- Multiple unfolding methods yield consistent results,
- Bayesian chosen to be primary method
- Full response matrix is used in unfolding
  (no fits, no gaussian assumption)
- No assumption on functional form
- Unfolded result consistent also with “smeared” result

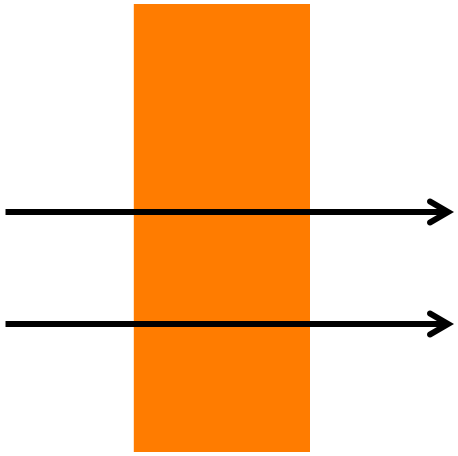
Modeling data

Generate events

Find Jets → Compare with unfolded analysis

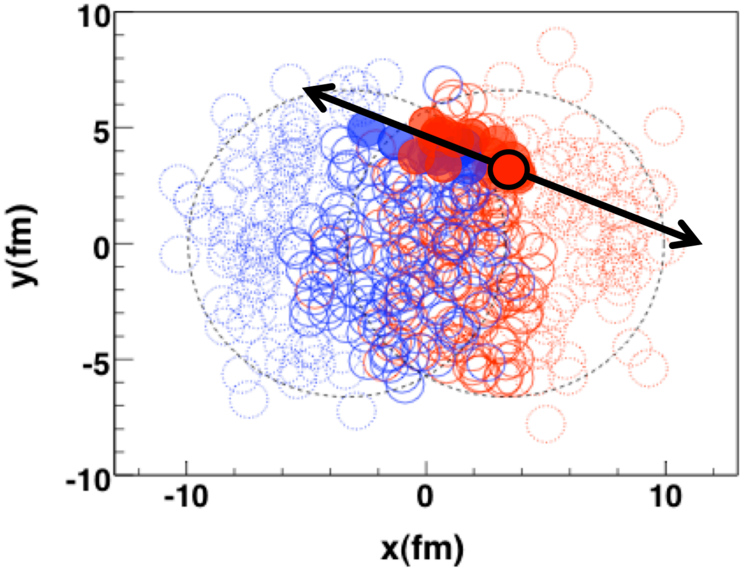
Smear → Compare with reco-level analysis

Brick-type toy-model



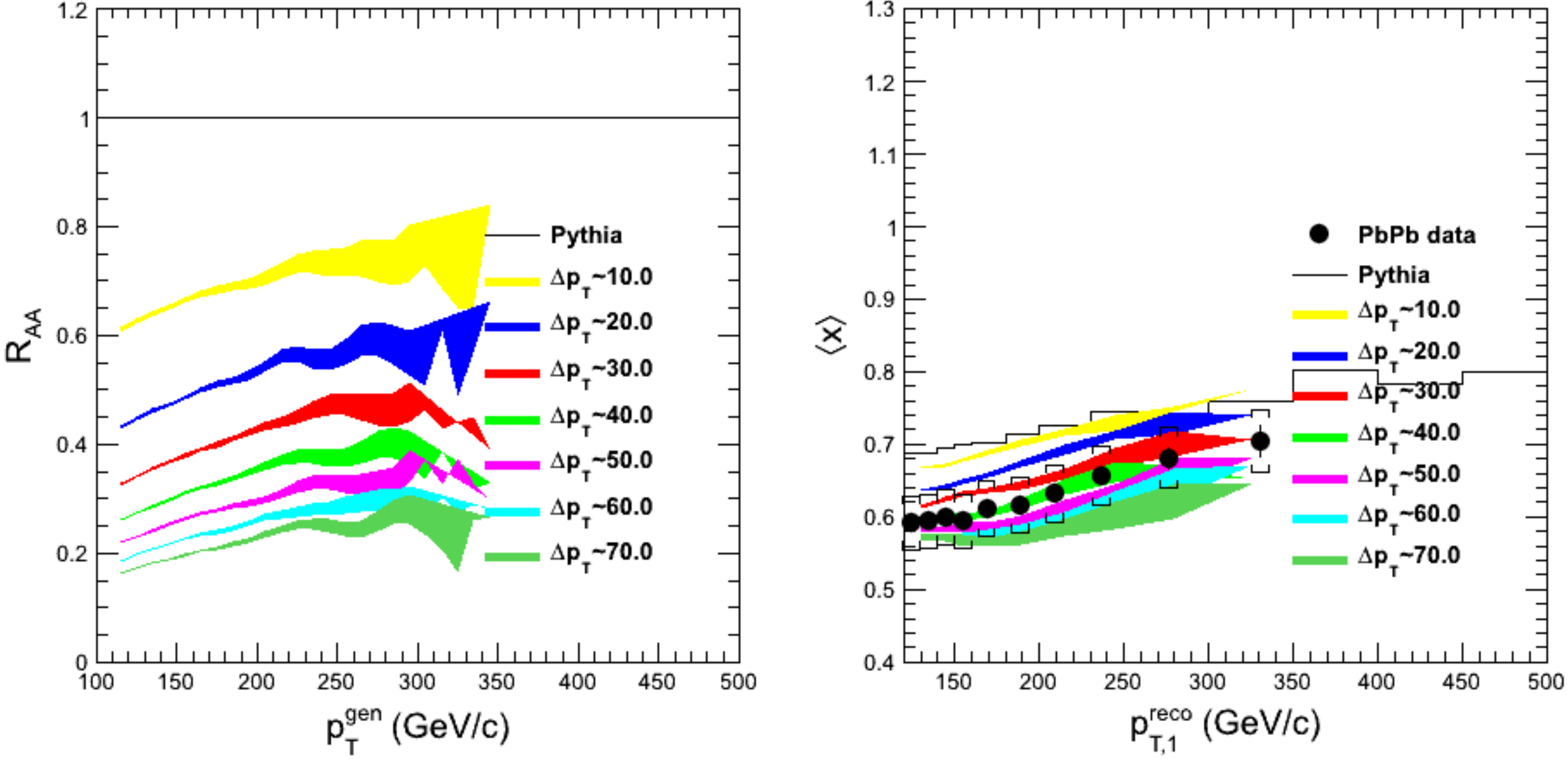
(e-loss is independent between the two jets)

Glauber-inspired toy-model



Modeling data

Brick-type toy-model

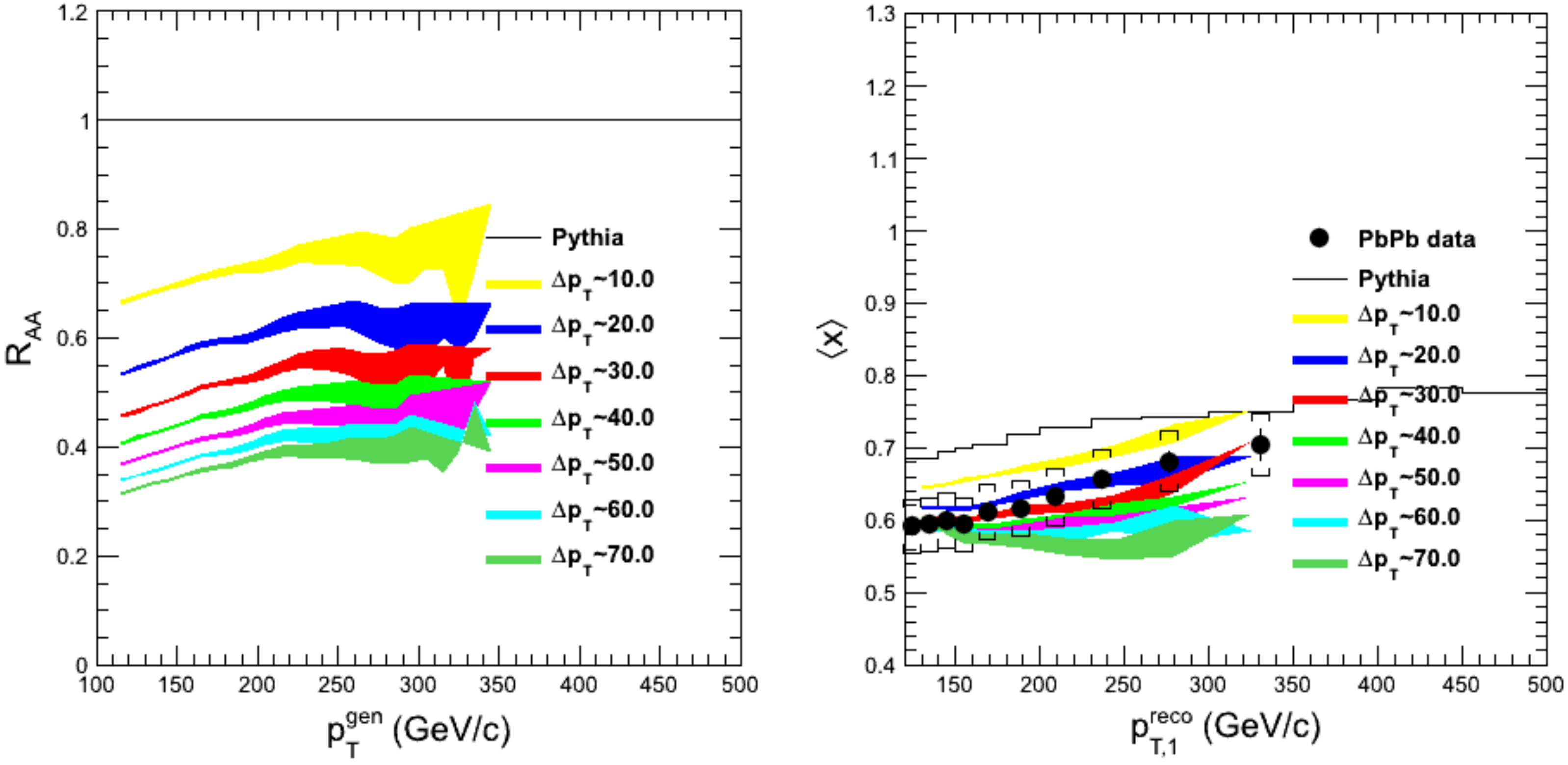


(e-loss is independent between the two jets)

- The different results may help to calibrate the comparison
- Jet RAA is compared in GenJet level, whereas imbalance compared in RECO (smeared) level
- Having smaller uncertainties would allow stronger exclusion

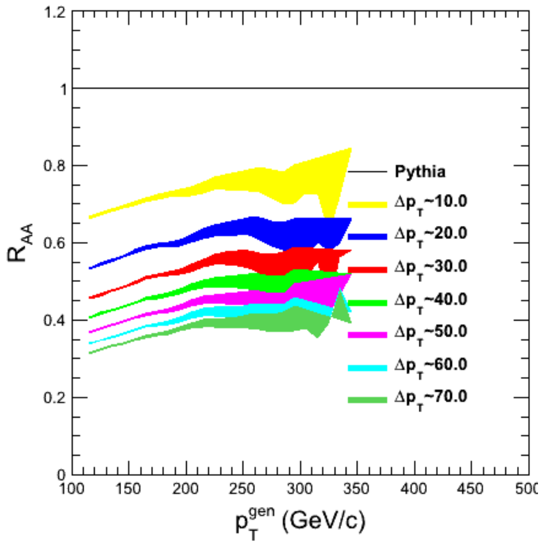
Modeling data

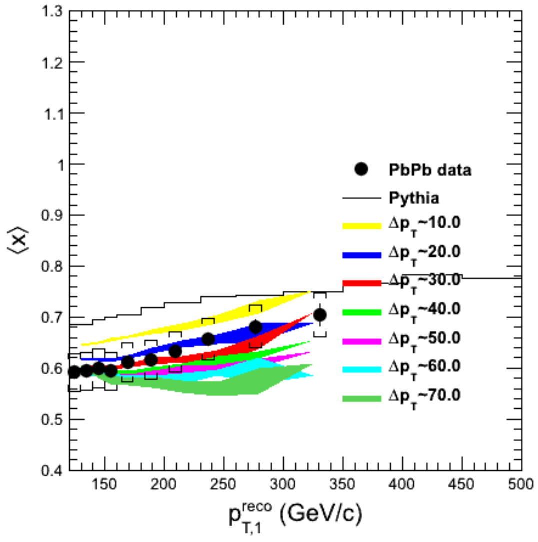
Glauber-inspired toy-model

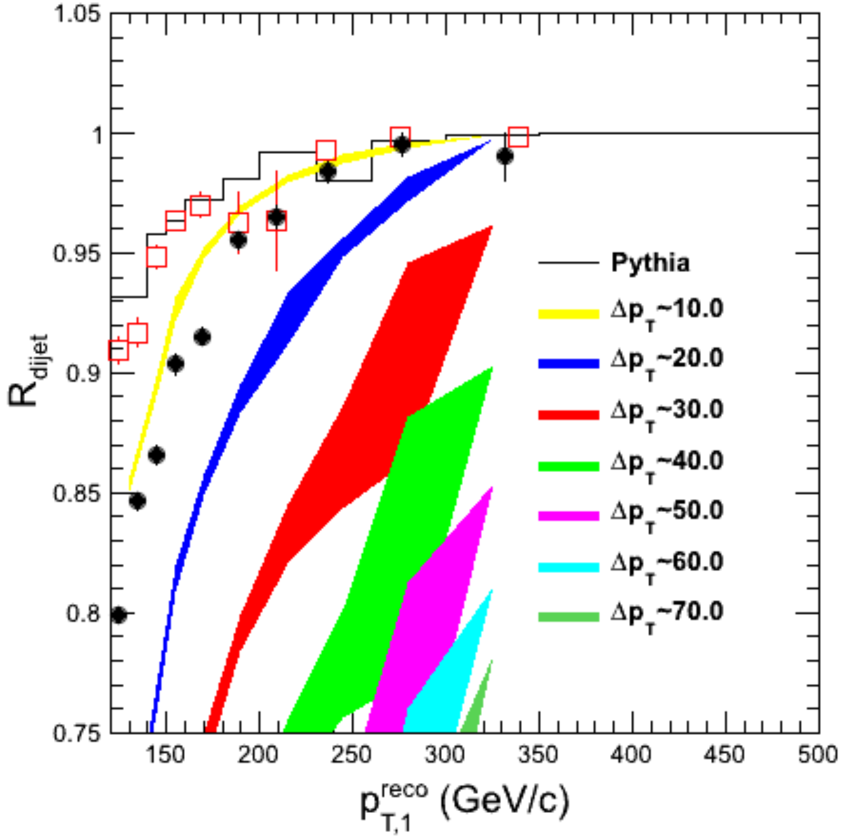


- The different results may help to calibrate the comparison
- Jet RAA is compared in GenJet level, whereas imbalance compared in RECO (smeared) level
- Having smaller uncertainties would allow stronger exclusion

Modeling data







Matched-dijet fraction also affected by the assumed quenching

- The different results may help to calibrate the comparison
- Jet RAA is compared in GenJet level, whereas imbalance compared in RECO (smeared) level
- Having smaller uncertainties would allow stronger exclusion

Discussion

There is room for improvement in both methods and calibrations of the reconstruction

Interesting phenomena is being observed already, which both
reveals physics, and
poses challenges for the reconstruction performance

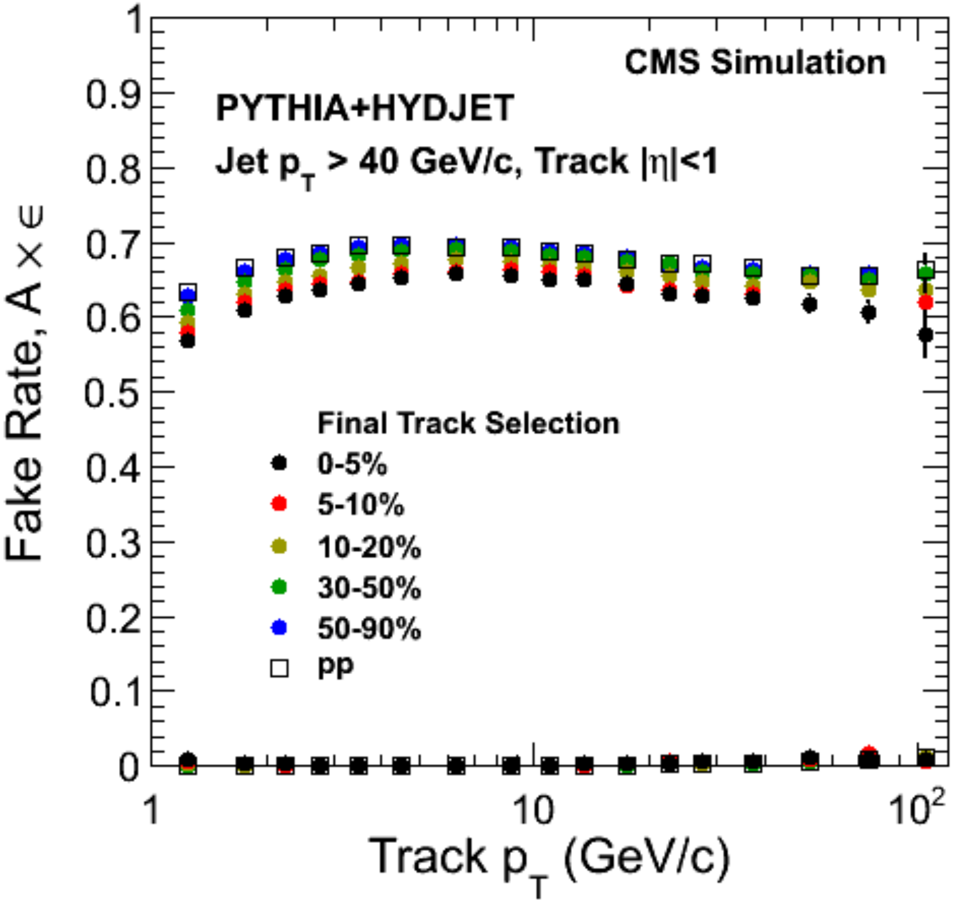
A greater variety of MC may help reduce the systematic uncertainties

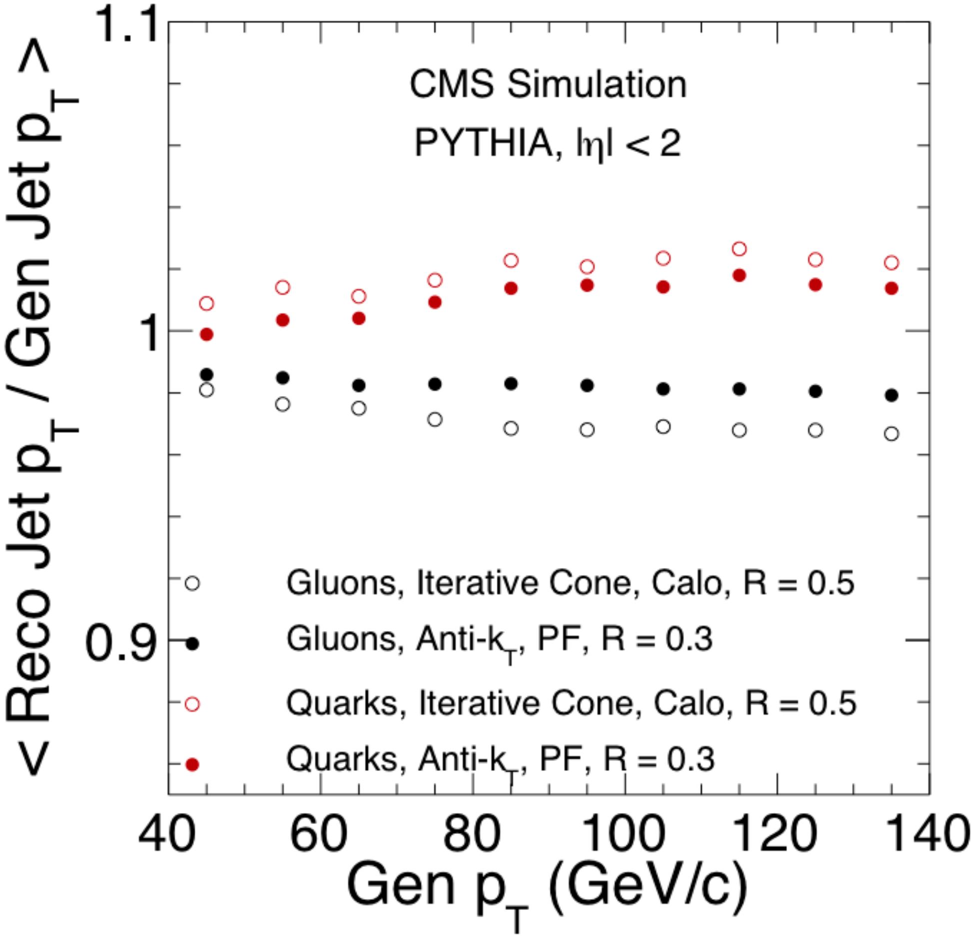
- Signal (dijet, photon-jet etc) MC
  - Tunable fragmentation
- Underlying event MC
  - Fluctuations, flow
- Interaction between the two
  - Input geometry model
  - Common framework (e.g. LHEvent – Pythia hadronization interface)

Thanks

Back-up slides

Jet response to parton types in PYTHIA

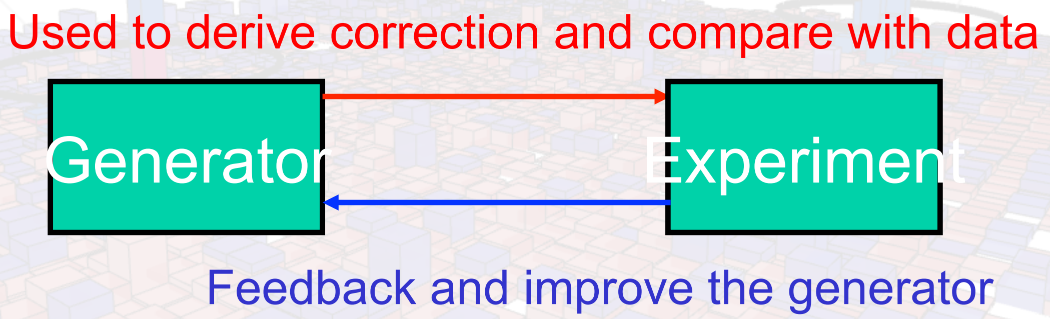




CMS-PAS-HIN-11-004

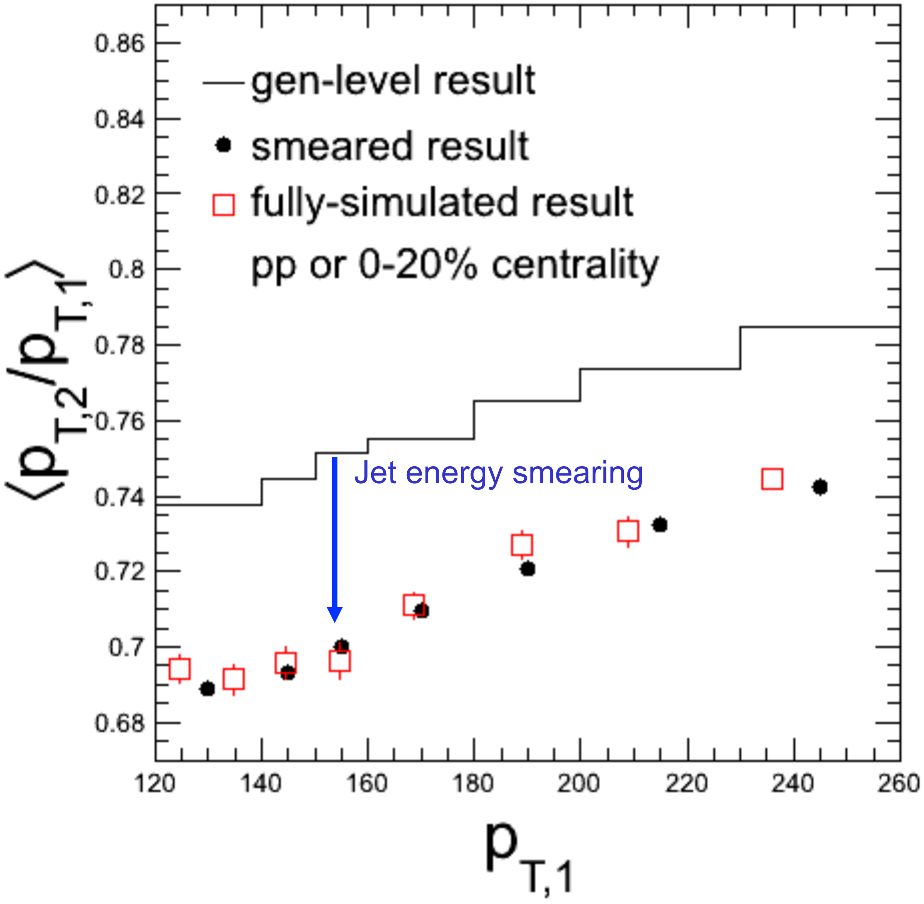
Inputs to the MC discussion

- Need realistic MC generator (for both jet and UE)
  - Iterative feedback cycle is very important (like PYTHIA v.s pp data in high energy community)
  - CMS is willing to use and check if you offer a jet event generator

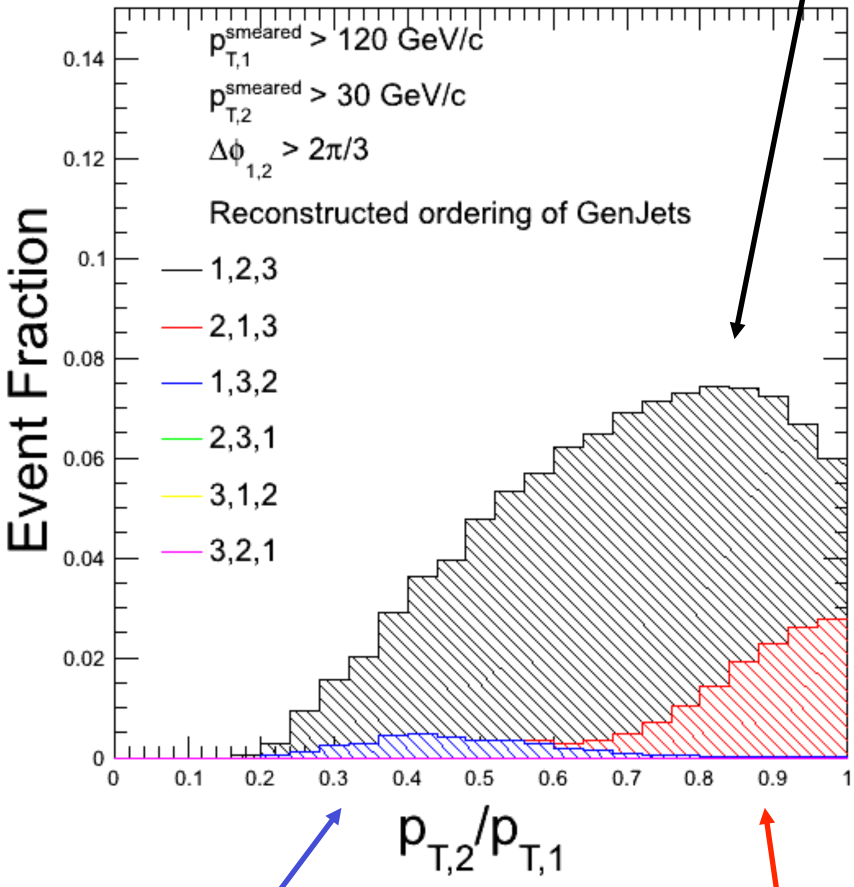


- To compare with CMS data:
  Need to include reconstruction effect properly
  - Genjet → energy loss → apply smearing

Inputs to the MC discussion



Generator level leading and subleading jets matches reco level

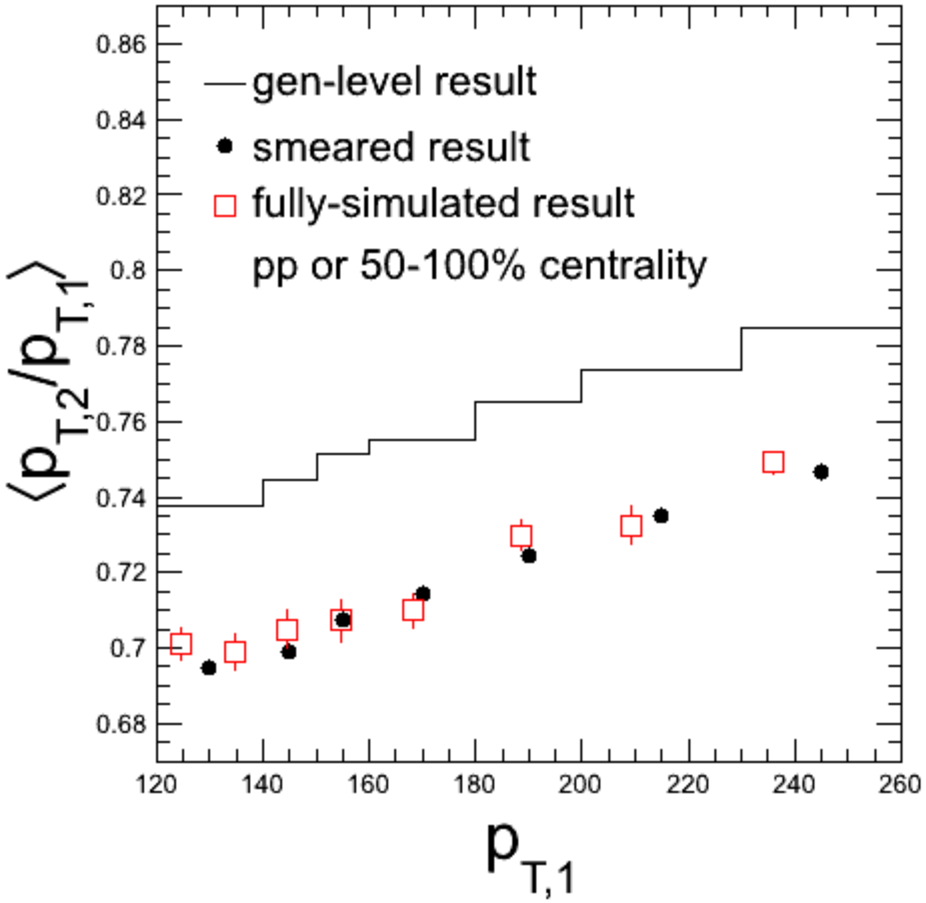


Subleading jet out of acceptance

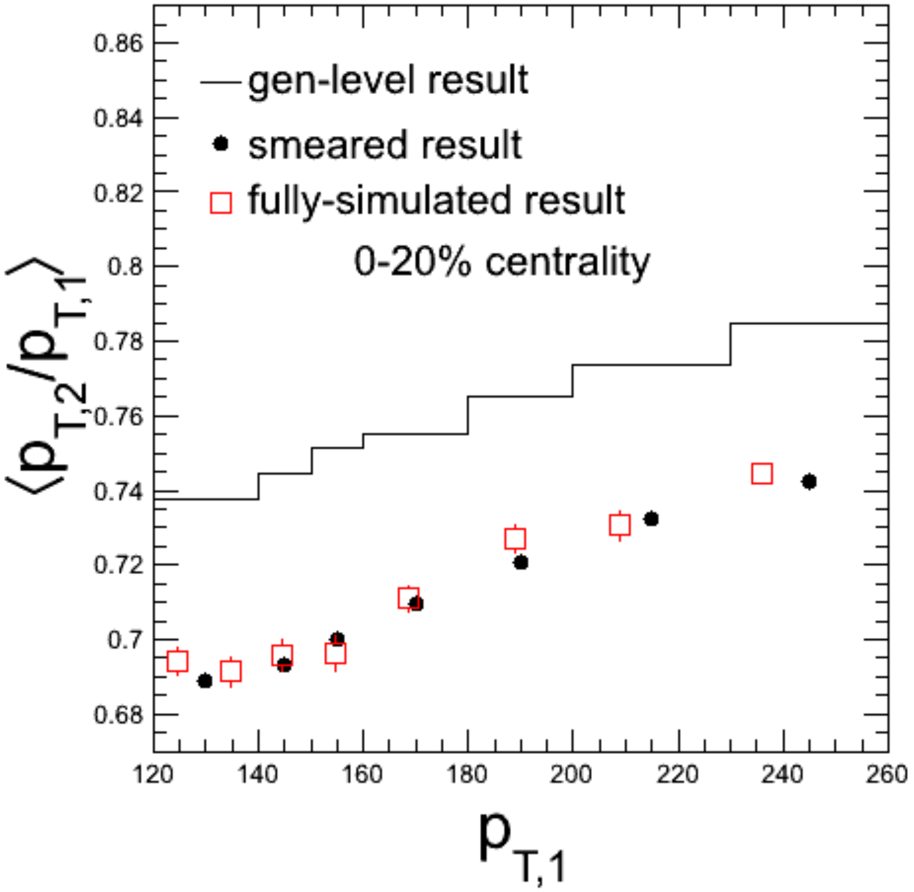
Swapped leading and subleading jet

Smearing effects in model comparison

pp & peripheral



central



b-jet results from CMS

Identification of b-quark jets with the CMS experiment

CMS Physics Analysis Summary BTV-11-004

Inclusive b-jet production in pp collisions at $\sqrt{s} = 7$ TeV

JHEP 1204 (2012) 084, **arXiv:1202.4617**

Measurement of $B\bar{B}$ Angular Correlations based on Secondary Vertex Reconstruction at $\sqrt{s} = 7$ TeV

JHEP 1103 (2011) 136, **arXiv:1102.3194**

Measurement of the b-jet to inclusive jet ratio in PbPb and pp collisions at $\sqrt{s_{NN}} = 2.76$ TeV with the CMS detector

CMS Physics Analysis Summary HIN-12-003

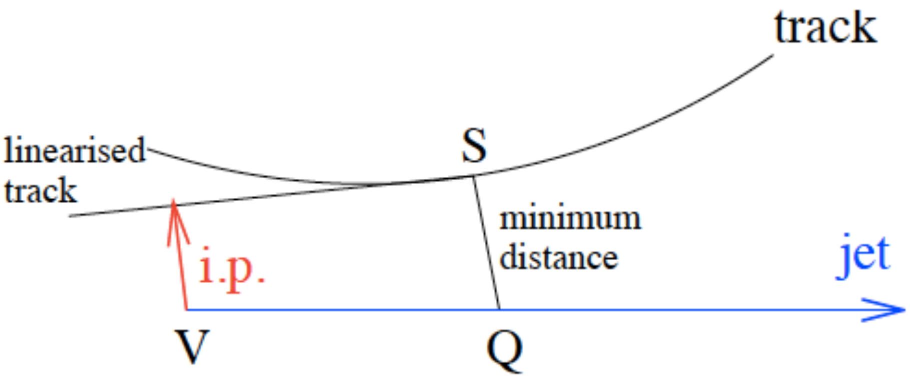
b-jet identification

Long lifetime of b (~1.5 ps) leads to measurable (mm or cm) displaced secondary vertices (SV)

Subsequent charm decay may lead to a tertiary vertex

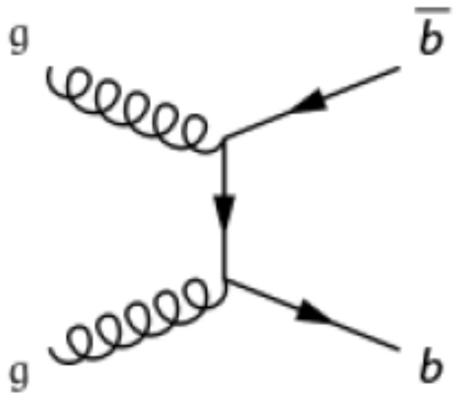
Several classes of b-jet taggers using:

- Reconstructed SV's, employing discriminating variables such as SV mass, flight distance, etc.
- Impact parameter (IP) of tracks associated to the jet, w/o requiring a reco'd SV
- Muons in jets, exploiting the large branching ratio (20%)

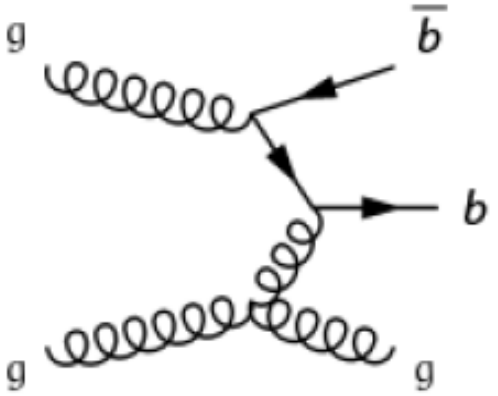


bottom production

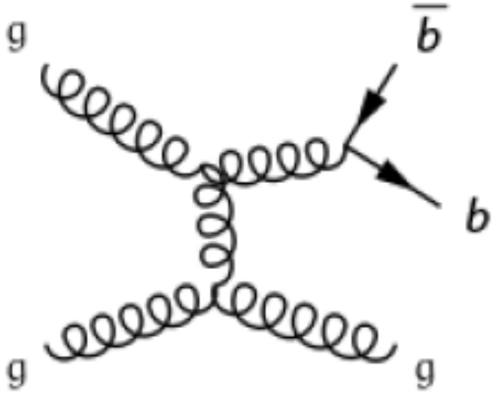
Flavor Creation (FCR)



Flavor Excitation (FEX)

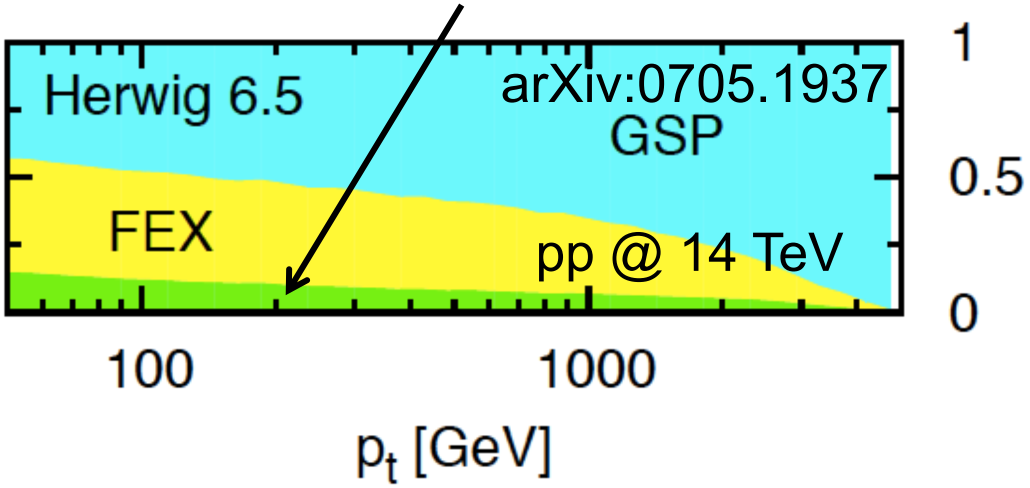


Gluon Splitting (GSP)



- LO b-$\bar{b}$ production (FCR) not dominant at the LHC

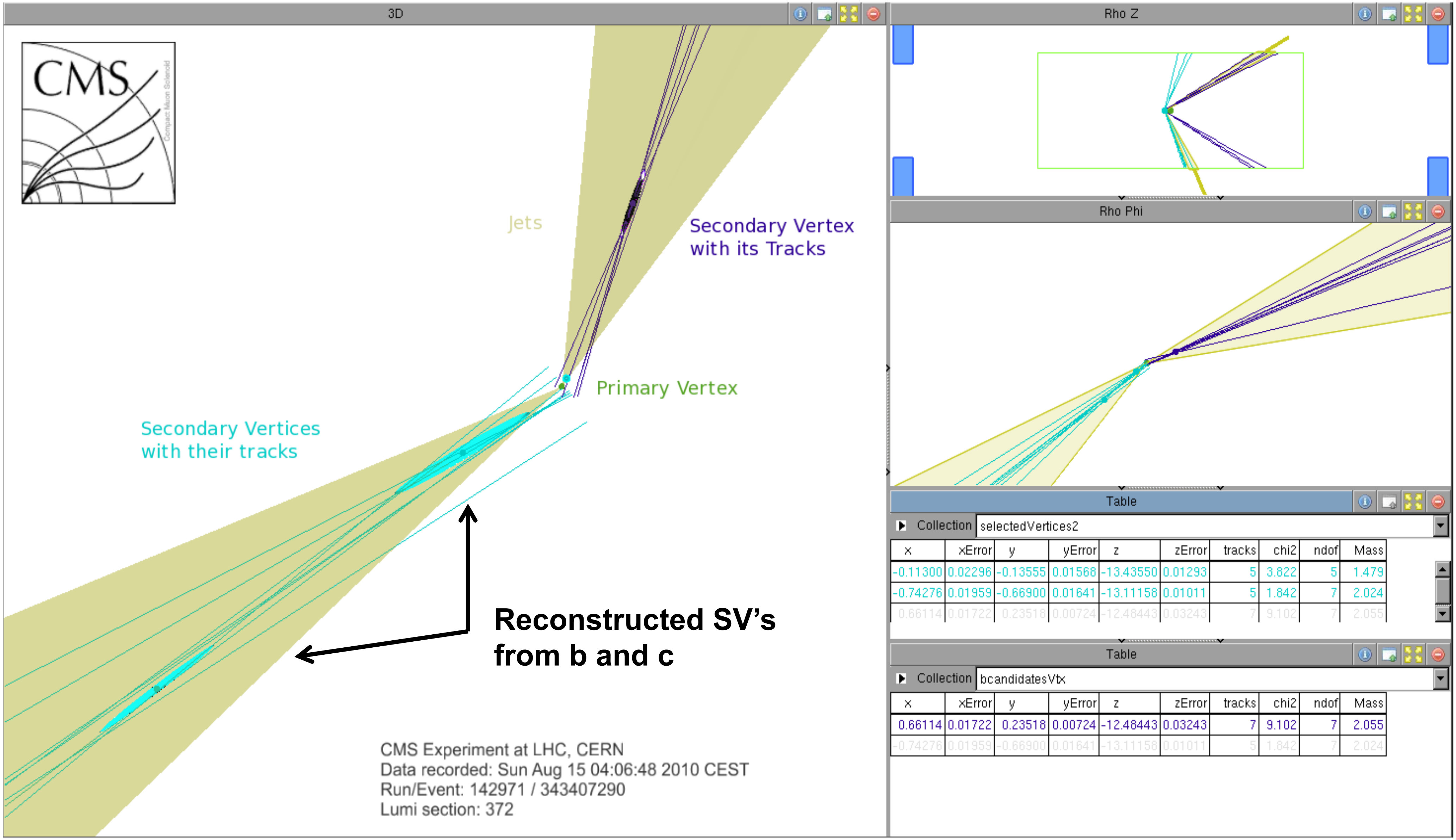
At NLO



Excitation of sea quarks → b(b) + light dijet, w/ b(b) at beam rapidity

Gluon splitting into b and $\bar{b}$ which can be reconstructed as a single jet

flavor creation candidate (pp @ 7 TeV)



Reconstructed SV's from b and c

CMS Experiment at LHC, CERN
Data recorded: Sun Aug 15 04:06:48 2010 CEST
Run/Event: 142971 / 343407290
Lumi section: 372

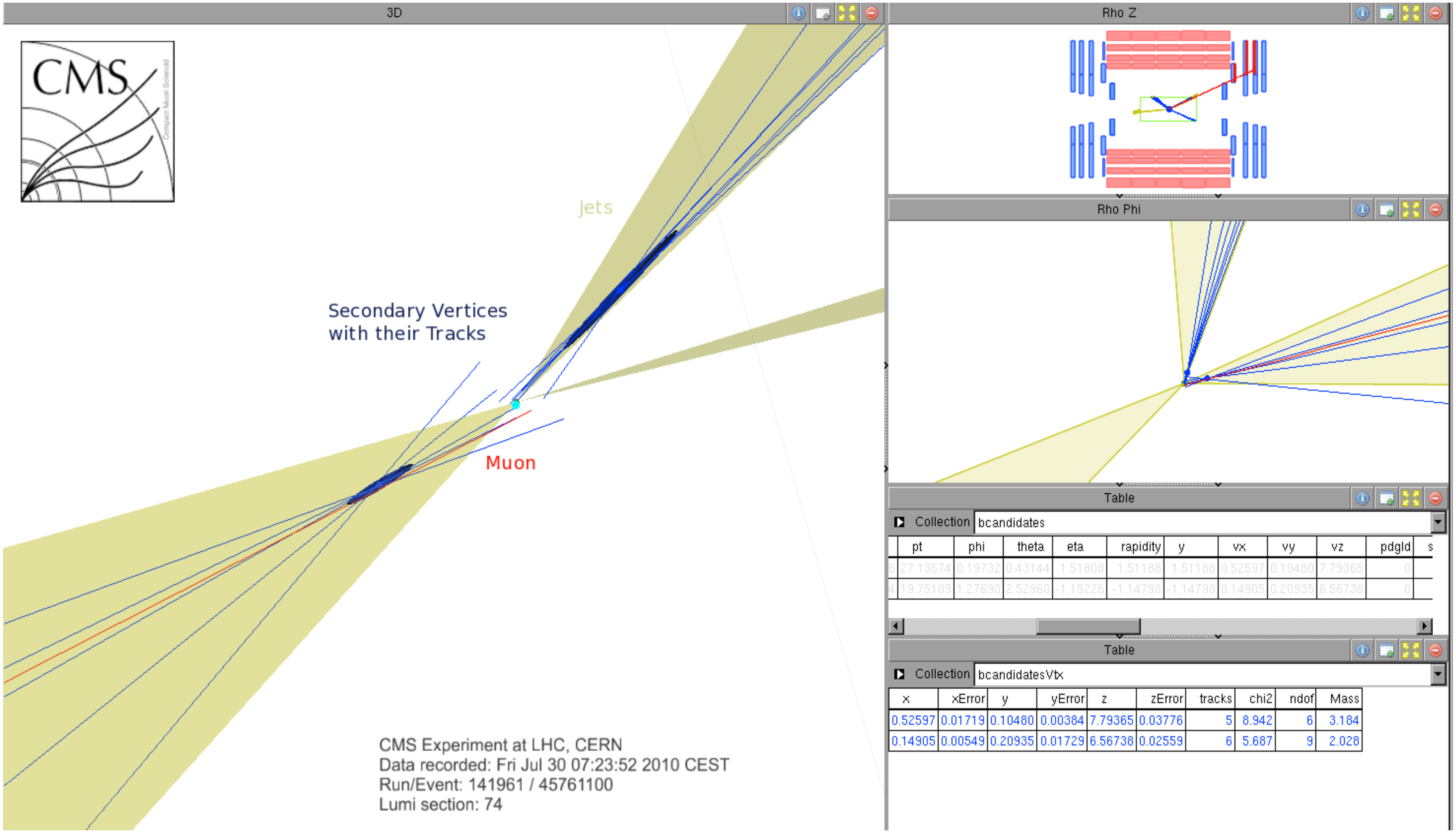
Collection selectedVertices2

x	xError	y	yError	z	zError	tracks	chi2	ndof	Mass
-0.11300	0.02296	-0.13555	0.01568	-13.43550	0.01293	5	3.822	5	1.479
-0.74276	0.01959	-0.66900	0.01641	-13.11158	0.01011	5	1.842	7	2.024
0.66114	0.01722	0.23518	0.00724	-12.48443	0.03243	7	9.102	7	2.055

Collection bcandidatesVtx

x	xError	y	yError	z	zError	tracks	chi2	ndof	Mass
0.66114	0.01722	0.23518	0.00724	-12.48443	0.03243	7	9.102	7	2.055
-0.74276	0.01959	-0.66900	0.01641	-13.11158	0.01011	5	1.842	7	2.024

gluon splitting candidate (pp @ 7 TeV)



CMS Experiment at LHC, CERN
Data recorded: Fri Jul 30 07:23:52 2010 CEST
Run/Event: 141961 / 45761100
Lumi section: 74

Collection: bcandidates

	pt	phi	theta	eta	rapidity	y	vx	vy	vz	pdgId	s
6	27.13574	0.19732	0.43144	1.51808	1.51188	1.51188	0.52597	0.10480	7.79365	0	
4	19.75109	1.27690	2.52960	-1.15226	-1.14798	-1.14798	0.14905	0.20935	6.56738	0	

Collection: bcandidatesVtx

x	xError	y	yError	z	zError	tracks	chi2	ndof	Mass
0.52597	0.01719	0.10480	0.00384	7.79365	0.03776	5	8.942	6	3.184
0.14905	0.00549	0.20935	0.01729	6.56738	0.02559	6	5.687	9	2.028

Some answers to Nestor's questions

- Can we agree in a common jet definition in terms of algorithm (with radius and merging scheme) and particle content a la Les Houches accords? That it could be the same in pp, pPb and PbPb? It would help to compare the experimental results and to understand the differences due to detector effects, charged/total,... Even if these definitions are not the ones used for the final results for all experiments, having some 'comparable' results would be most profitable.

- Anti-kT already seems to be serving as a common merging scheme which many experiments are satisfied with.
CMS has used R = 0.2, 0.3, 0.4, 0.5

- For two reasons, it may be difficult to fix a single R:
Different R values may be probing different physics
Different measurements, in different types of environment, may drive the experiment to prefer a specific R

- It is desirable to have all results for all R values as mentioned, but on the practical side, the more algorithms there are, the bigger the commissioning phase for jet reconstruction for an analysis gets

Some answers to Nestor’s questions

- (a) Should we theoreticians simply provide the quenching code to the experiments and they extract the medium parameters running it?; (b) can we do it ourselves with some model background?; (c) both?

- c, both:

- First, b: If the experiment’s results are not unfolded for reconstruction effects, the experiment should provide the performance of the reconstruction (and a clear prescription of how to account for this in modeling) so that every theoretician can test the result of the model against reconstructed-level data

- Also, a: It is of huge desire to have various quenching models, particularly ones representing the fragmentation in a realistic manner, so that the effects on the experimental jet response can be tested further, helping to reduce uncertainties

Some answers to Nestor’s questions

- If (b,c), how can we validate our background? Would it be possible that all collaborations, irrespective of how they finally do the subtraction/reconstruction, provide some numbers (for example, \rho, \sigma_{jets} as provided by some fixed version of FastJet) that we can use to validate our model for the background? I understand that those numbers may depend on experimental cuts, particle flow method or not, decays included or not,... Can we get an agreement? For example, if all experiments get in some region all charged particles with pt>ptmin, could all collaborations provide some numbers for charged with pt>ptmin in that region?

- This is a good motivation for the experiments to measure dN/deta and v2 results, which can be treated as a validation of the background.

- As already mentioned, rho, sigma are highly dependent on detector properties and reconstruction method. This is also the case for the “random cone” energies. It is difficult to compare these across experiments, however these are good validation tools for the simulations within a specific experiment

Some answers to Nestor’s questions

- Which observables look more promising for extracting medium parameters i.e. which ones are those for which background effect seems smaller? How should be proceed to propose new ones?

- The search for better background estimation/subtraction is still ongoing.

The goal is to reduce fragmentation and flow vulnerability.
A “golden” subtraction method is not commissioned so far.