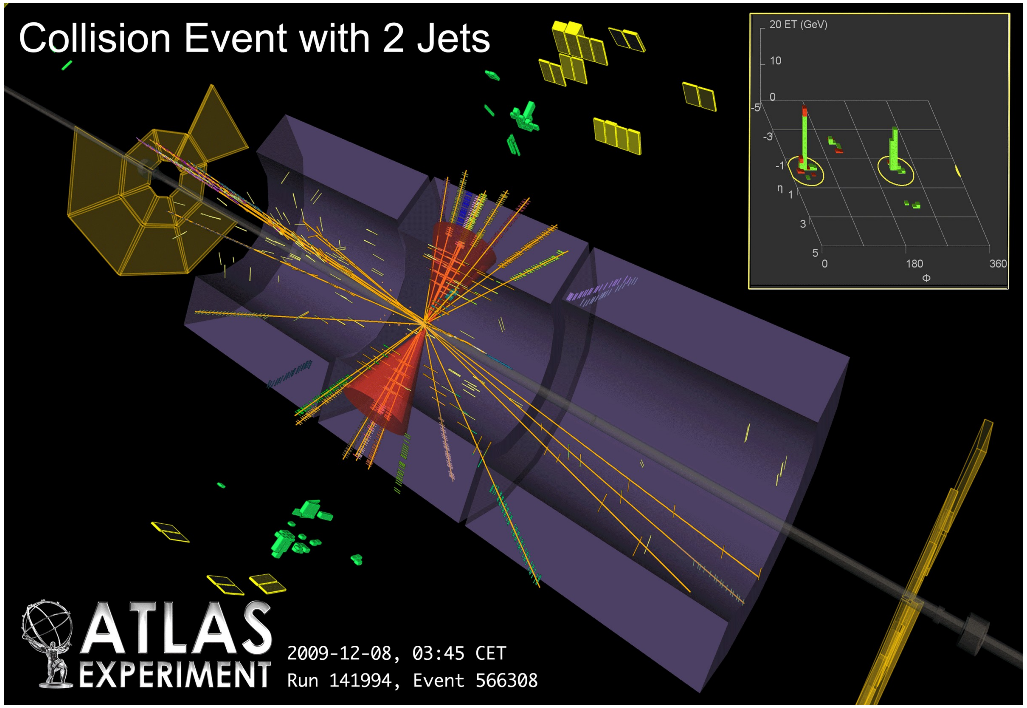
Jet reconstruction

- What are jets?
- How can we relate the measured jets to the underlying physics process?
- Calibration of jets
- Tagging of jets

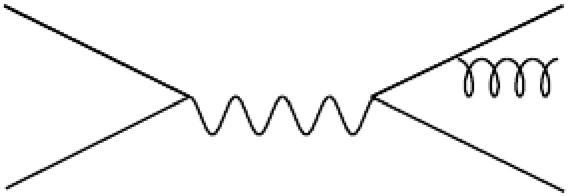


What are jets?

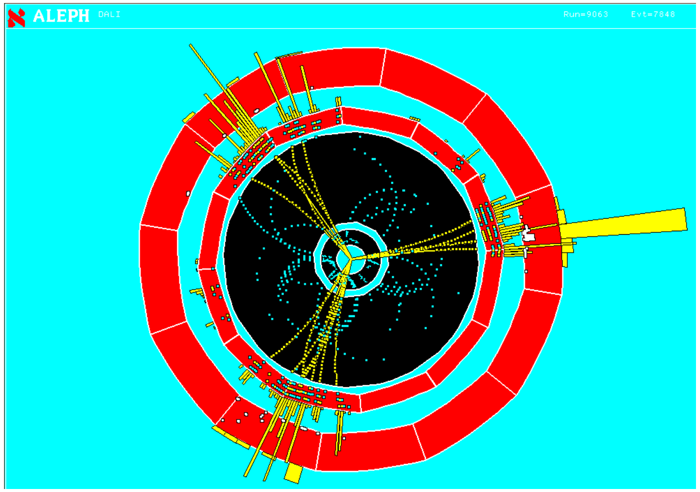
Jets for non- particle physicists

Jets for theorists

- jets of partons: gluons, quarks



ALEPH: e+e- → 3 jets



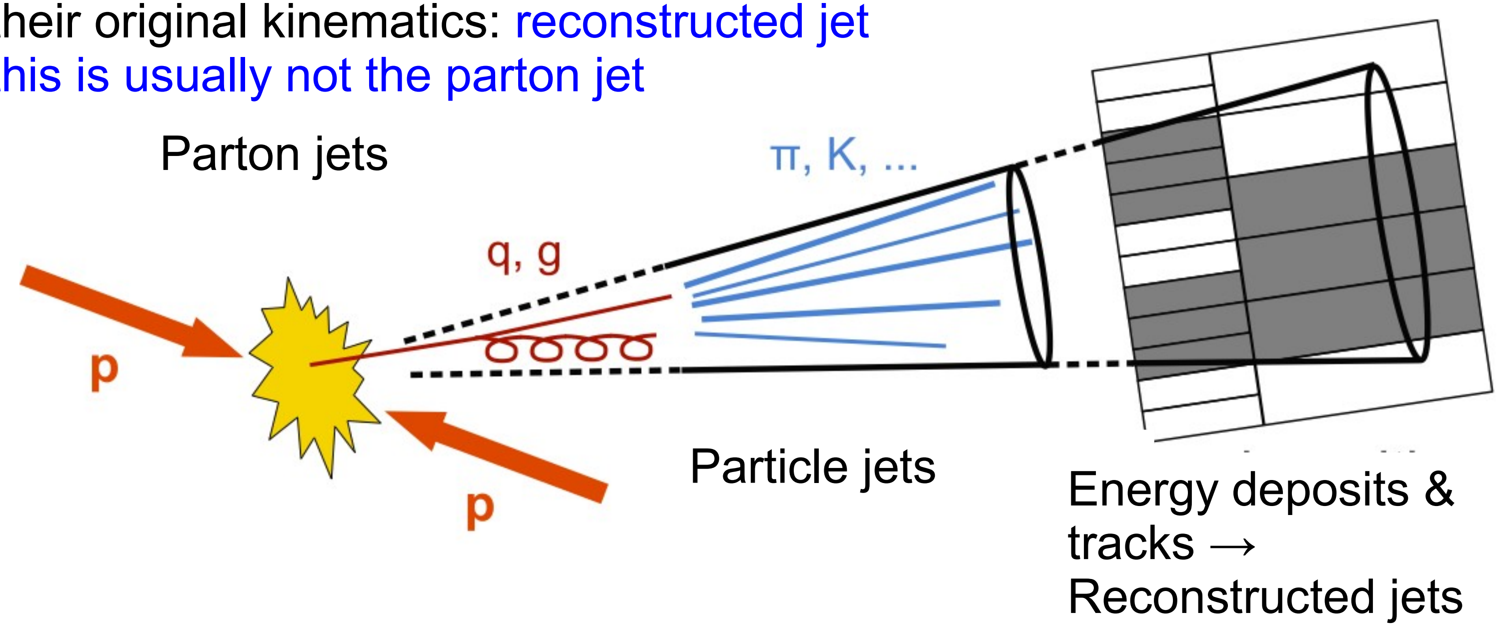
Jets for experimentalists

- bunch of particles generated by hadronization of a common source: quark, gluon fragmentation → the particles in this bunch have correlated kinematic properties
- observables in the detector: protons, neutrons, pions, photons, electrons, muons, plus other particles with lifetime > 10 ps
- non-interacting particles do not generate a signal – mostly neutrinos

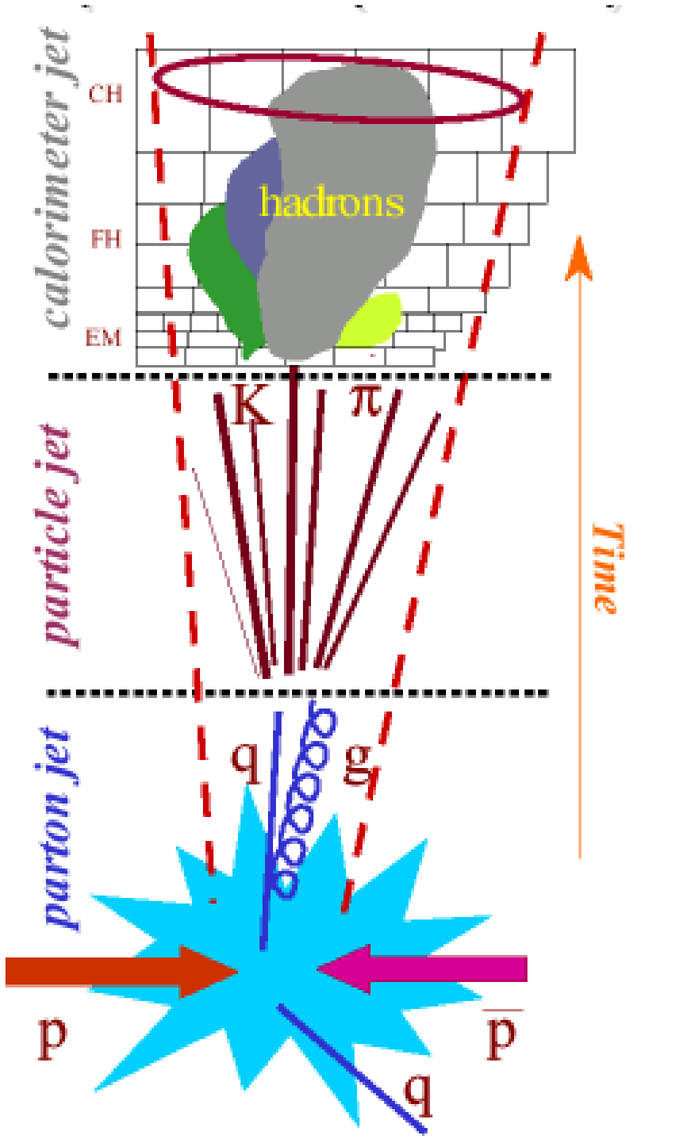
What are jets?

Jet reconstruction

- Theory: cluster partons (often a jet with just one parton): parton jets
- Model/Simulation: collect the final state particles (p, n, γ, π, μ, e and other particles with lifetime > 10 ps) into objects (jets) representing the original parton kinematic, re-establishing the correlations: particle jet
- Experiment: attempt to collect the detector signals from these particles to measure their original kinematics: reconstructed jet
  this is usually not the parton jet



Jet in an experiment



Jet Reconstruction Challenges

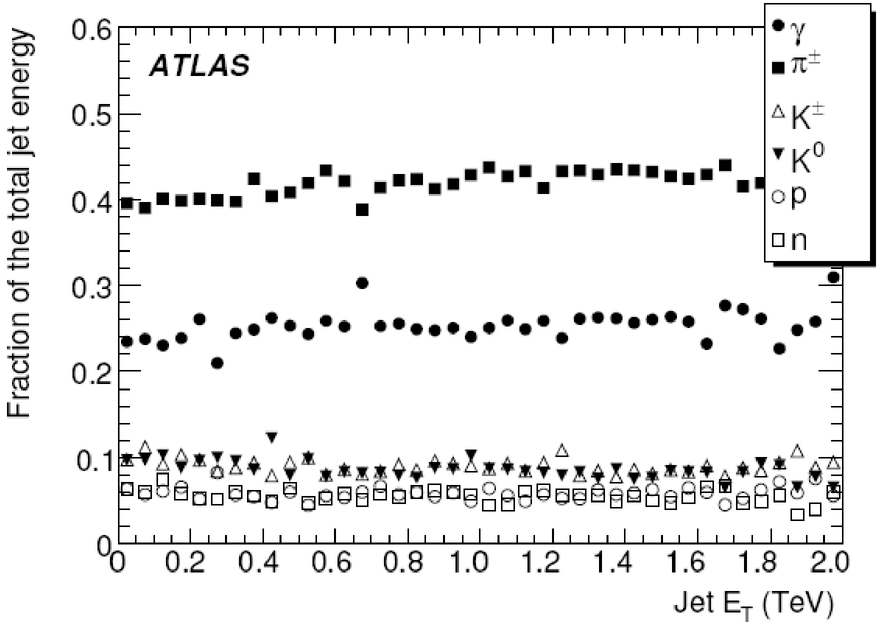
- longitudinal energy leakage
- detector signal inefficiencies (dead channels, HV…)
- pile-up noise from (off- and in-time) bunch crossings
- electronic noise
- signal definition in calorimeter(clustering, noise suppression…)
- dead material losses (front, cracks, transitions…)
- detector response characteristics (e/h ≠ 1)
- jet reconstruction algorithm efficiency
- lost soft tracks due to magnetic field
- tracks from underlying event
- tracks from pile-up events (more than one collision)
- jet reconstruction algorithm efficiency

What are jets?

Jets for experimentalists

- observables in the detector: protons, neutrons, pions, photons, electrons, muons, other particles with lifetime > 10 ps
- non-interacting particles do not generate a signal – mostly neutrinos

Fraction of total jet energy mostly carried by pions ($\pi^{\pm}$ or $\pi^{0} \rightarrow \gamma\gamma$)



Jets – why bother?

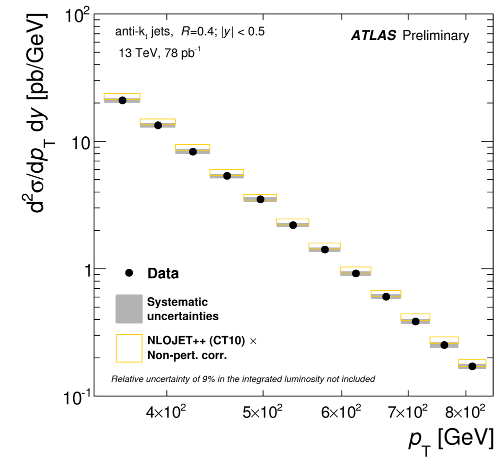
Jets are no clear observables as electrons, muons, missing energy …
but they are very important ingredients

SM Physics with jets
- jet production (X+jets), sensitive to strong coupling constant
- test of QCD at large energy scales

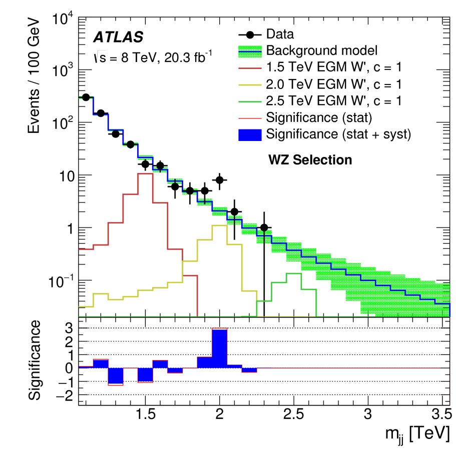
SM Physics of jets
- jet structure – find sub-jets
- test of QCD at small energy scales

QCD and jets are the key to New Physics
- new physics is likely to be born in a QCD process
- new physics often results in jets in final states
- most of the time, QCD is the major background

inclusive jet x-section at 13 TeV
ATLAS-CONF-2015-034

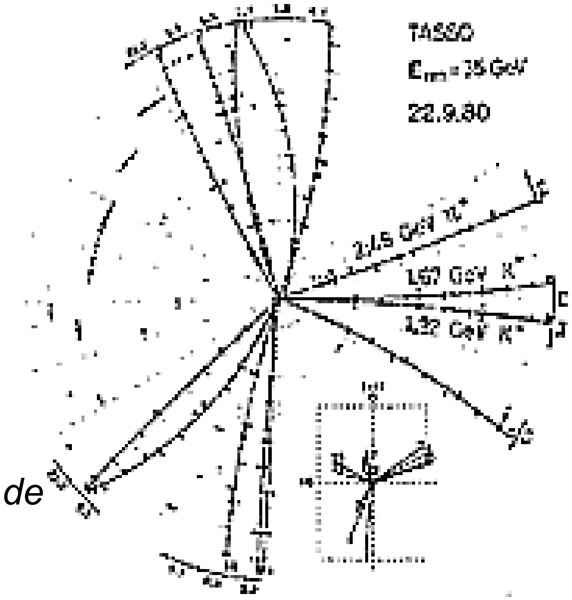


search for high mass boson
arXiv:1506.00962



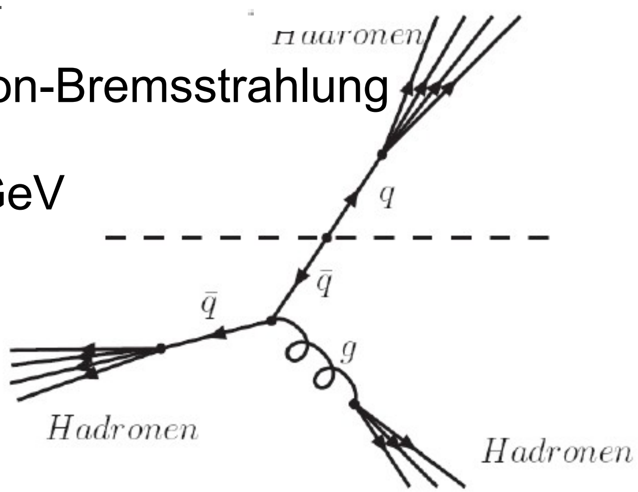
TASSO: discovery of the gluon

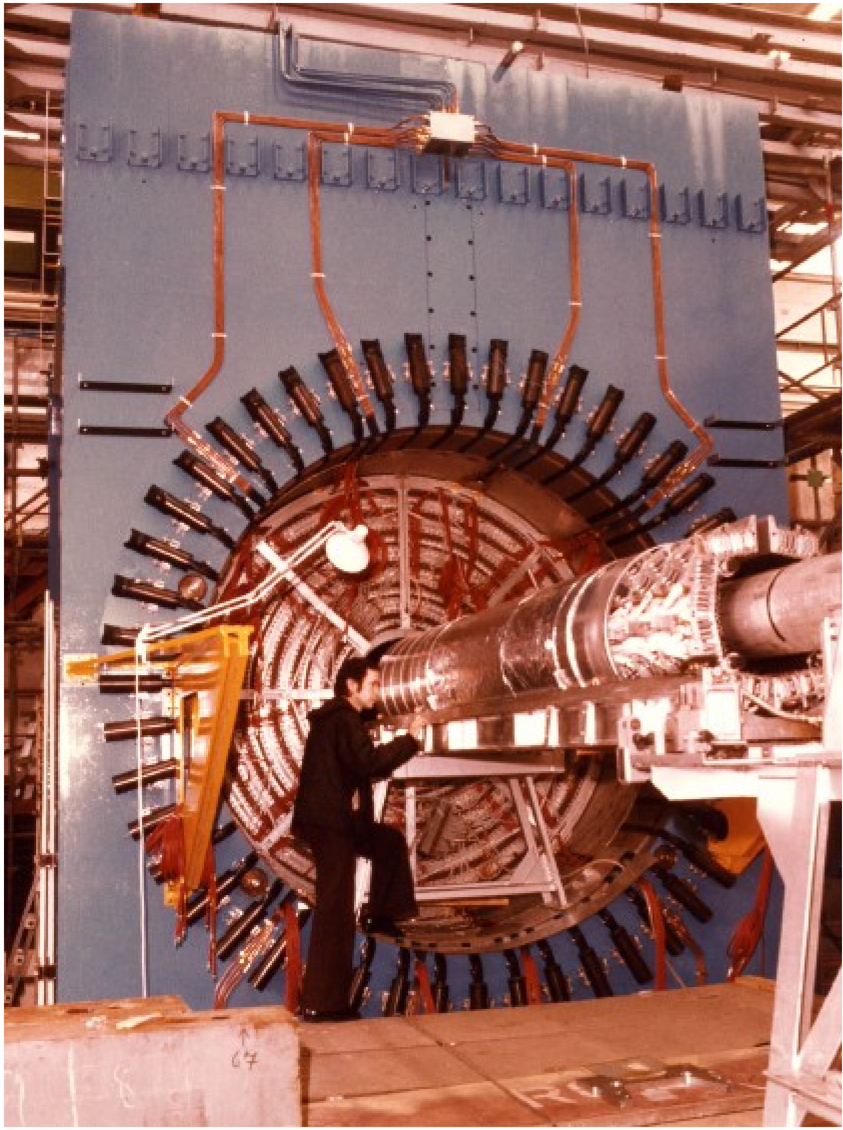
1979 observation: 3-jet event measured as TASSO



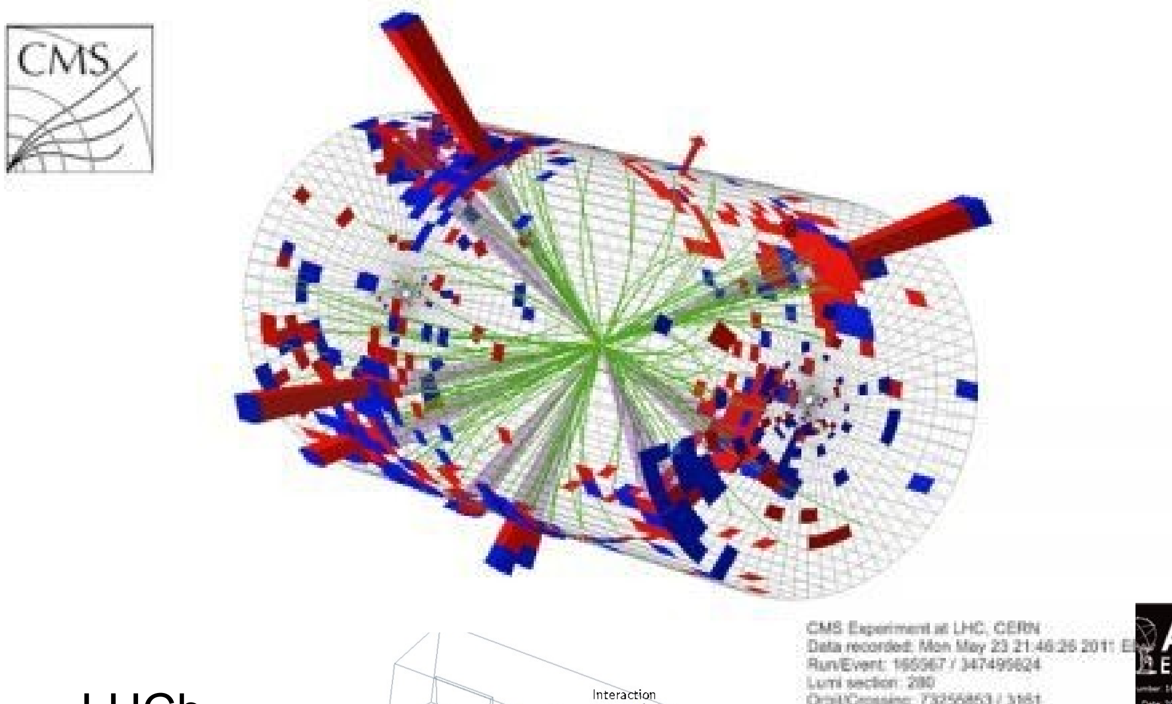
www.desy.de

Interpretation: Gluon-Bremsstrahlung
e+e- → q qbar g
cms energy 27.4 GeV



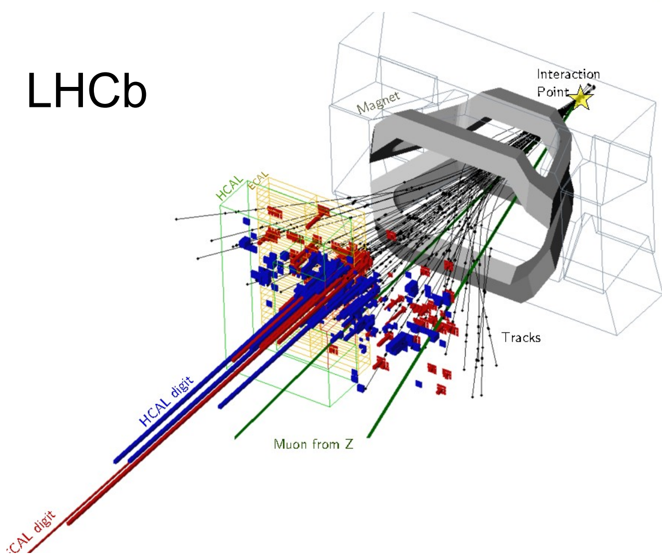


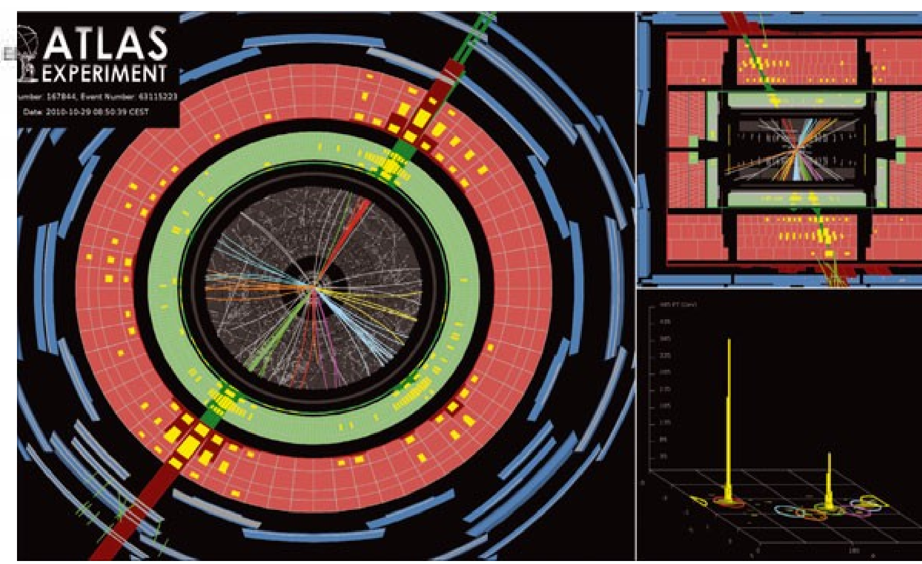
Jets at LHC



Jets:
collimated, energetic bunches of particles

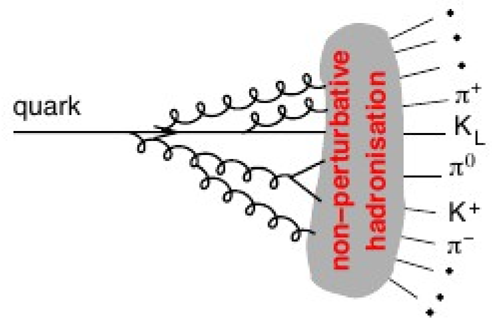
LHCb



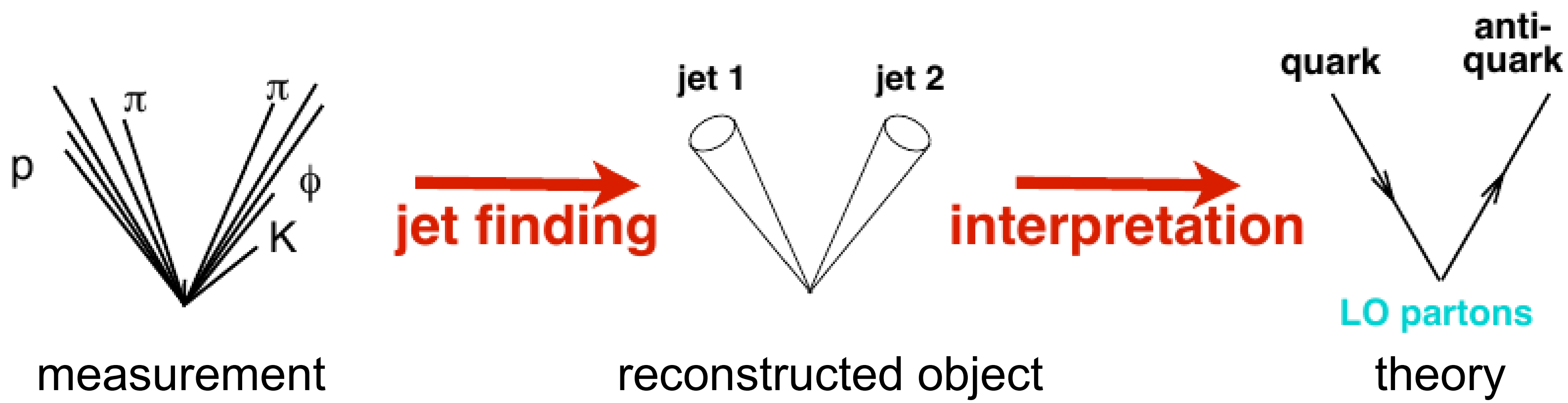


Key aspects

Jets get their structure through fragmentation and hadronisation



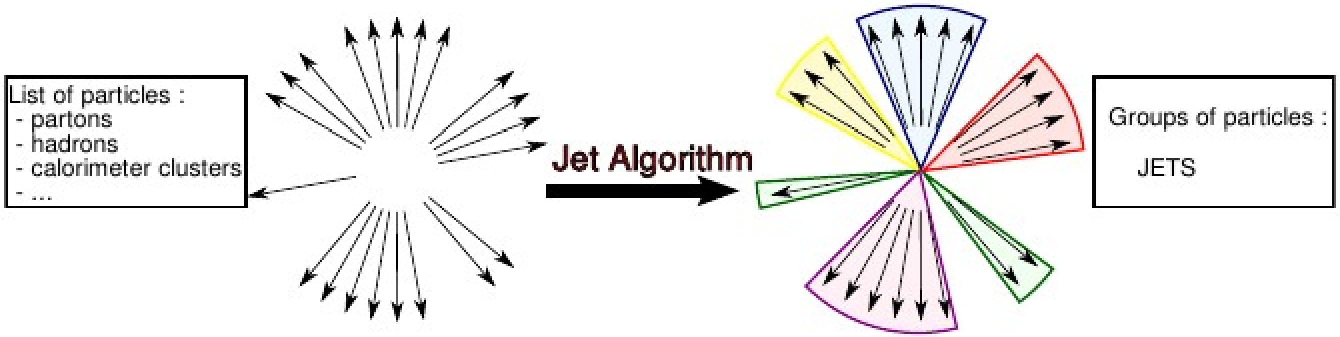
How do we reconstruct jets – correlation to underlying physics



Different levels of jets

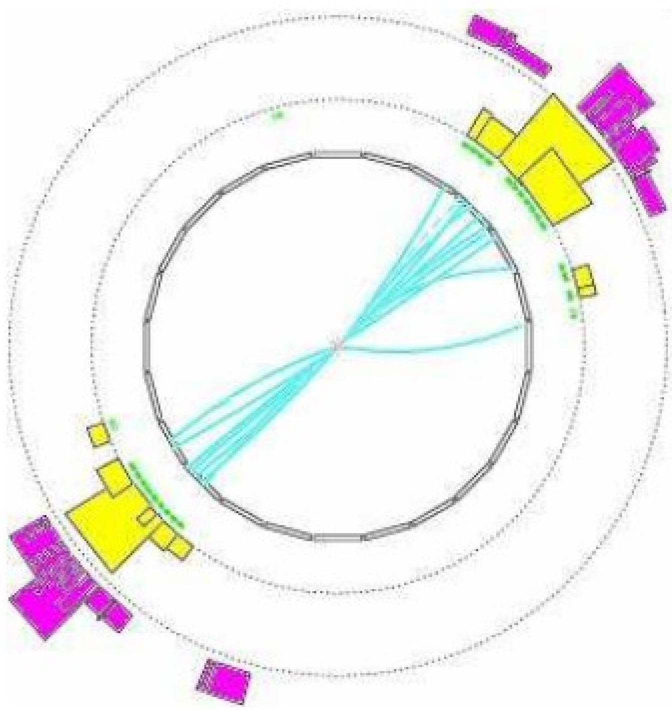
Levels of jets

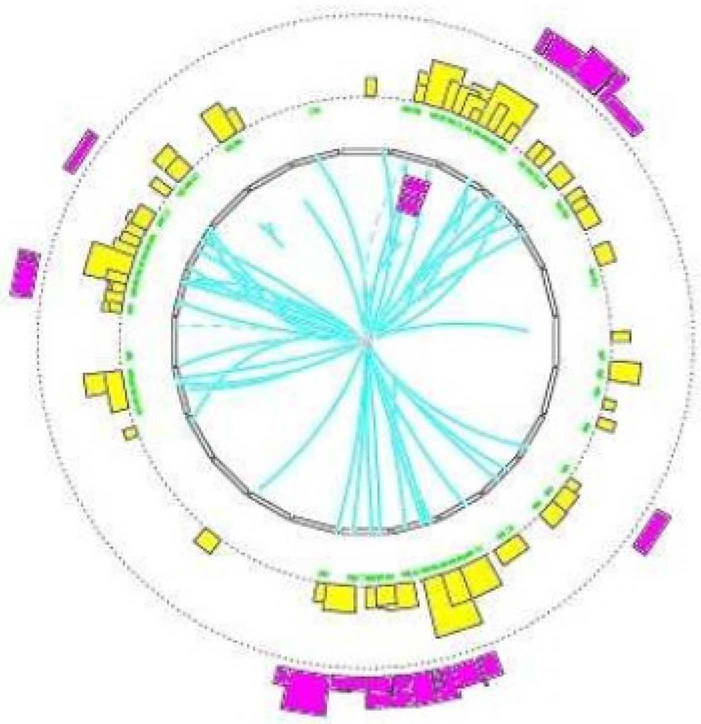
- parton jets – clean definition QCD perturbative
- hadron jets contains objects with low transverse momentum, non perturbative QCD
- reconstructed (experimental) jets – thresholds, noise ..
- ultimate goal: relate experimental to parton jets



- level independence: parton, hadron, experiment
- infrared & collinear safe

Seeing jets



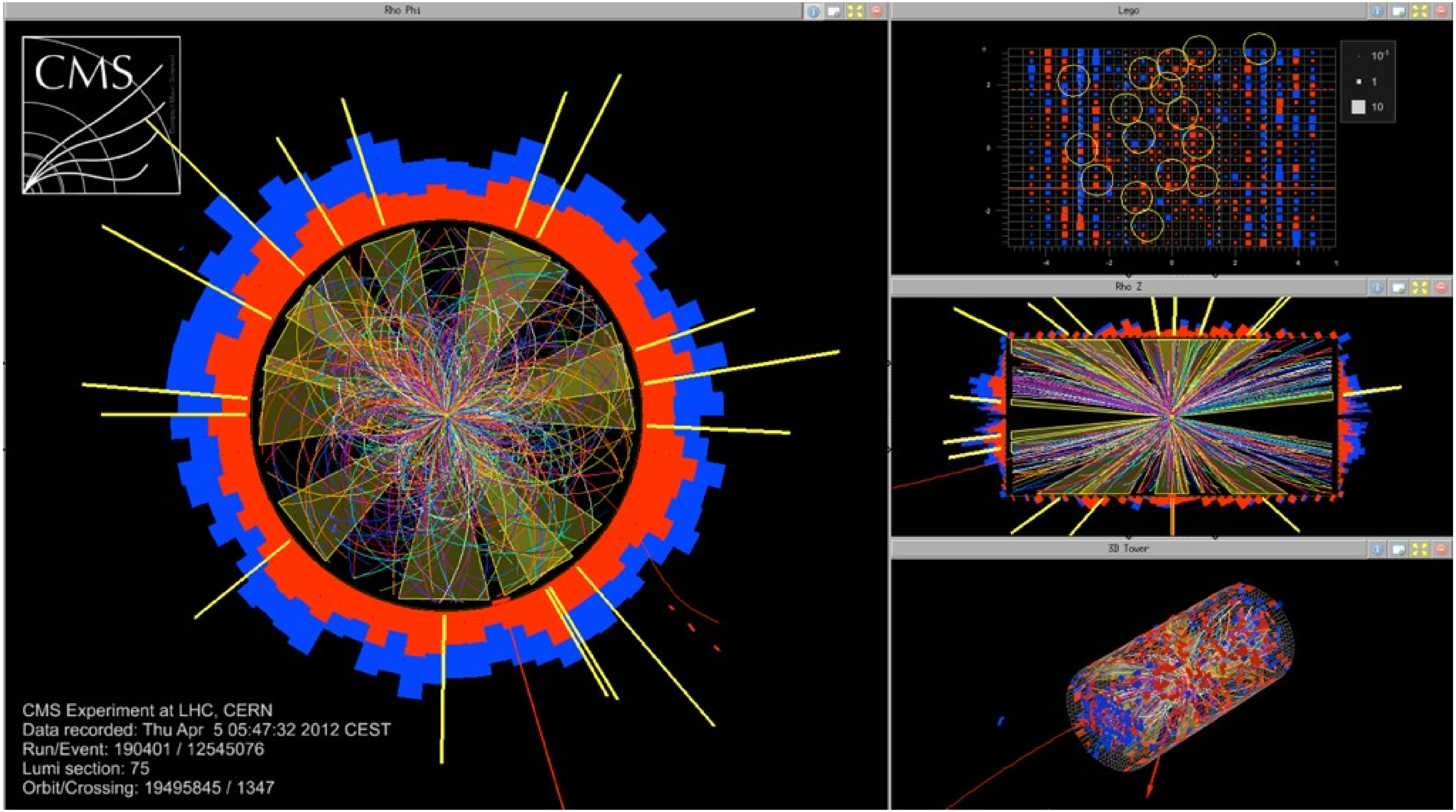


Clearly two jets

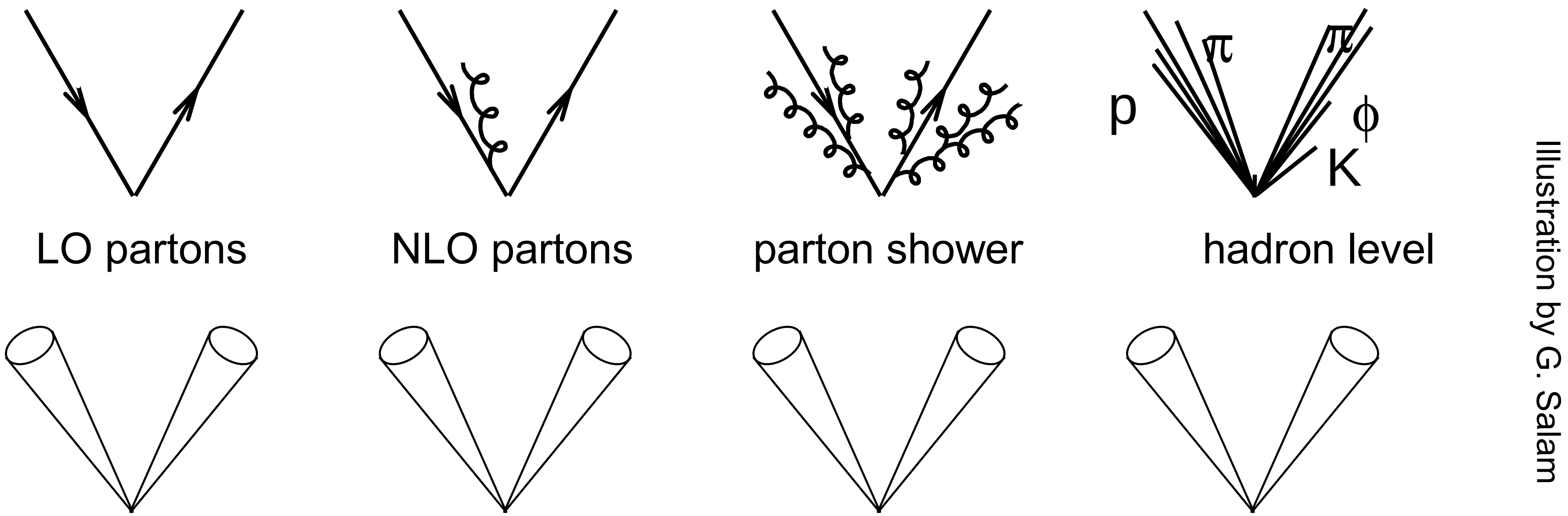
How many jets?

need to define clever algorithms to use the jet information

how many jets??



Jet definition: validity



In order to allow comparison to theory, jet definition should not be sensitive to soft effects from hadronisation and energy loss

Important: jet definition should not depend on QCD effects
→ jets should be collinear safe and infrared safe

Jet algorithm

jet reconstruction is unavoidably ambiguous and highly complex

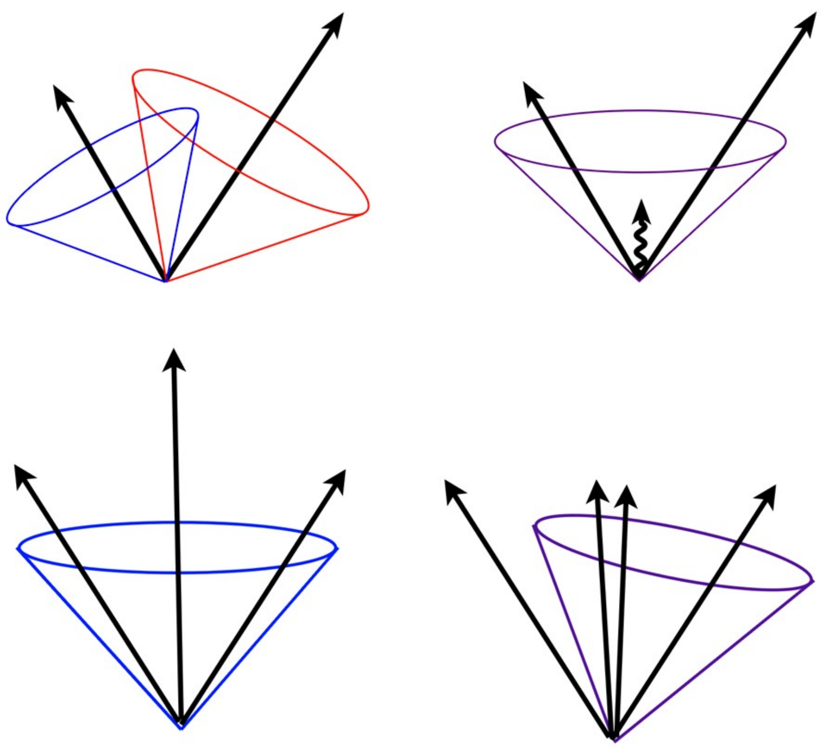
1. which particles get put together into a common jet?
   - tracks
   - energy clusters
   - tracks and energy clusters
   - particle flow algorithm
2. choice of jet algorithm and parameters
   - size of jet, seed definition
3. how do you combine the momenta of the particles?
   - recombination scheme
   - most commonly used: direct 4-vector sums (E -scheme)

1-3 define the jet

physics results (particle discovery, masses, PDFs, coupling) should be independent of your choice of jet definition

Infrared and collinear safety

jet definition is ambiguous but
jets should be invariant with respect to certain modifications of the event:
infrared and collinear safe



infrared safe:
configuration must not change when adding a further soft particle

infrared unsafe:
after emission of soft gluon jets are merged: 2jets → 1 jet

collinear safe:
configuration does not change when substituting one particle with two collinear particles

examples: signal split into two towers
decay $\pi^0 \to \gamma\gamma$
collinear emission of a gluon
→ if jet energy and/or direction change algorithm is collinear unsafe

Jet definition: collinear and infrared safe (IRC)

For an observable’s distribution to be calculable in fixed-order perturbation theory, the observable should be infrared safe, i.e.

→ insensitive to the emission of soft or collinear ($\theta \rightarrow 0$) gluons .
→ insensitive to splitting of signals

this guarantees the cancellation between real and virtual emission diagrams of the infrared and collinear divergences in every order of pQCD

Examples:

- multiplicities of gluons: not IRC – modified by soft and collinear splitting
- energy of hardest particle: not IRC – modified by collinear splitting
- energy flow into a cone is IRC – soft emissions don't change energy flow
  collinear emissions don't change the direction

Jet algorithm guidelines

- detector technology independence
- jet efficiency should not depend on detector technology

  → final jet calibration and corrections ideally unfolds all detector effects

- minimal contribution from spatial and energy resolution to reconstructed jet kinematics
  unavoidable intrinsic detector limitations set limits

- stability within environment
  detector noise should not affect jet reconstruction within reasonable limits
  energy resolution limitation
  avoid energy scale shift due to noise
  stability with changing (instantaneous) luminosity at LHC
  control of underlying event and pile-up signal contribution

- "easy" to calibrate and high reconstruction efficiency
  identify all physically interesting jets from energetic partons in perturbative QCD
  jet reconstruction in resonance decays
  → high efficiency to separate close-by jets from same particle decay

Jet families

Two categories of jet algorithms:

1) Cone jets (Cone, SisCone, MidCone) traditionally for hadron colliders
top-down approach, simple geometrical picture
- simple geometrical motivation: draw cone radius R around starting point
- find stable cones around seed, stable: jet 4-momentum along cone axis
- split/merge jets: procedure for overlapping cones

2) Clustering: sequential recombination (Jade, k_T, anti-k_T) traditionally e+e-
bottom up approach, try to undo fragmentation and hadronisation
- proto-jets
- successive pairwise combination of proto-jets
- define distance between constituents
- combine first closest constituent

Sum particles inside jet: different prescriptions, most natural: sum 4-vectors

A jet is an object defined by an algorithm. If parameters are right it may approximate a parton

For a complete discussion, see: http://www.lpthe.jussieu.fr/~salam/teaching/PhD-courses.html

Cone algorithms

Extremely slow if no seeds are used

Iterative cone algorithm

1) start with p_T ordered list of objects
   2) choose first object (highest p_T) as seed
   3) collect all objects within a cone of radius R around seed, typical R=0.5, 0.7
      4) recalculate jet axis and use it as new seed
      5) repeat from 3) until stable axis
      there might be no stable solution!
   6) declare constituents as a jet and remove them from the input list
   7) repeat from 2) until list empty

faster: apply threshold on input list objects (save computing time and reduce noise)
→ algorithm is neither infrared nor collinear safe

split/merge: if two jets overlap compute fraction of p_T in overlap region
split if f > fcut (typically f>0.75)

Seeded algorithm

CMS: iterative cone with removal

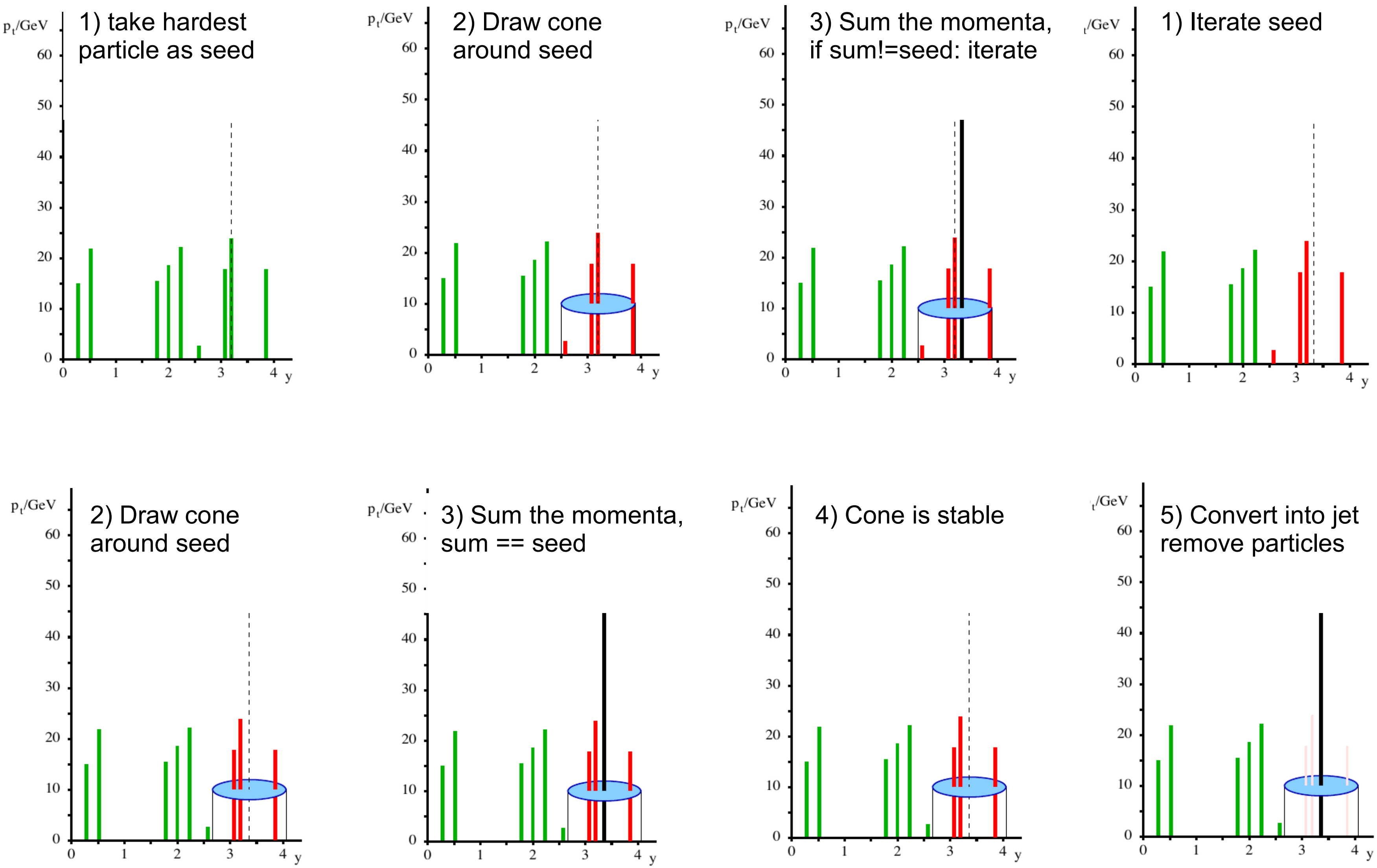
- remove objects assigned to the cone from list
- find next surviving seed and new cone
- stop if no more seeds or stable cones

ATLAS: overlapping cones (ATLAS)

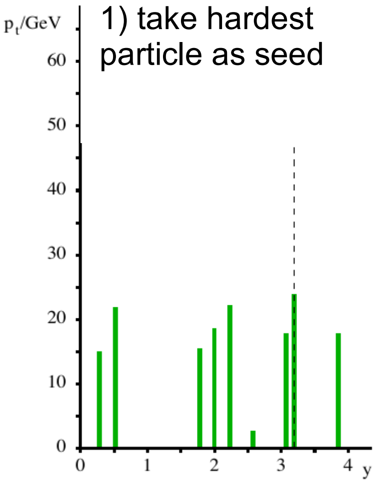
- find next seed and new cone until all seeds are exhausted
- apply split and merge procedure
- split and merge
  - merge jets if p_T of overlapping constituents above fraction f of p_T of lower p_T jet f = 50% in ATLAS
  - otherwise split lower p_T jet
- assign split constituents to higher p_T jet

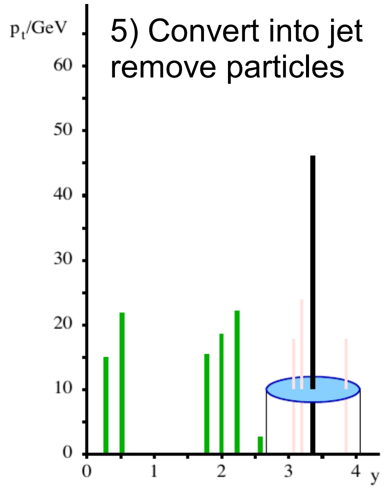
both algorithms are not collinear and infrared safe
variations: e.g., Mid-point cone at Tevatron, SisCone

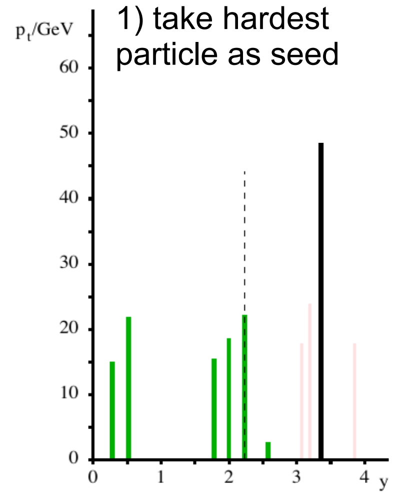
Example: CMS iterative cone with removal

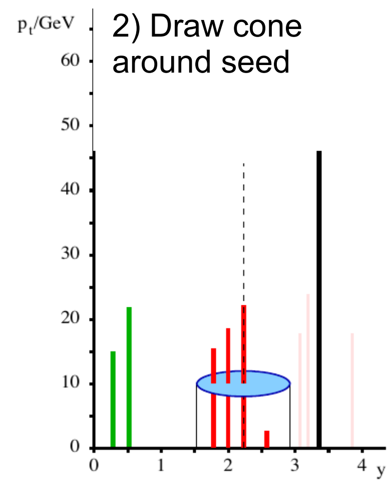


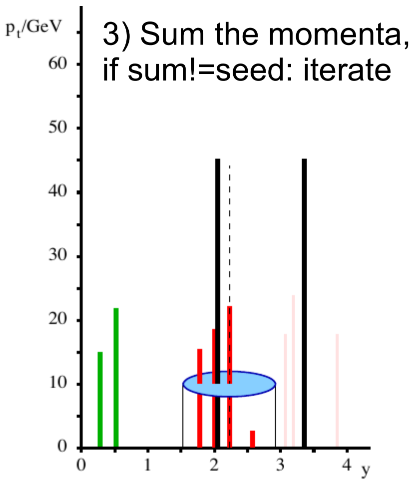
Example: CMS iterative cone with removal



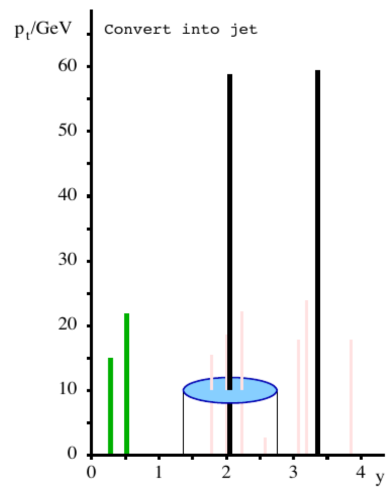


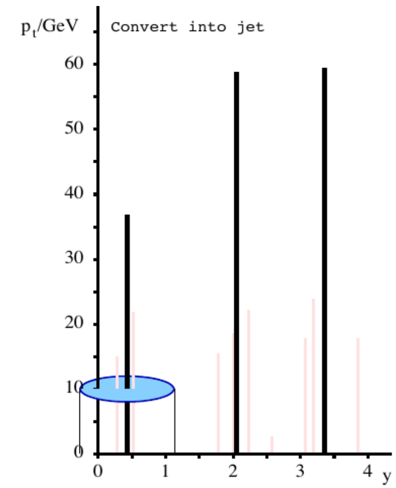






iteration until stable



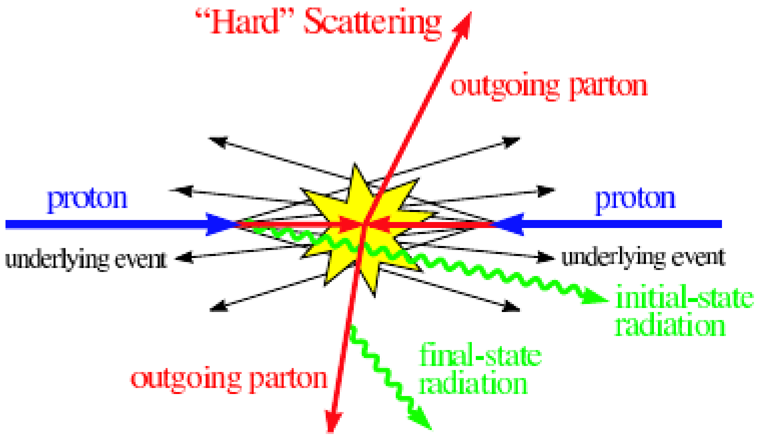


but choice of hardest particle is collinear unsafe
there is a work around SISCone (Seedless cone algorithm)

Features of cone algorithm

– low energy hadrons are not all included in jets
  → energy missing for event reconstruction
+ large part of underlying event/pileup rejected
+ jet shapes are usually round → corrections for detector effects easier

corrections (for all jet algorithms)
- efficiencies
- acceptance (detector edge effects)
- pileup (multiple primary collisions)
- underlying event
- noise



underlying event: everything except the products of the hard collision

Midpoint cone algorithms

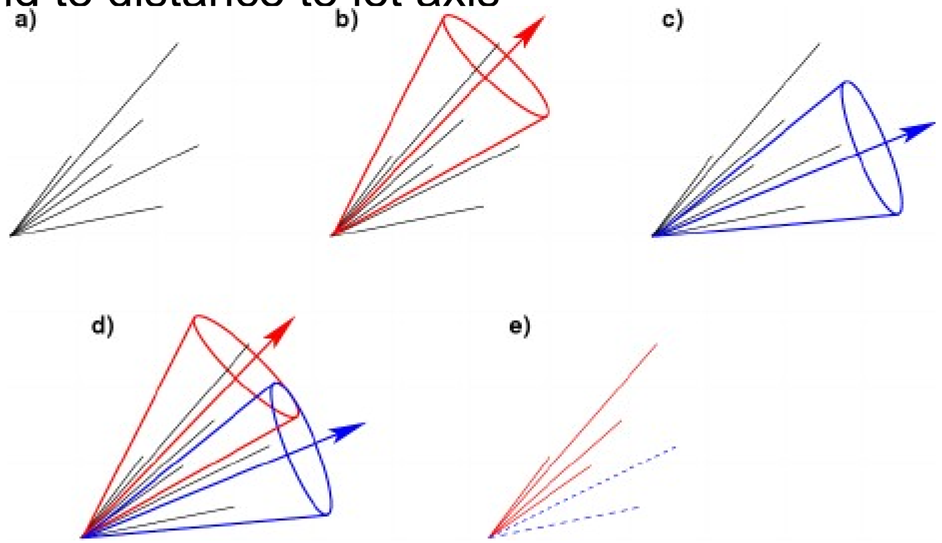
similar to iterative cone, but:

objects assigned to proto-jets are not removed from input list
(overlapping proto-jets are possible)

for each pair of proto-jets closer than R the midpoint is used as additional seed

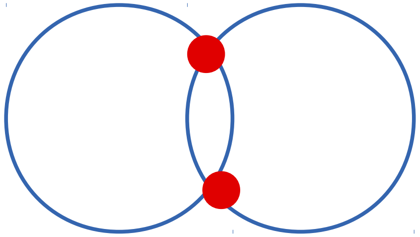
overlapping energy of 2 proto-jets is larger than 50% of smaller one
→ merge proto-jets
→ else: split overlapping constituents according to distance to jet axis

+ midpoint cone is infrared safe up to NLO
– midpoint cone is not collinear safe
– shape is not circular



SIS Cone: seedless infrared-safe cone algorithm

1) find all stable cones with size R
   - for each two objects within 2R check the two cones with both objects on the circumference for stability
   - stable: all particles defining the cone axis have d<R
     all particles not belonging to cone have d>R



2) some clusters might not be stable due to nearby jets →
   remove all objects from stable jets and repeat search for stable cones

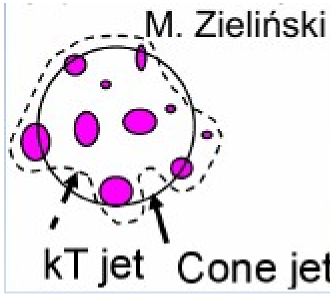
3) splitting and merging similar to Midpoint Cone (but: scalar sum pT ordering)

code:http://projects.hepforge.org/siscone/

+ only slightly higher execution time compared to Midpoint Cone
+ collinear and infrared safe as long as there are no thresholds
+ acceptable computing performance ($\sim N^2 \ln N$)
  previous algorithms: $\sim N2^N \rightarrow 10^{17}$ years for N=100
– jet shapes not circular

SIS Cone used at LHC for cross checks

Recursive algorithms (Jade, kT, anti-kt)



attempt to undo parton fragmentation:
iterative pairwise clustering to build larger objects

but: hadrons are colour singlets with integer charge
quarks and gluons are colour triplets/octets, quark charge 1/3, 2/3
→ quark/gluon "reconstruction" can never be exact

recursive algorithms:

- define distance measure dij for pair of particles (metric)
- define combination algorithm

combination algorithm

- calculate dij for all pairs and find dij,min
- if dij,min < dcut: recombine i and j
- if dij,min > dcut: i is a jet and removed from list
- repeat until all pairs have a distance > dcut

definition of the distance parameter gives general behavior of the jet algorithm
dcut is a resolution parameter, it defines whether jets are coarse or fine objects
in general: jet shape not circular

key ingredient: definition of distance measure

Jade – invariant mass: first algorithm

$d_{ij} = m_{ij}^2 \approx 2E_i E_j (1 - \cos\theta_{ij})$

kT/Durham algorithm – widely used at LEP (Catani, Dokshitzer, Olsson, Turnock & Webber 1991)

$d_{ij} = 2\min(E_i^2, E_j^2)(1 - \cos\theta_{ij})$

kT at hadron colliders:

un-seen beam remnants → introduce beam distance

$d_{iB} = p_{Ti}^2$ squared transverse momentum wrt beam

use longitudinally invariant variables, eg pT, rapidity (y) and azimuth (φ)

exclusive kT algorithm (Catani, Dokshitzer, Seymour & Webber 1993)

$d_{ij} = \min(p_{Ti}^2, p_{Tj}^2)\Delta R_{ij}^2, \quad \Delta R_{ij}^2 = (y_i - y_j)^2 + (\varphi_i - \varphi_j)^2$

one cutoff parameter dcut – how to choose cut, no absolute scale as in e+e-

inclusive kT algorithm (Ellis&Soper 1993)

normalise with R → all jets are at least separated by R

$d_{ij} = \min(p_{Ti}^2, p_{Tj}^2)\Delta R_{ij}^2/R^2$

two parameters R and dcut

hadron colliders: only # jets above threshold are IR safe

Reformulate recombination algorithm: inclusive kT

kT at hadron colliders: inclusive kT algorithm (Ellis&Soper 1993)

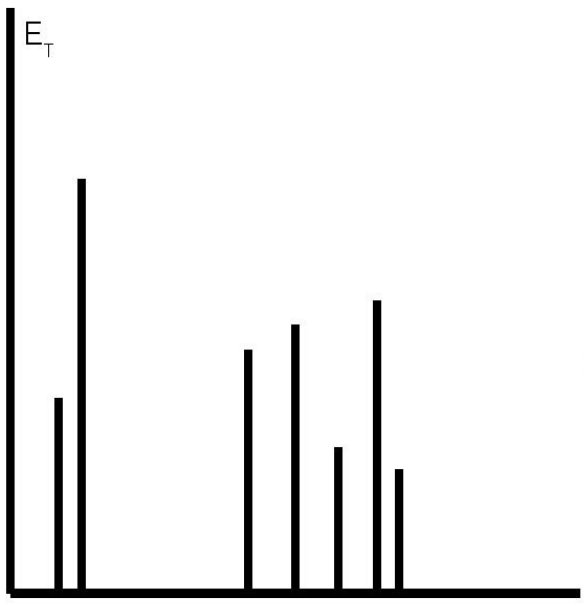
two parameters R and $p_{T,min}$

metric: $d_{ij} = \min(p_{Ti}^{2}, p_{Tj}^{2})\Delta R_{ij}^{2}/R^2$ $\Delta R^2 = (y_i - y_j)^2 2 + (\varphi_i - \varphi_j)^2$

distance of object to beam: $d_{ijB} = p_{Ti}^{2}$

combination algorithm

- calculate d_{ij} for all pairs
- if $d_{ij} < d_{iB}$: recombine i and j
- if $d_{ij} > d_{iB}$: i is a jet and removed from list
- repeat until no particles are left
- only use jets with $p_{T,} > p_{T,min}$



Features of the kT algorithm

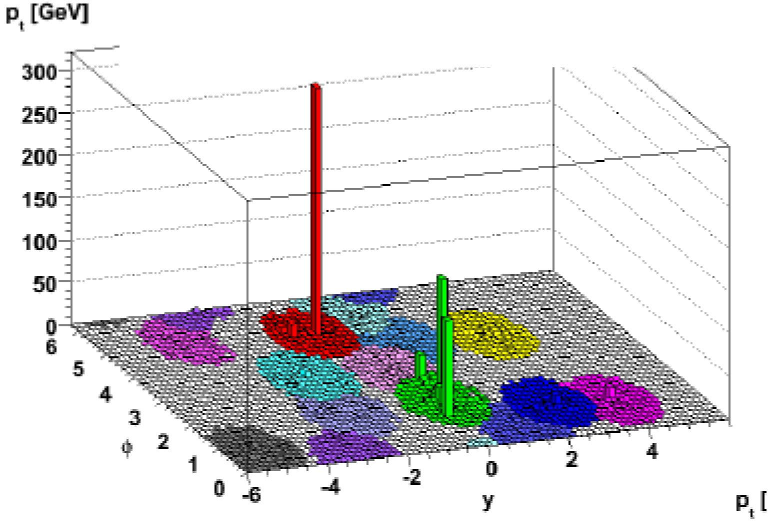
+ collinear and infrared safe
+ every hadron is uniquely assigned to a jet
+ few hadrons that belong to a given parton are missing

– significant noise from underlying event and multiple interactions

– jets have complicated shapes

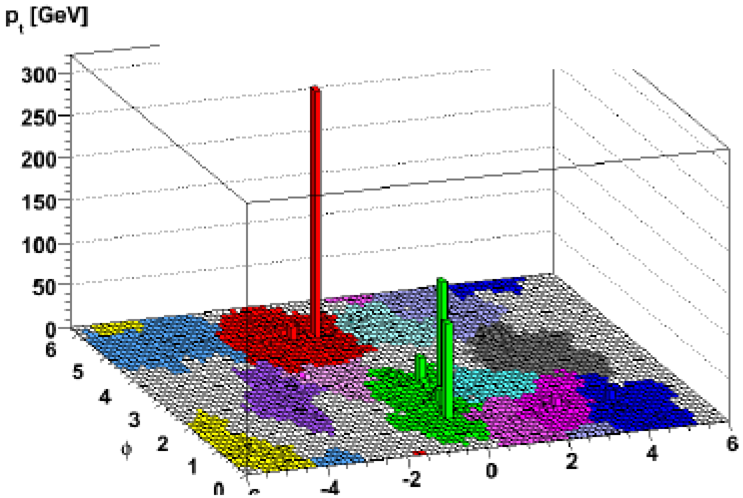
Cone vs kT algorithms

Cone



(Some) cone algorithms give
circular jets in y − φ plane
easier for experiments e.g. for acceptance
corrections

kT



kt jets are irregular
because soft particles
cluster together first

sequential recombination algorithms

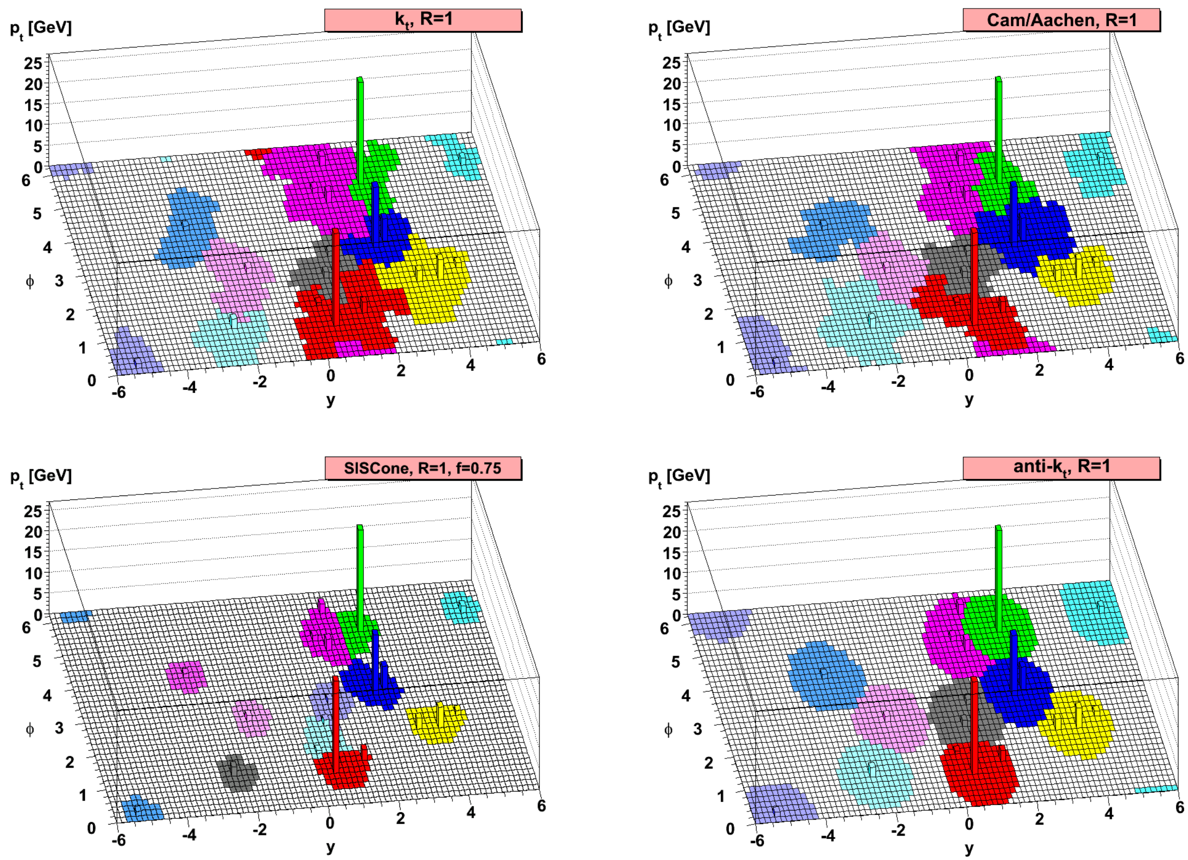
several families, they differ in choice of distance measure

- kT algorithm – just discussed
  hierarchical in pT
- Cambridge Aachen (simplest algorithm)
  recombine pair of objects closest in Δrij
  repeat until all ΔRij > R — remaining objects are jets

- anti-kT
  $d_{ij} = 1/\max(p_{Ti}^2, p_{Tj}^2)\Delta R_{ij}^2/R^2$, $d_{iB} = 1/p_{Ti}^2$
  first cluster high energy - high energy and high energy – low energy particles
  disfavours clustering between pairs of soft particles
  → most clusterings involve at least one hard particle → clustering grows around hard cores
  → anti-kT gives cone like jets without using stable cones, they have a well defined area
  + infrared and collinear safe
  + insensitive to underlying event and multiple interactions
  good performance, main jet algorithm by LHC experiments

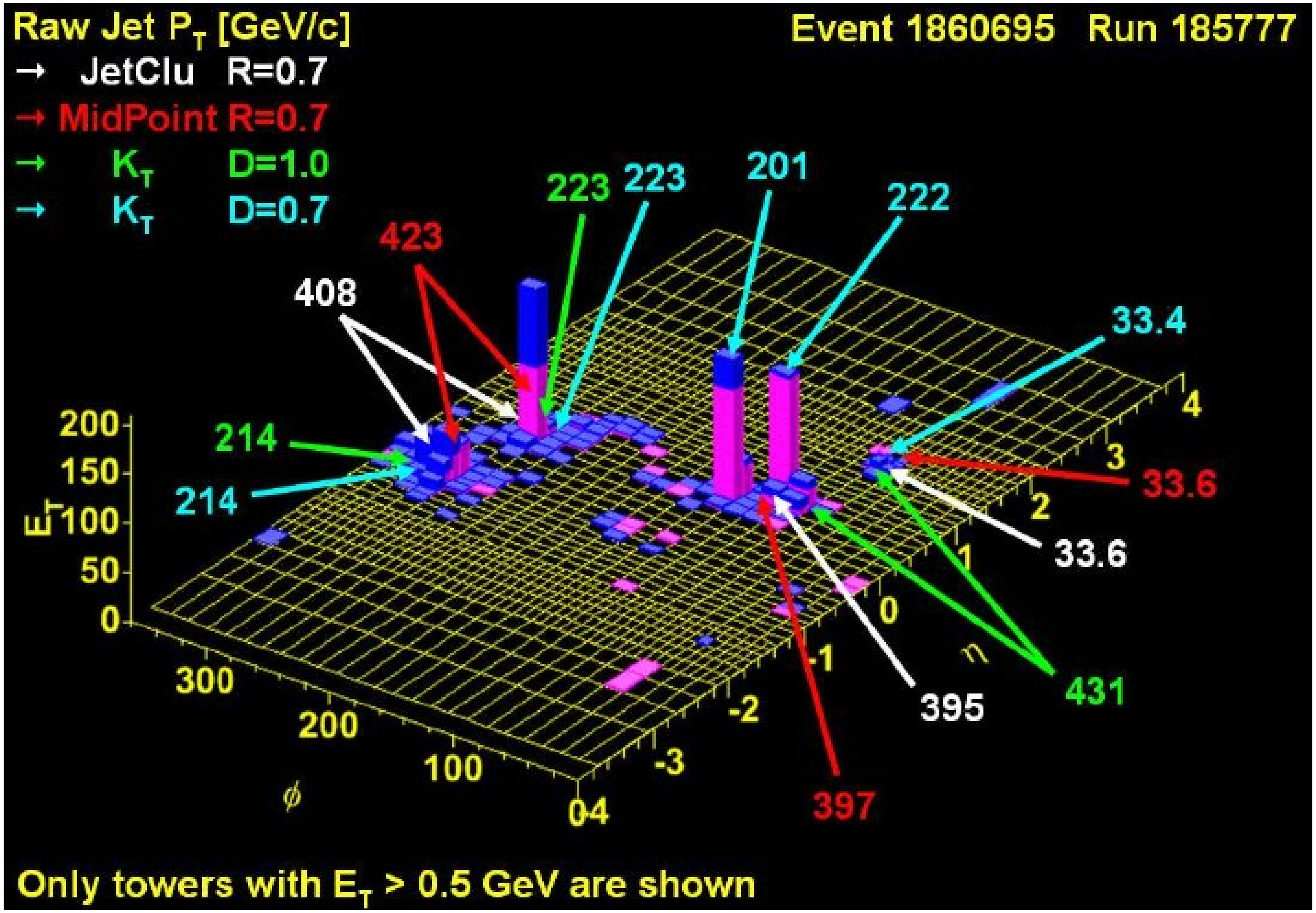
Typical shapes of IR and CL safe algorithms

JHEP 0804:063,2008

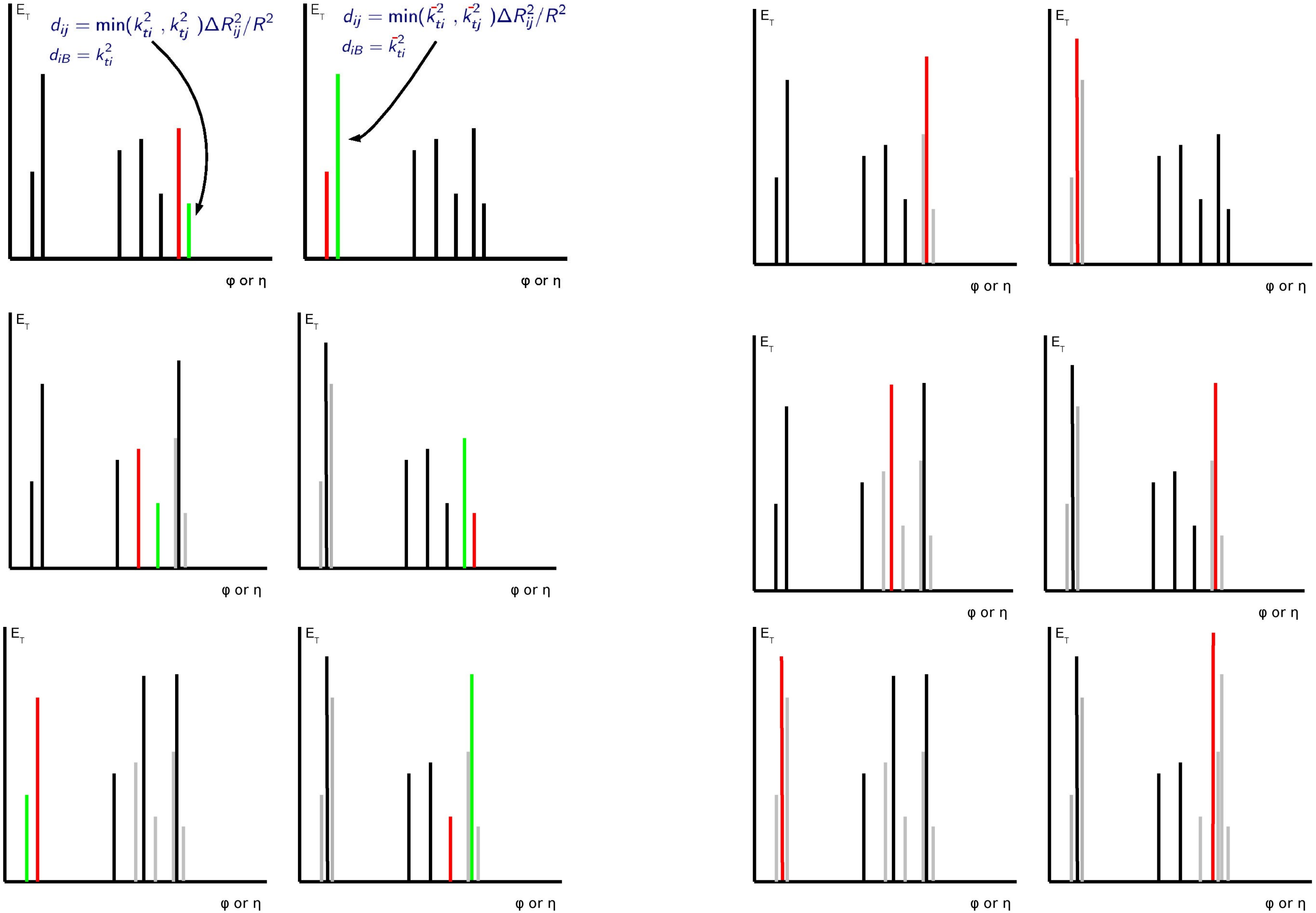


anti-kT gives cone like jets without using stable cones

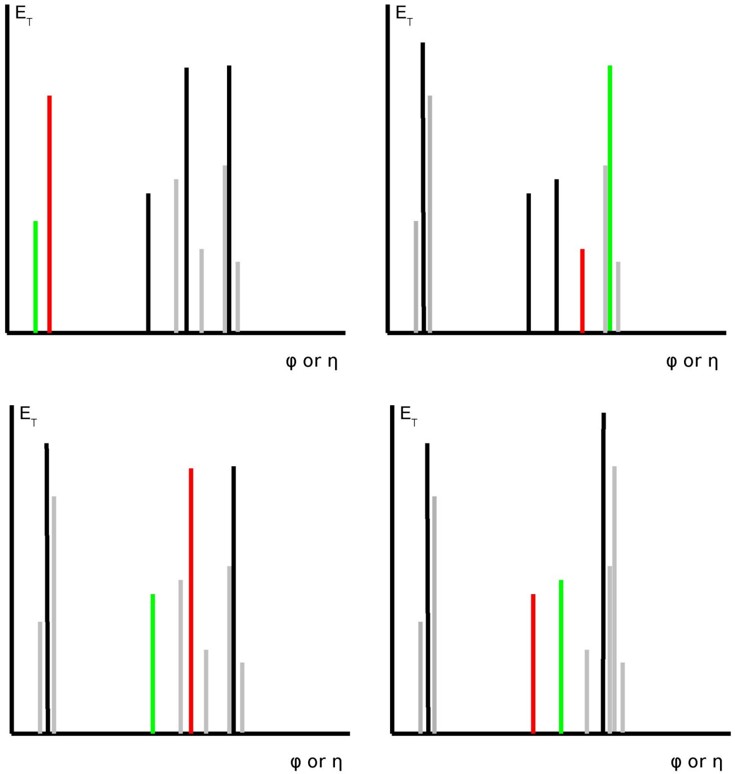
Result of different jet algorithms for one CDF event

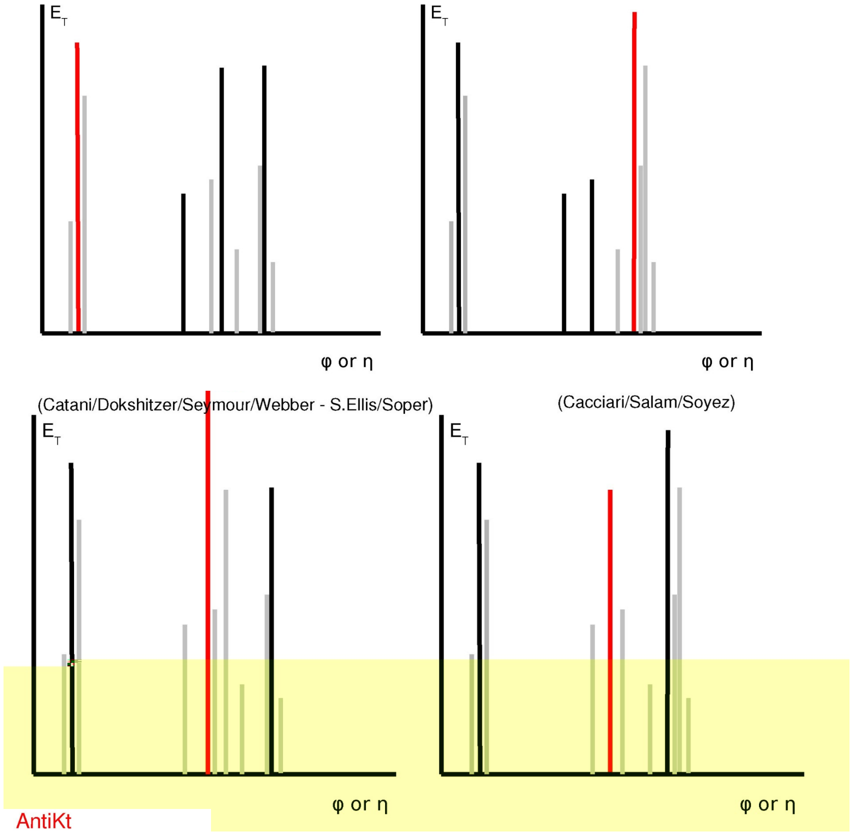


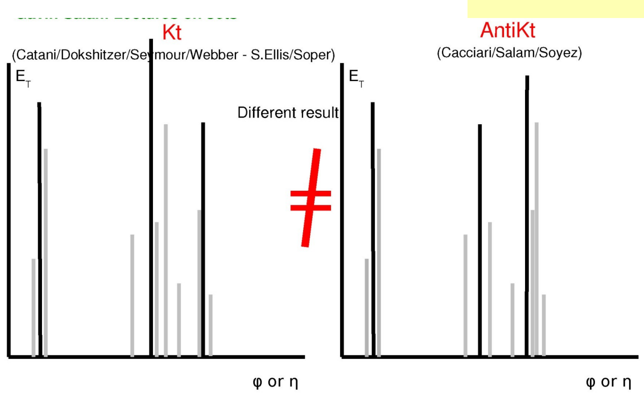
kT and anti-kT in action



kT and anti-kT in action

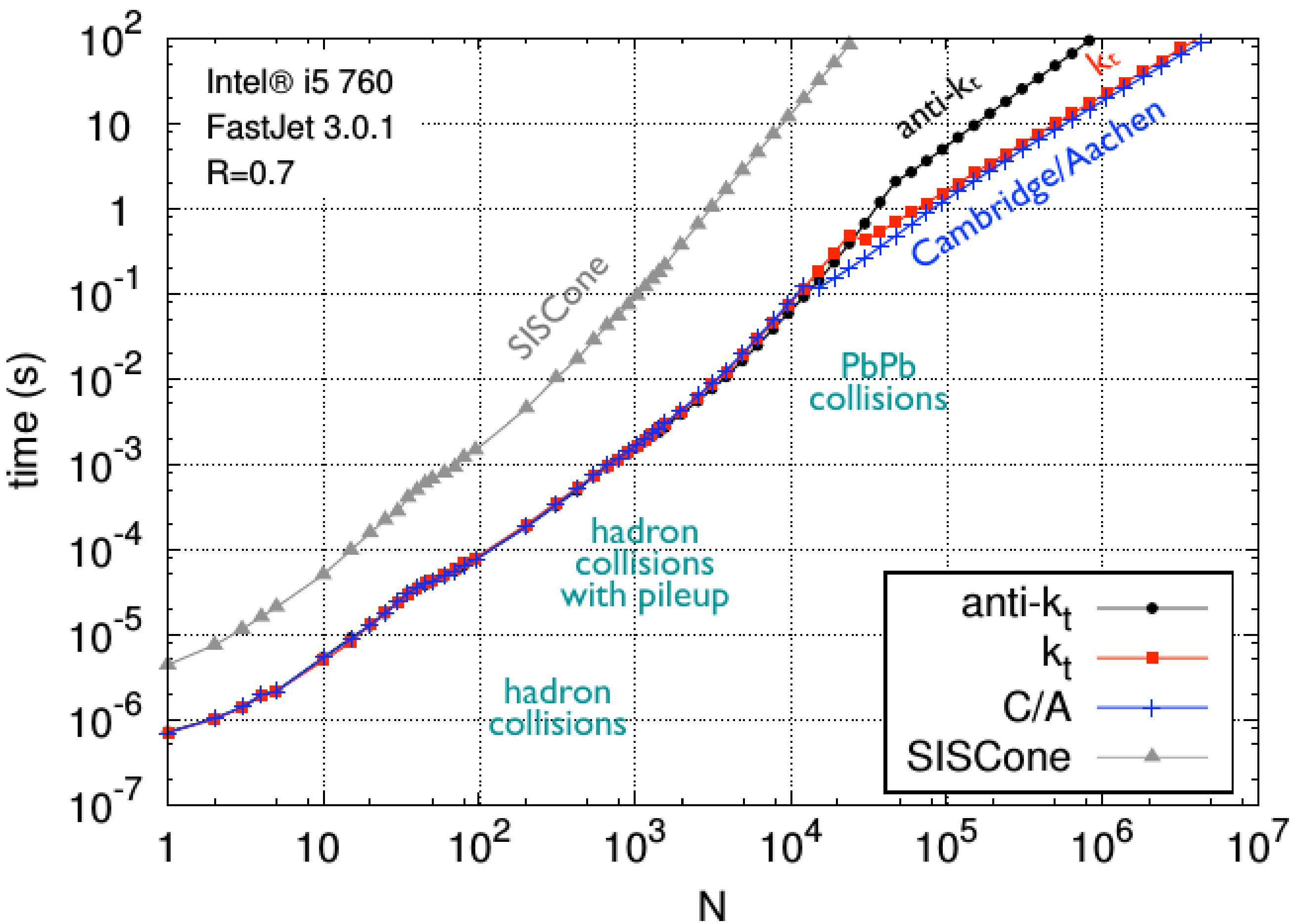






Performance

Time needed to cluster an event with N particles



Input objects for jets

different choices (example CMS):

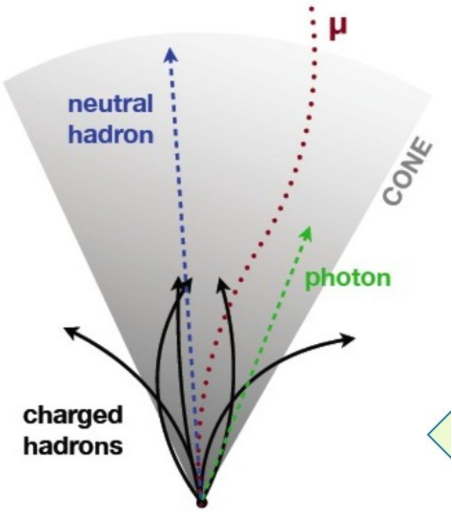
1) Calo jets:jets reconstructed from towers in the
ECAL and HCAL

2) Track jets reconstructed from tracks of charged particles
+ independent of calorimeter measurement
– neutral energy is missing

3) Tracks plus clusters (JPT: jet plus track)
replace calorimeter measurement with track measurement if available

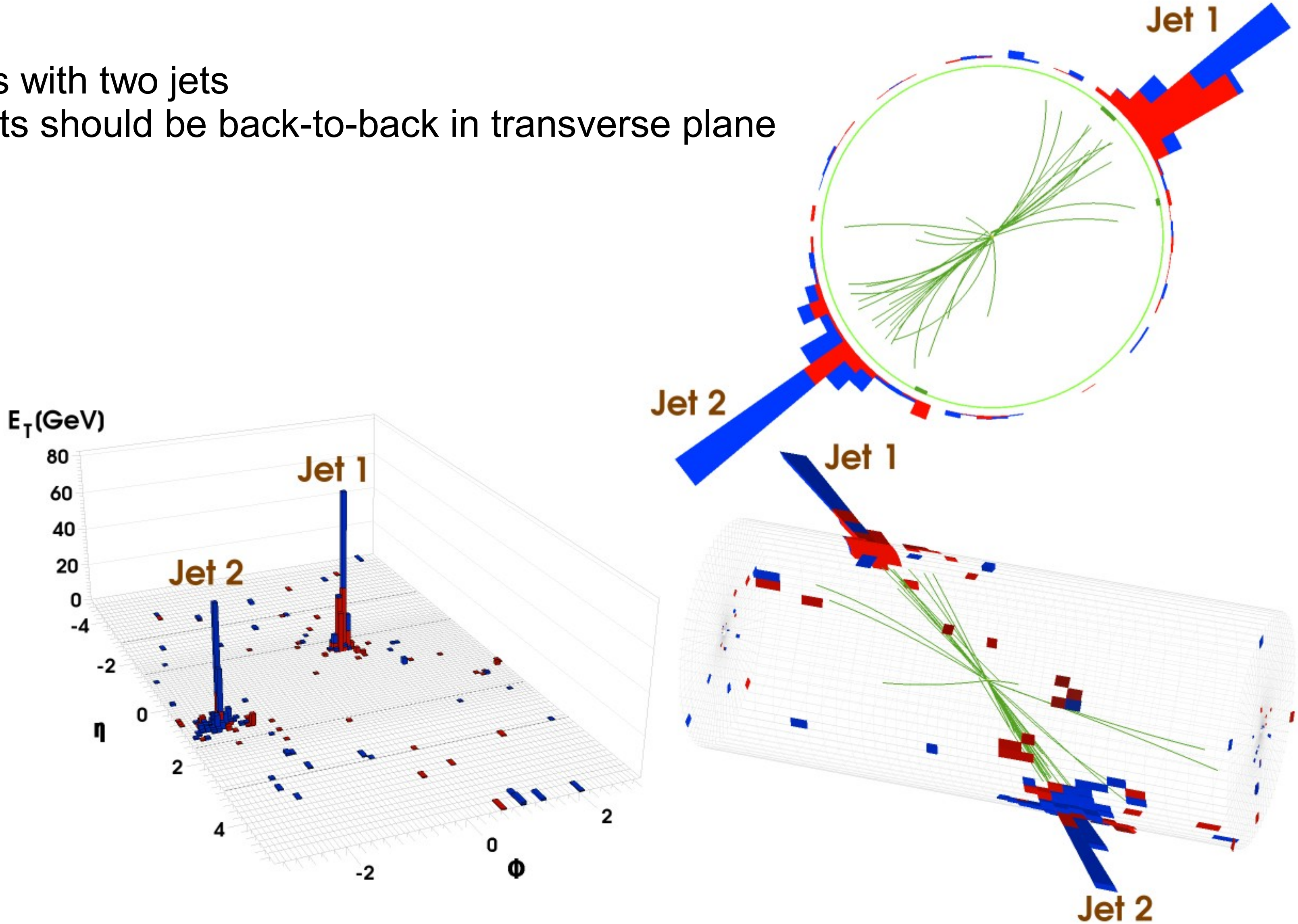
4) Particle flow: use calibrated particles (best knowledge)

→ using different inputs allows to study and constrain systematic effects



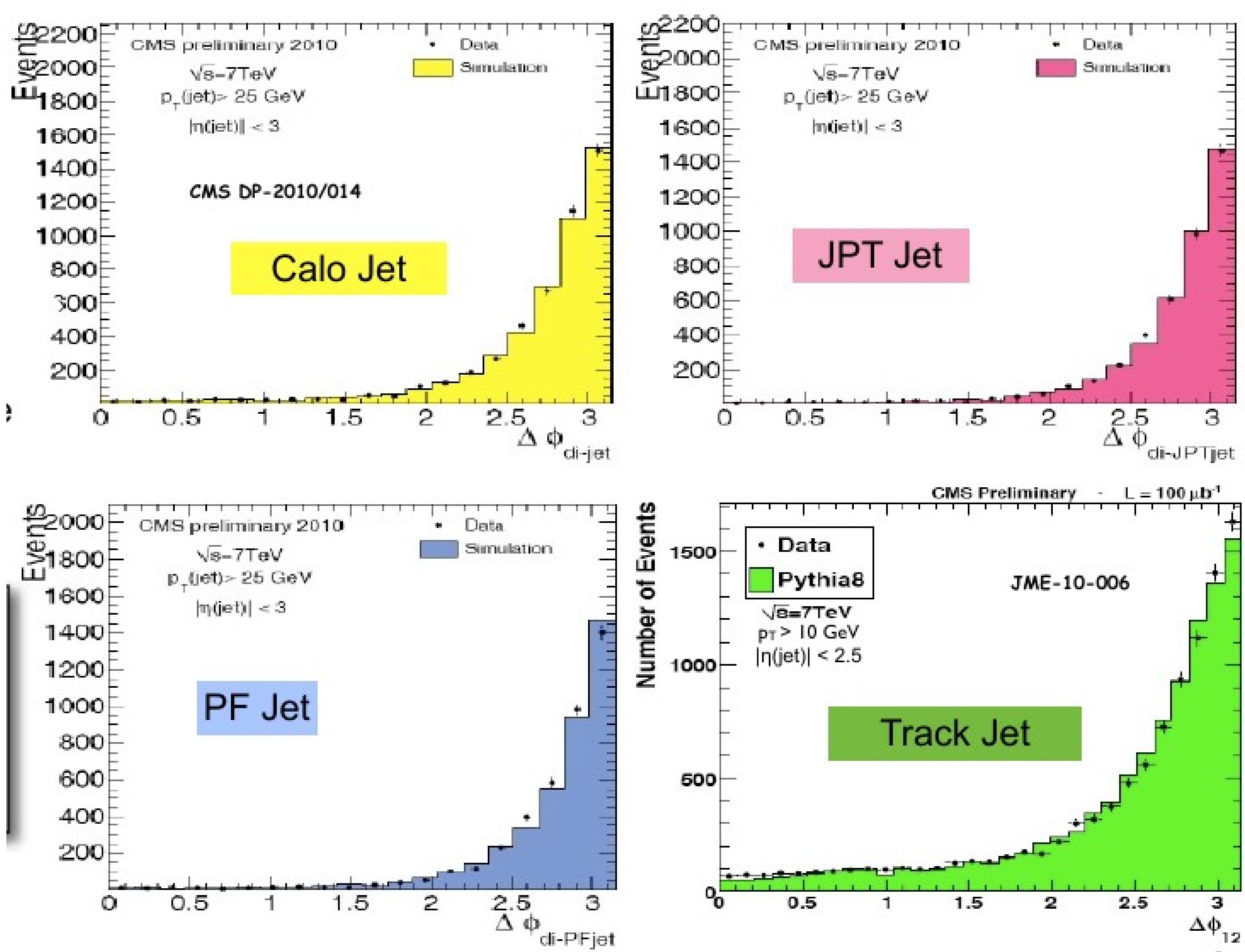
Dijets – compare data and MC

events with two jets
LO: jets should be back-to-back in transverse plane



Separation of di-jet events in phi

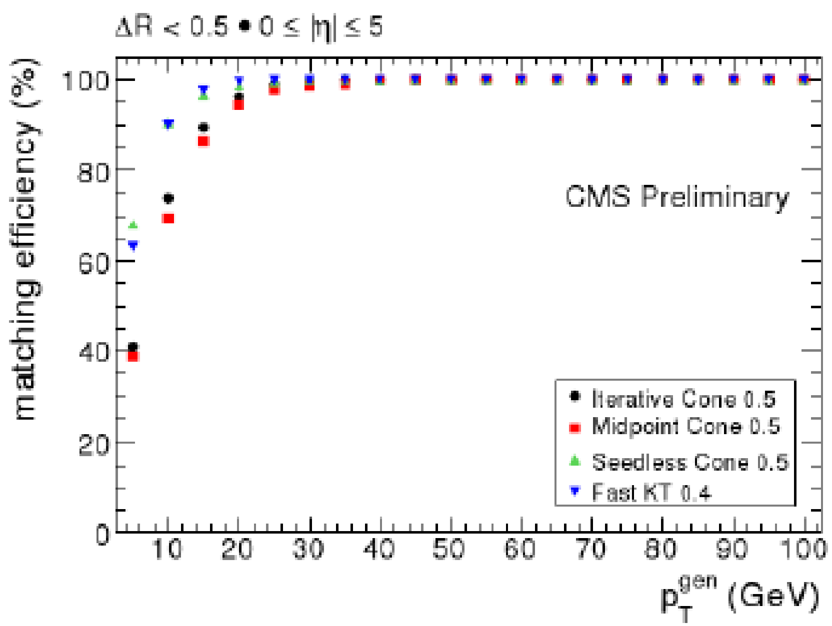
good agreement btw MC and data for all jet types

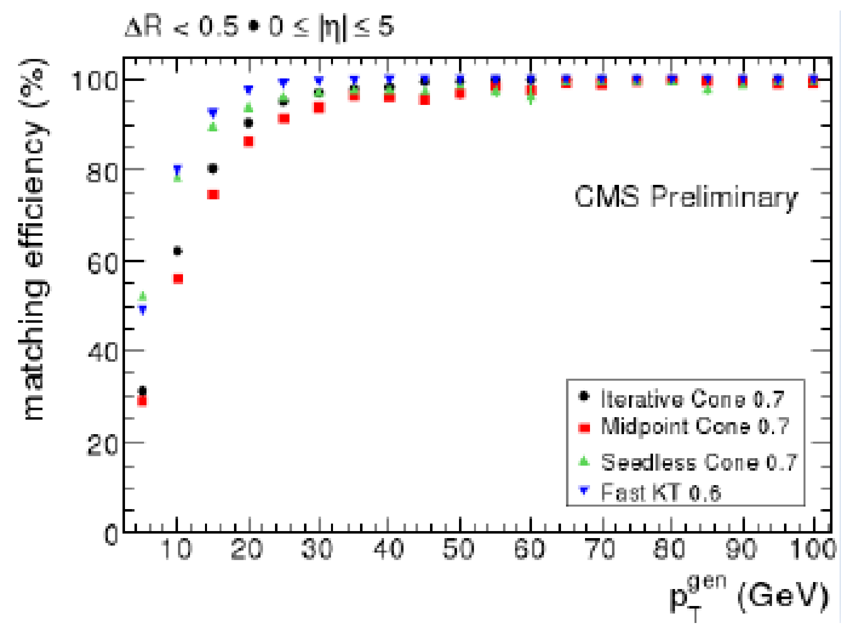


Matching efficiency

one meaningful measure of reconstruction efficiency
ϵ =particle jets matched to reconstructed jet within Δ R =0.5/ total number of particle jets

strong dependence on position resolution, ΔR and cone size
SIS Cone and kT tend to have better performance than Midpoint and Iterative Cone
smaller cone size parameter → better matching efficiency





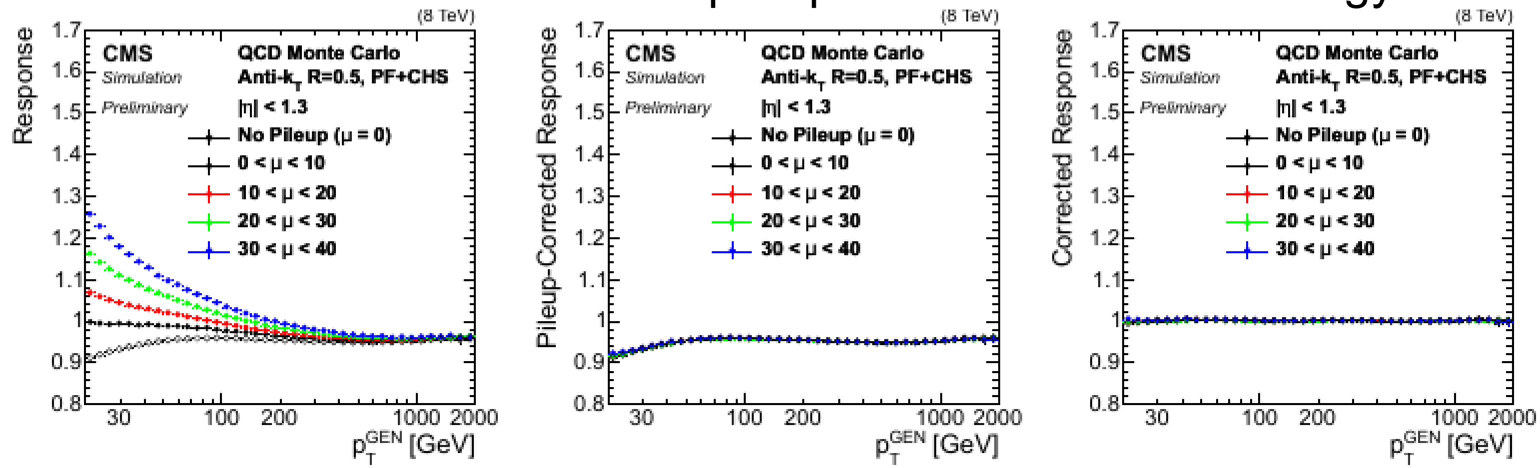
Jet energy correction (CMS)

correct reconstructed jet energy to obtain a measurement of the true particle

- event-by-event pileup correction to reduce pileup fluctuation
  → jet measurement independent of number of primary vertices
- multiplicative energy scale correction, determined from simulation:
  → correct for missing energy (neutrinos) leakage into/out of jet and mis-measurements
- additional correction in data for data-MC differences
  → input from data
  MC-data calibration samples: dijet, Z+jets, and isolated photon+jet events
  → uncertainty in energy scale given by the uncertainty on the measurement of the jet response in data
- optional: correct for jet type: gluon vs quark jets, b or c quarks jets …

reconstructed | after pileup correction | after energy scale correction

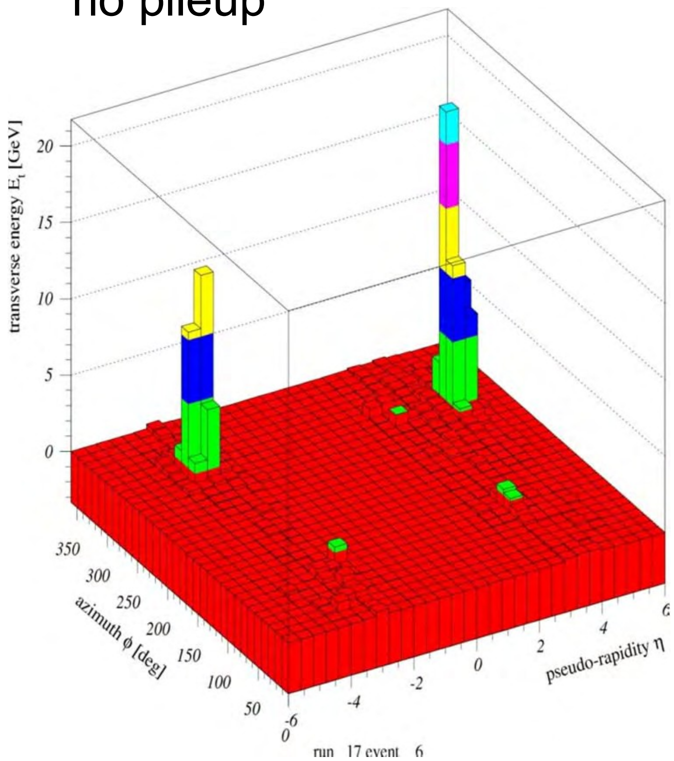
(pT(rec)-pt(gen))/pt(gen)



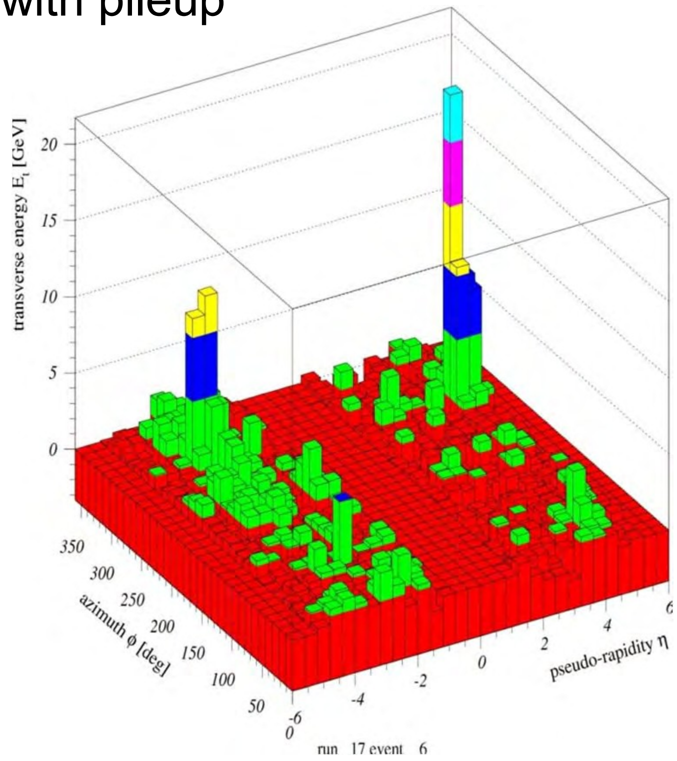
pileup correction

pile-up: average additional energy due to additional proton-proton interactions from in-situ measurements

no pileup



with pileup

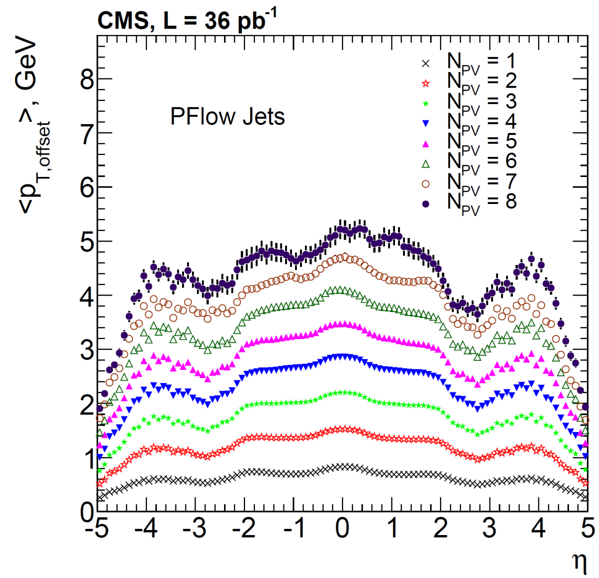


pileup correction

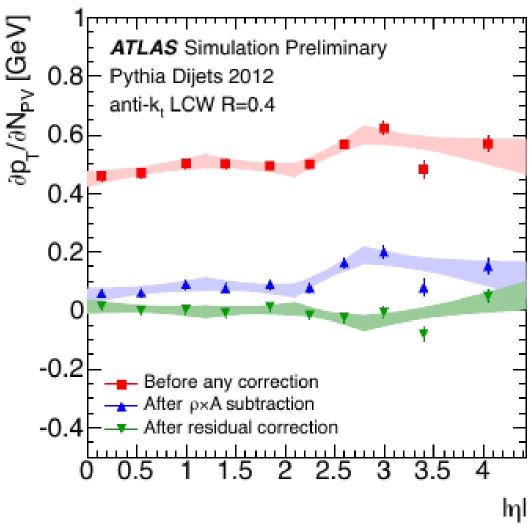
$p_T^{corr} = p_T^{jet} - \rho\, A^{jet}$

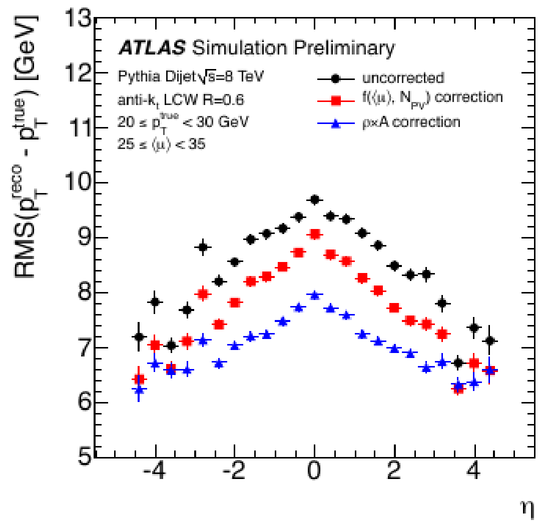
ρ: median p_T density for each event

A^{jet} area of jet



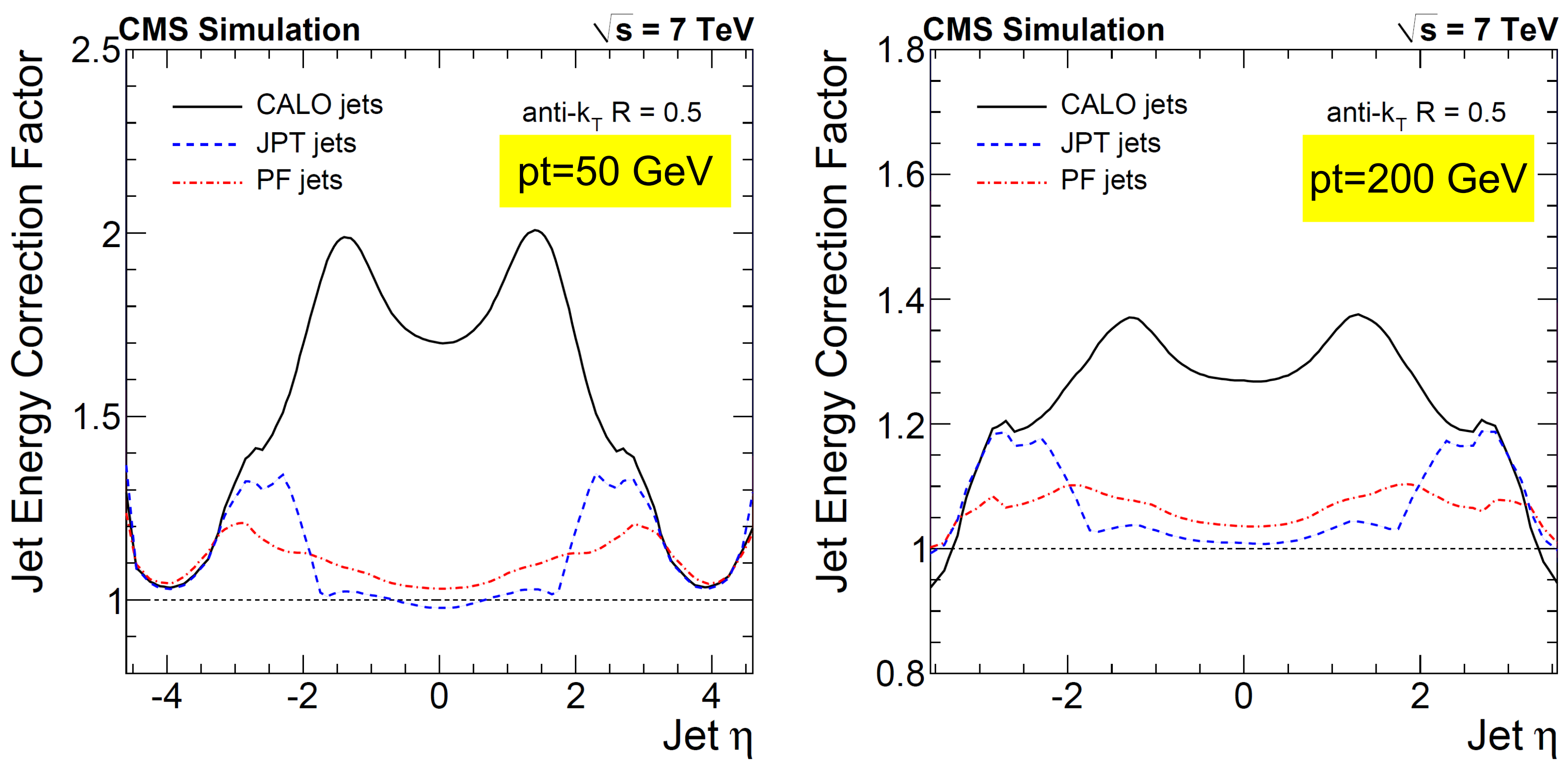
pileup correction rises approx. linearly with # PVs





Relative correction: example CMS

Typical size of correction to equalize detector response versus η
for particle-flow jets< 20%
if jets are only built from CALO objects, the corrections are significantly larger



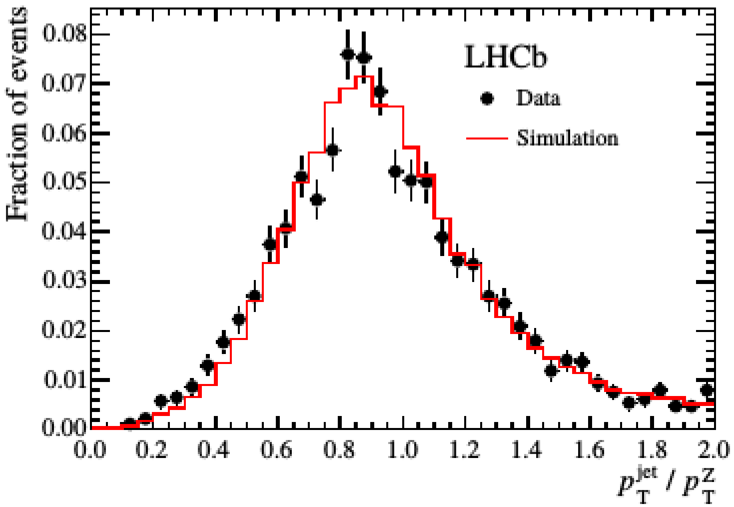
Comparison data-MC

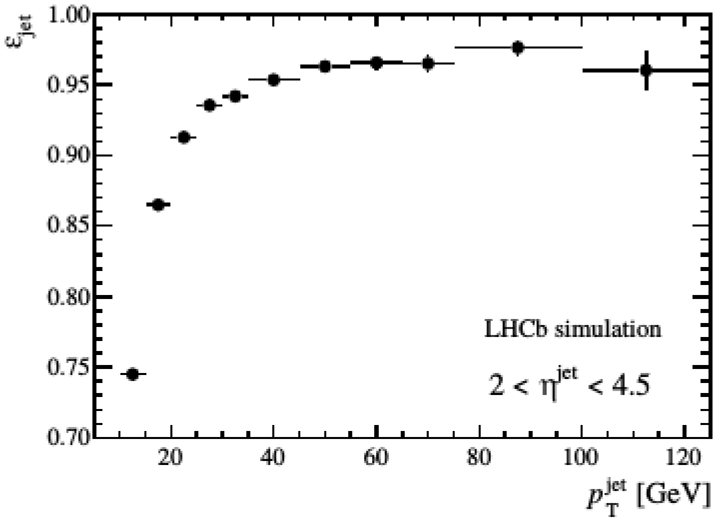
LHCb: events with a reconstructed Z and one jet:
back-to-back → jet and Z should balance in pT

→ good agreement between data and simulation
resolution about 20%, flat in pT

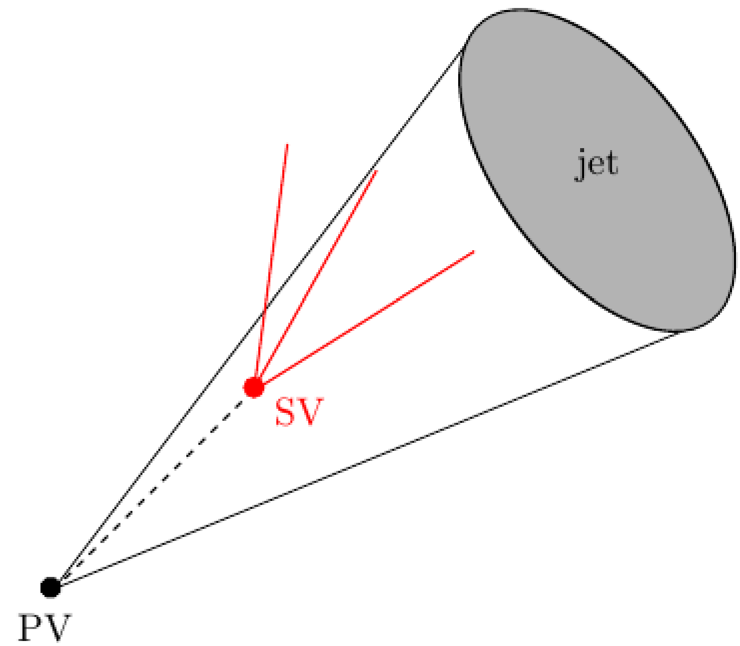
about 15% of jet pT is not measured

jet reconstruction efficiency about 95%





Tagging of jets



Tag secondary vertices in the jet

→ use multivariate analysis (MVA) to distinguish

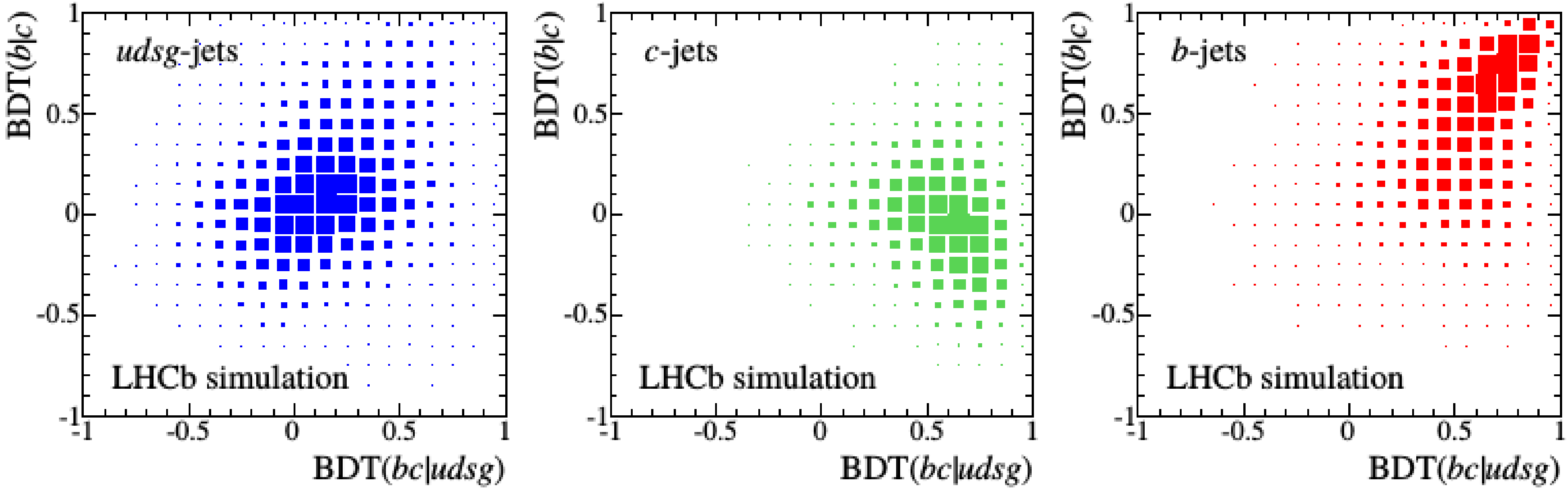
1) jets from light from heavy quarks
2) jets from c quarks from jets from b quarks

example of variables sensitive to flavour

distance of SV to jet axis
tracks in jet
track of SV
reconstructed mass of SV
charge of SV
flight distance
transverse momentum of SV

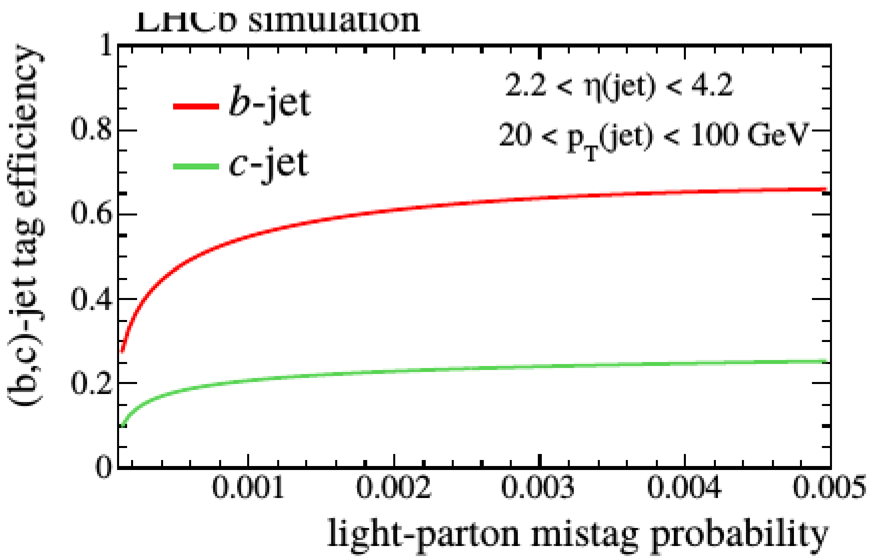
eg boosted decision tree (BDT), artificial neural net (ANN) ...

Tagging of jets



BDT(bc|udsg): distinguish light from heavy quarks
BDT(b|c): distinguish b from c quarks

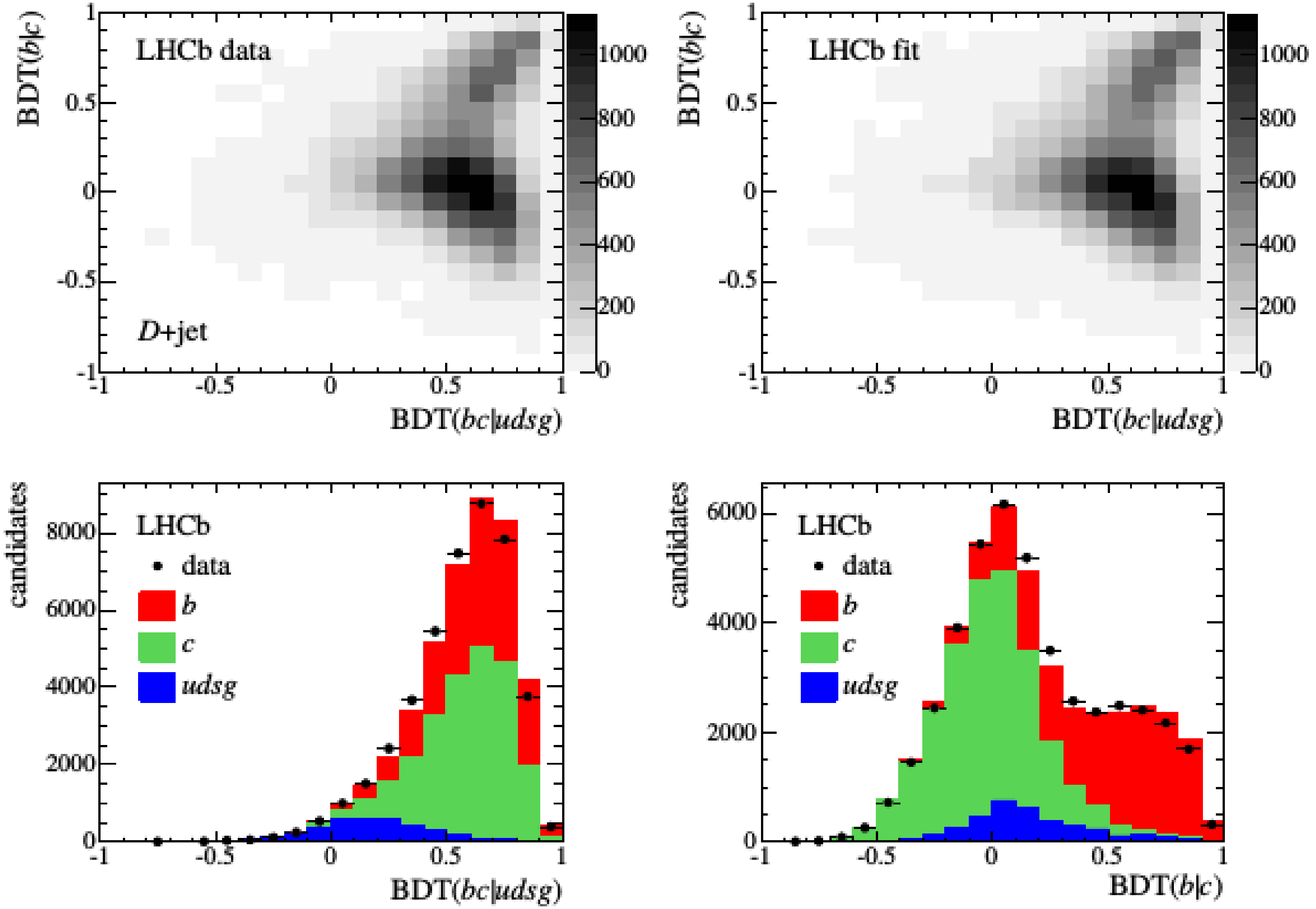
b-tagging efficiency 60%
c-tagging efficiency 20%



validate response on data

jet-tagging: validate on data

c, b-enhanced $(D + j)$



you can do more: separate gluon and quark jets

the fragmentation process is different for gluon and quark jets
(different colour and charge interaction) can we distinguish quark from gluon jets?

simulation:
gluon jets tend to
- be wider,
- have higher multiplicities
- more uniform energy distribution

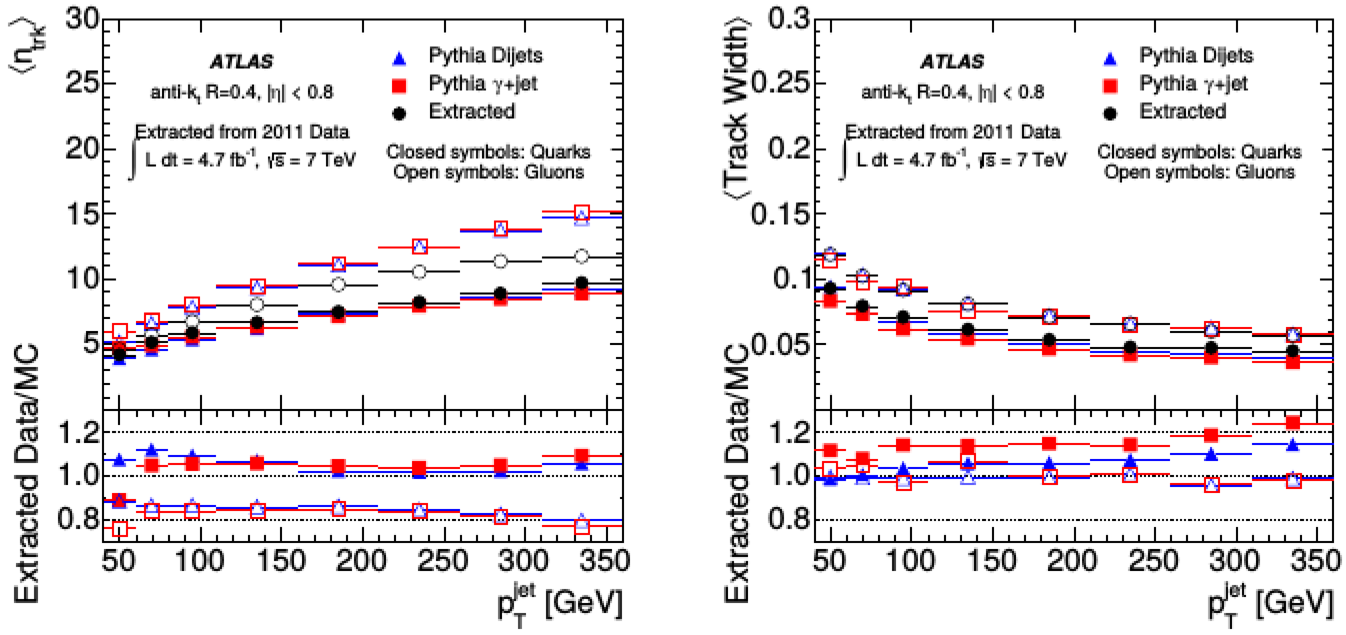
quark jets tend to
- produce narrow jets
- have hard constituents which carry most of the energy

separation of jets initiated by quarks or by gluons is a powerful tool in searches for new physics
eg SUSY decays produce many light quarks, h → ZZ → llq qbar
→ jet identification helps to suppress backgrounds

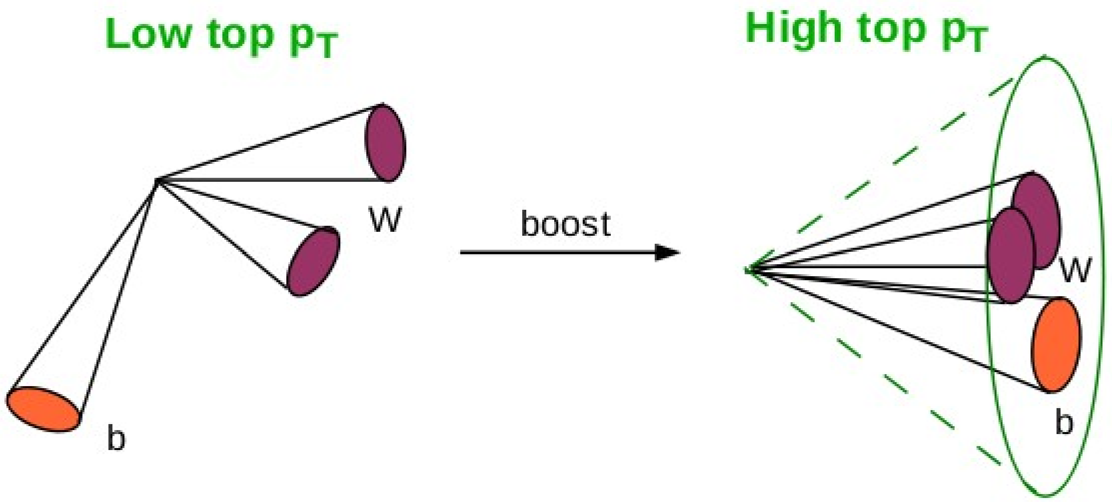
you may also look at jet substructures ...

Separate gluon from quark jets

track width: w =Sum(pT,i × ΔR(i,jet))/Sum(pT,i)



You can do more: jet substructure



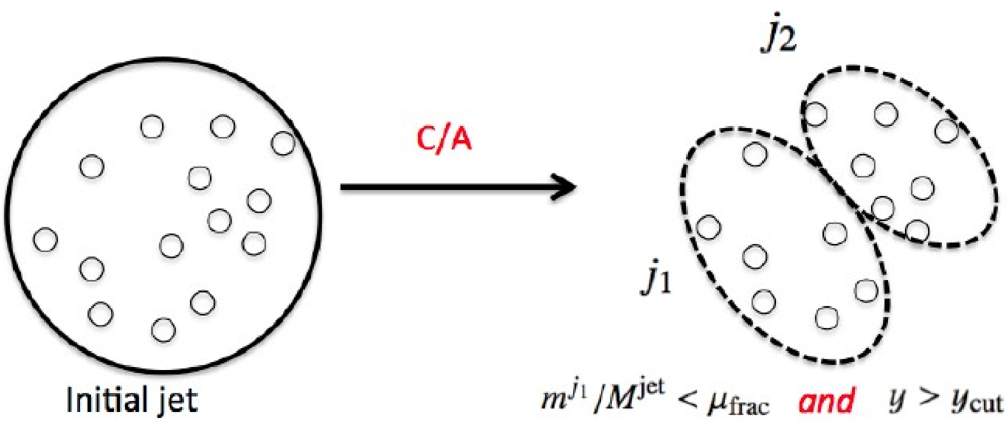
LHC energies: highly boosted objects → more than one parton contained in a jet
→ look at/understand jet substructures
distinguish massive boosted objects from gluons or light quarks

- jet mass: deduced from 4-momentum sum of all jet constituents
- kT splitting scales: recluster jet constituents with kT algorithm
  kT tends to cluster the hardest constituents last → dij for two remaining proto-subjets
  define splitting scale $\sqrt{dij}=\min(p_{Ti},p_{Tj})\Delta Rij$
  QCD jets: asymmetric splitting, heavy particle jets: symmetric splitting
- N-subjettiness: tau_N: measure of how well jet can be descibed as containing
  N or fewer subjets

You can do more: jet substructure

grooming techniques: redefine jet in high luminosity environment
remove soft QCD radiation from jet → enhances differences between QCD jets and jets from heavy boosted objects
techniques

- mass-drop: undo last step in C/A, require that the two subjets are symmetric and there is a significant mass difference $m^{j1}/m^{jet}\ \mu_{cut}$
- filtering: remove all constituents in jet that are outside of the three hardest subjets (C/A)
- trimming: create sub-jets, and removes them if p_{Ti}/p_T^{jet}<fcut
- pruning: similar to trimming removes in addition wide angle radiation

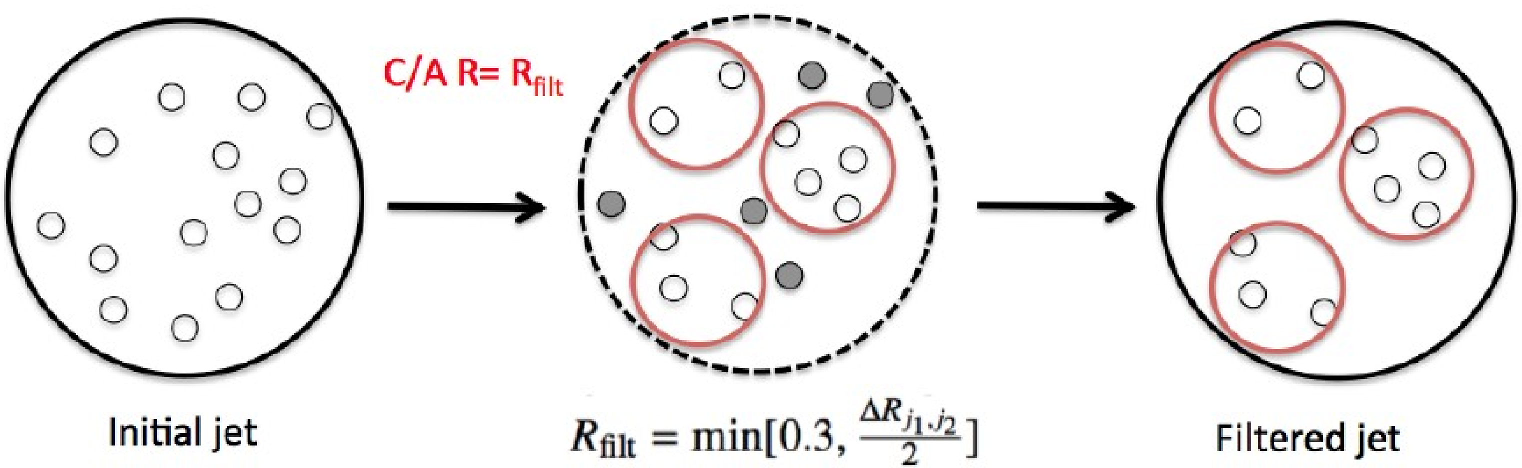


mass-drop:
drop jet if condition is not fulfilled

You can do more: jet substructure

grooming techniques: redefine jet in high luminosity environment
remove soft QCD radiation from jet → enhances differences between QCD jets and jets from heavy boosted objects
techniques

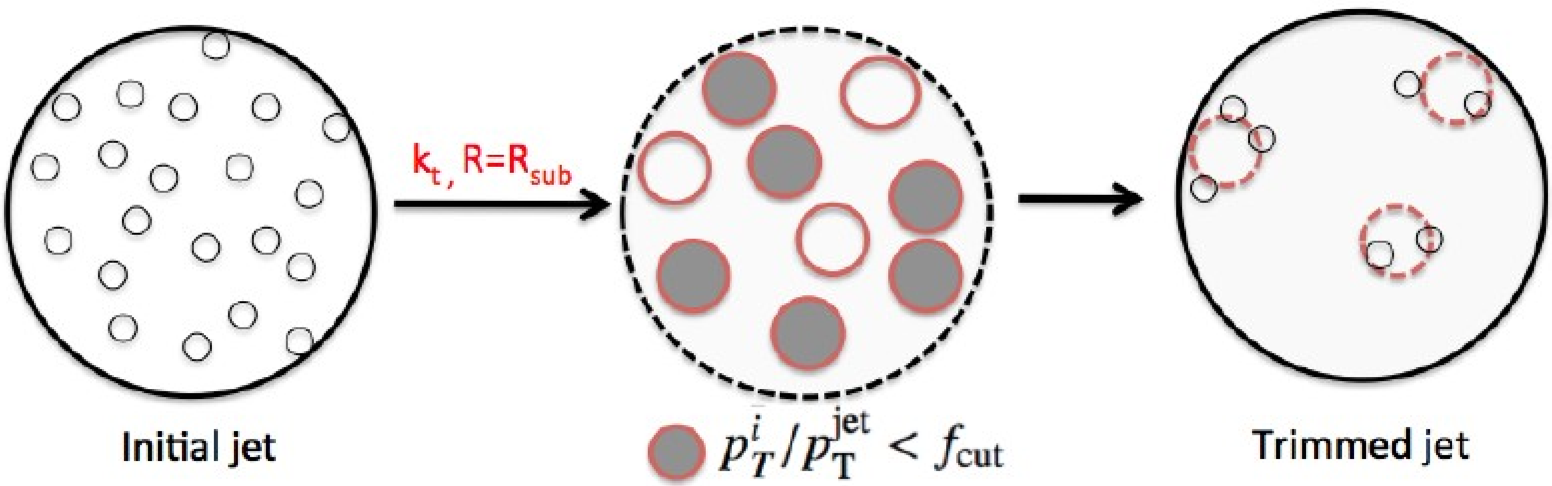
- mass-drop: undo last step in C/A, require that the two subjets are symmetric and there is a significant mass difference $m^{j1}/m^{jet}\ \mu_{cut}$
- filtering: remove all constituents in jet that are outside of the three hardest subjets (C/A)
- trimming: create sub-jets, and removes them if p_{Ti}/p_T^{jet}<fcut
- pruning: similar to trimming removes in addition wide angle radiation



You can do more: jet substructure

grooming techniques: redefine jet in high luminosity environment
remove soft QCD radiation from jet → enhances differences between QCD jets and jets from heavy boosted objects
techniques

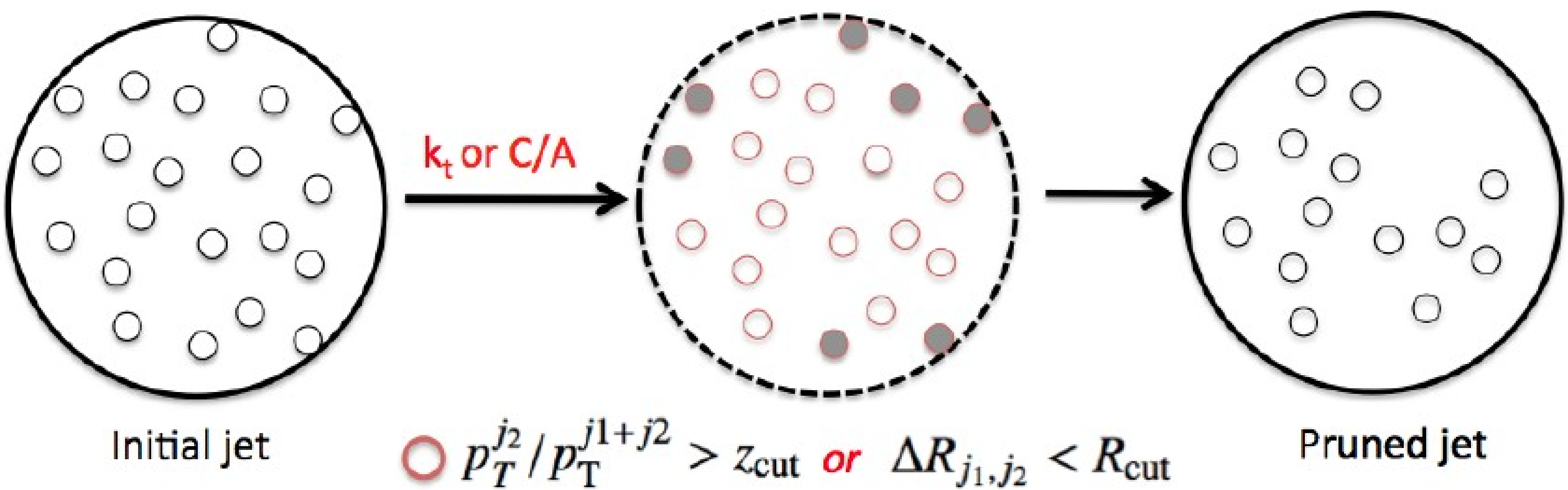
- mass-drop: undo last step in C/A, require that the two subjets are symmetric and there is a significant mass difference $m^{j1}/m^{jet}\ \mu_{cut}$
- filtering: remove all constituents in jet that are outside of the three hardest subjets (C/A)
- trimming: create sub-jets, and removes them if p_{Ti}/p_T^{jet}<fcut
- pruning: similar to trimming removes in addition wide angle radiation



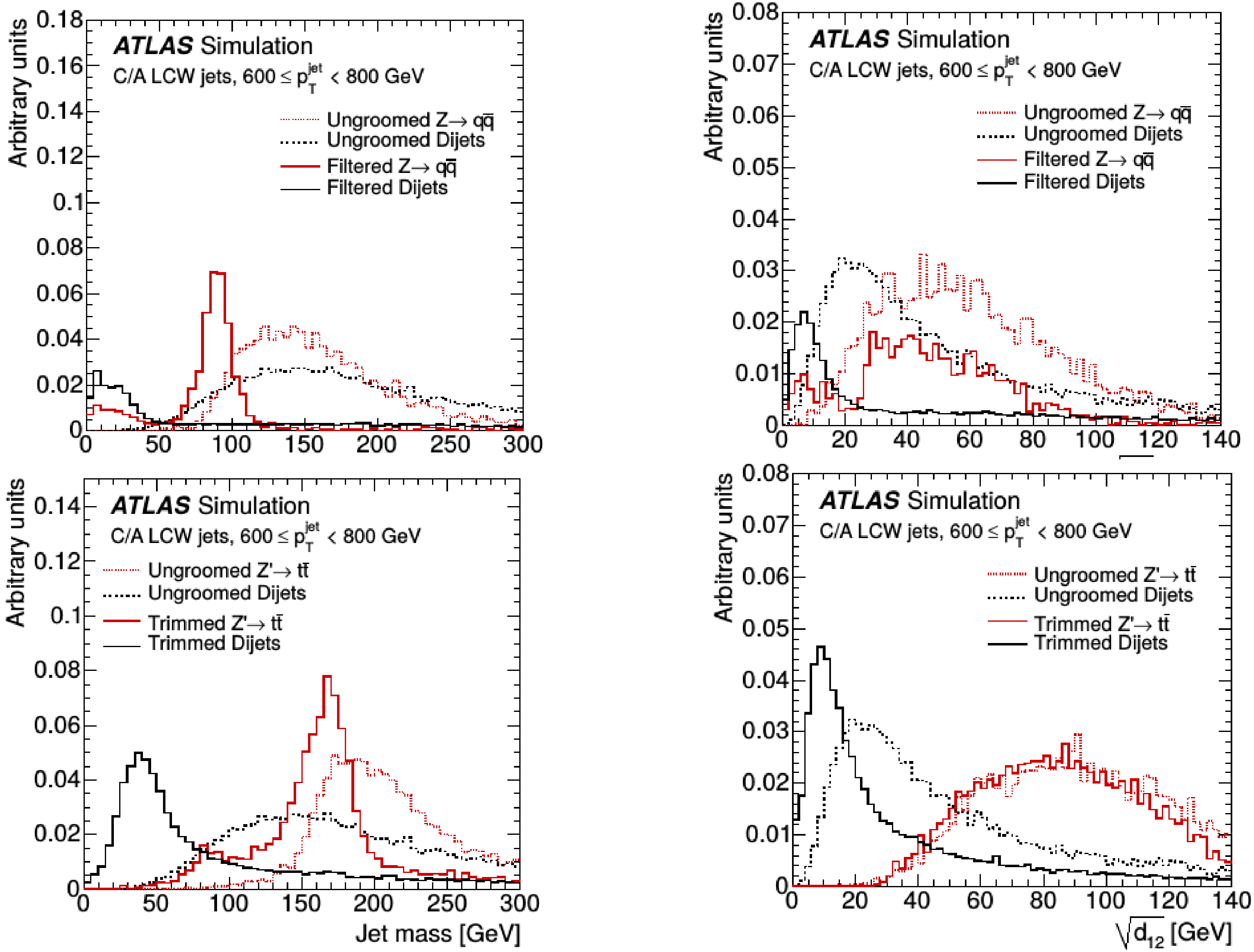
You can do more: jet substructure

grooming techniques: redefine jet in high luminosity environment
remove soft QCD radiation from jet → enhances differences between QCD jets and jets from heavy boosted objects
techniques

- mass-drop: undo last step in C/A, require that the two subjets are symmetric and there is a significant mass difference $m^{j1}/m^{jet}\ \mu_{cut}$
- filtering: remove all constituents in jet that are outside of the three hardest subjets (C/A)
- trimming: create sub-jets, and removes them if p_{Ti}/p_T^{jet}<fcut
- pruning: similar to trimming removes in addition wide angle radiation



You can do more: jet substructure



Considerations when working with jets

know what jet algorithm you are using – is it appropriate for your physics question?
→ simulation studies

be sure your algorithm is infrared & collinear safe

are the conclusions robust when you change the algorithm?
and when you change the parameters?

there is a lot you may do when looking at jet substructures