An introduction to modern Jet Algorithms

Paolo Francavilla

Foz do Arelho 23rd June 2009
Hadronic calibration workshop 2009

Introduction

Quarks and gluons are confined: we only see hadrons.

A jet of hadrons is the **final state signature of quarks or gluons** produced in the hard scattering.

The **gross jet properties** (energy, momentum) **reflect the properties of the quarks or gluons**

BUT jets are not just smeared partons, they have a **complex substructure.**

A jet (or a pattern of jets) is a complex QCD event shape, designed to reflect as closely as possible the short distance degree of freedom (quarks, gluons, H, W, Z)

The jet is defined by an algorithm. Available Algorithms?

Jet Algorithms

- **"Cluster" algorithms**

  Generally start from all the elementary objects available (TopoClusters, Truth Particles), and perform an **iterative pair-wise clustering** to build larger objects (using either geometric and kinematic properties of the objects).
  Sort of inverts the QCD parton shower idea.
  Successfully used in e^+e^- and e p.

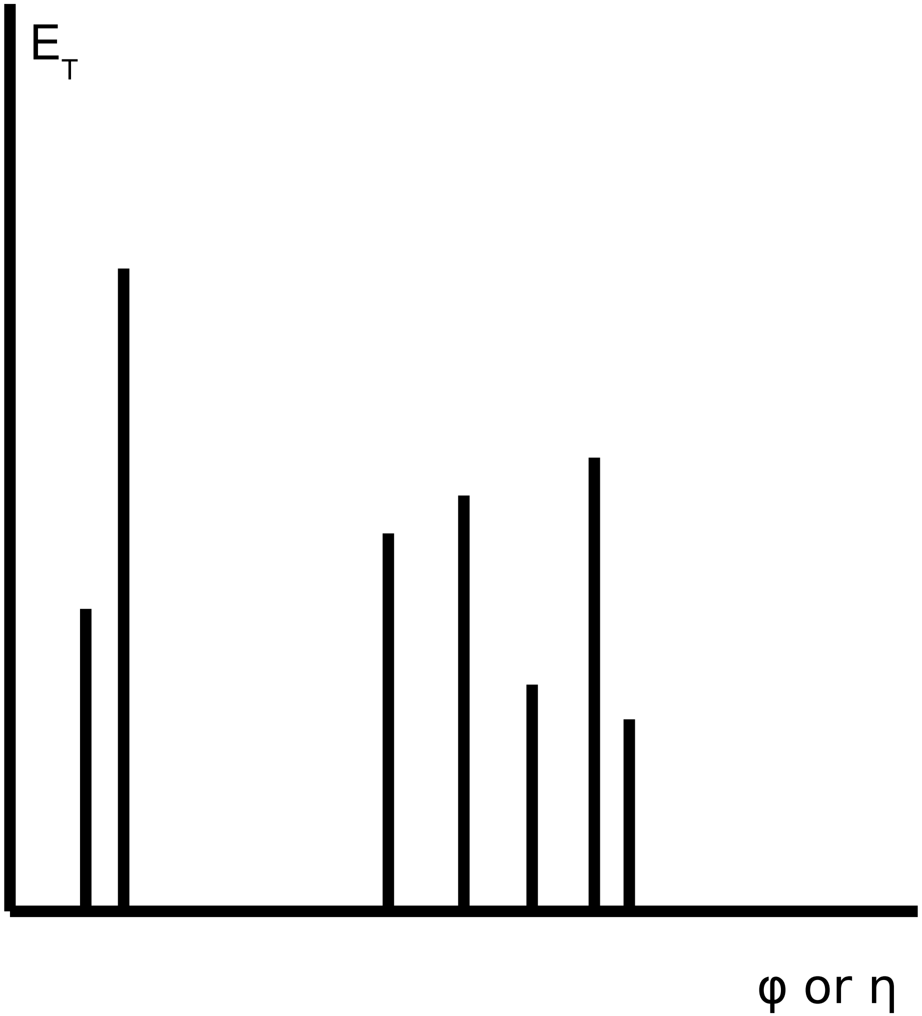
- **"Cone" algorithms**

  Generally seek to find **geometric regions which maximise the momentum** in a given area/shape (often not actually a cone!).
  Sort of mimics the "event display + eye" method.
  Traditionally used in hadron-hadron collisions.

In the last few years new jet algorithms were developed that overcome short-comings of algorithms used so far.

Clustering Algorithms

CTEQ-MCnet school 2008
Gavin Salam Lectures on Jets



Clustering Algorithms:

- Define a distance d_{ij} between two objects i, j :

$$\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$$

$$d_{ij} = \min(k_{ti}^2, k_{tj}^2)\Delta R_{ij}^2 / R^2$$

and a distance d_{iB} between one object i and the beam direction B:

$$d_{iB} = k_{ti}^2$$

- Find the smallest of d_{ij}, d_{iB} .
  If d_{ij} recombine i, j ;
  If d_{iB} , i is a jet.

Clustering Algorithms

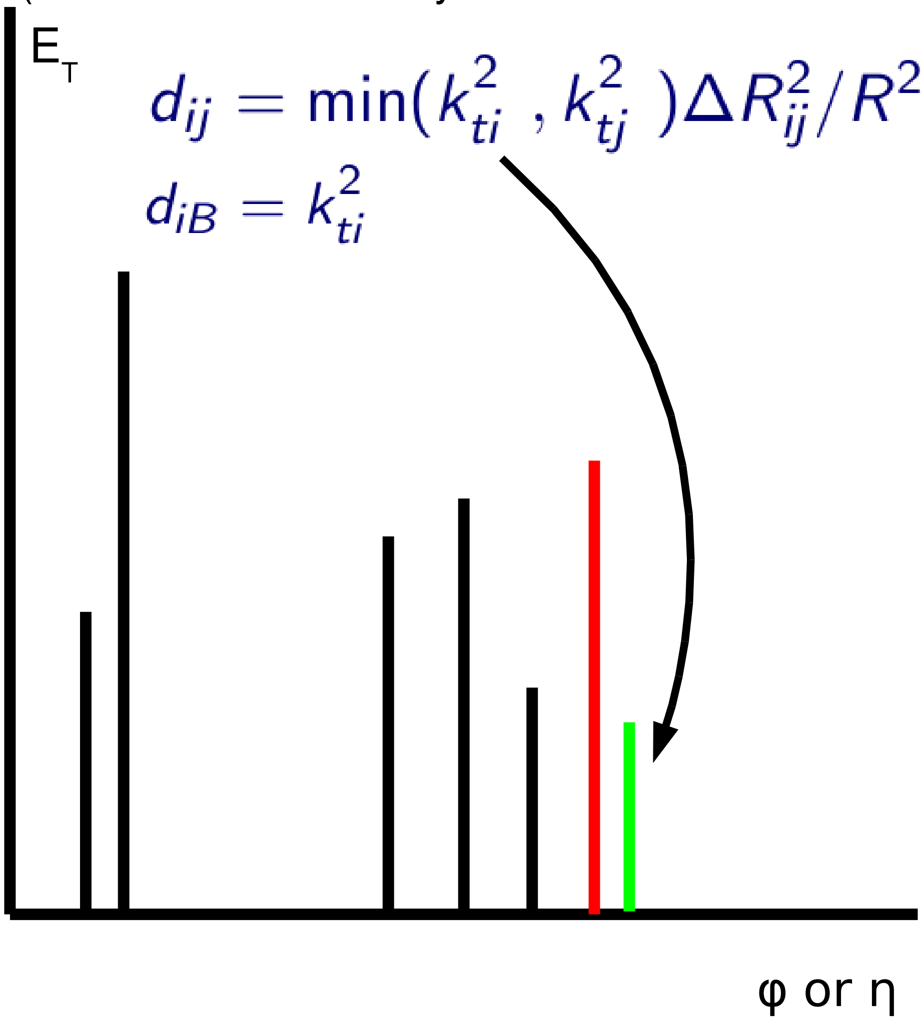
CTEQ-MCnet school 2008
Gavin Salam Lectures on Jets

Kt

(Catani/Dokshitzer/Seymour/Webber - S.Ellis/Soper)

$$d_{ij} = \min(k_{ti}^2, k_{tj}^2)\Delta R_{ij}^2/R^2$$

$$d_{iB} = k_{ti}^2$$

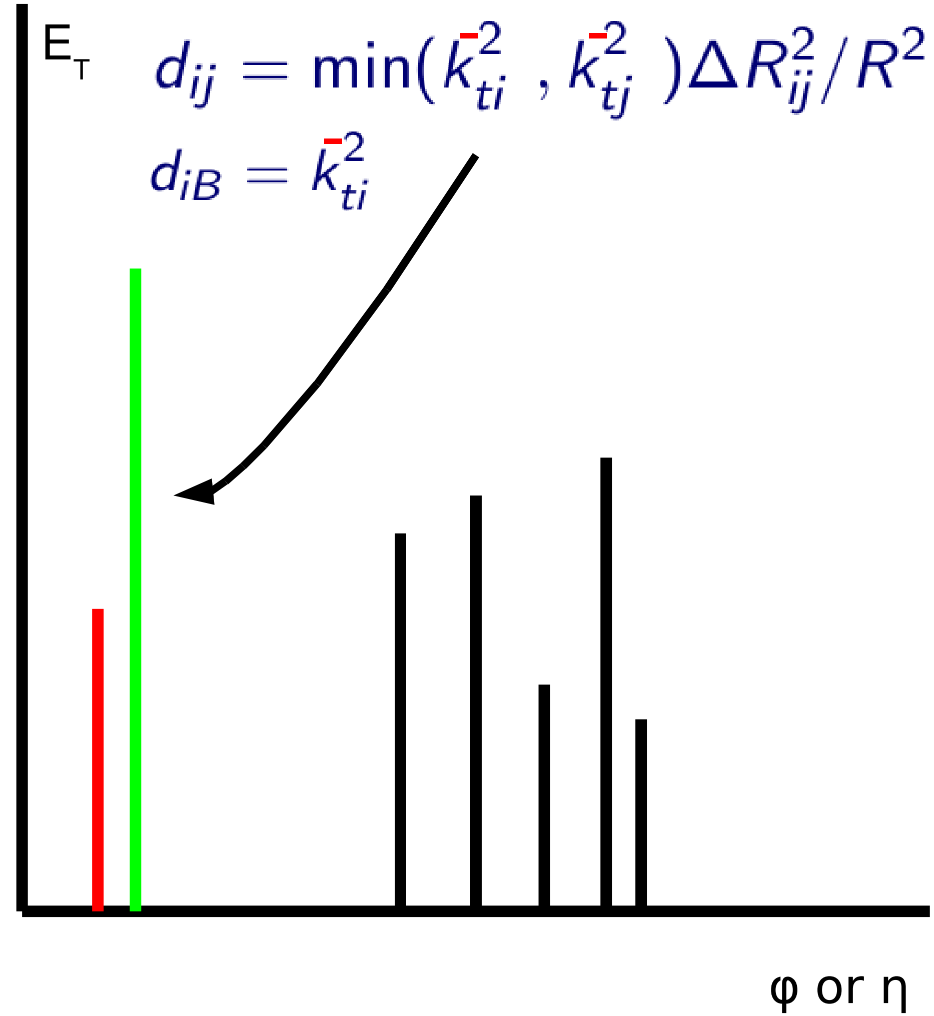


AntiKt

(Cacciari/Salam/Soyez)

$$d_{ij} = \min(k_{ti}^{-2}, k_{tj}^{-2})\Delta R_{ij}^2/R^2$$

$$d_{iB} = k_{ti}^{-2}$$

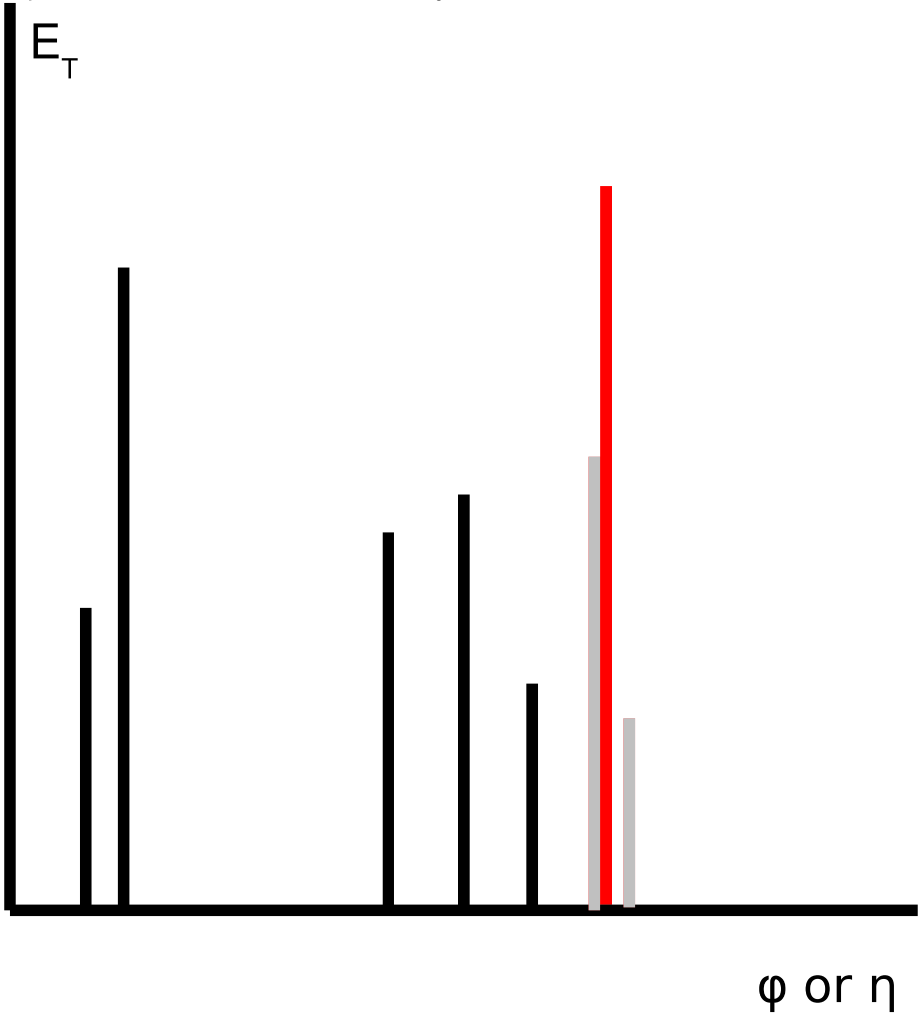


Clustering Algorithms

CTEQ-MCnet school 2008
Gavin Salam Lectures on Jets

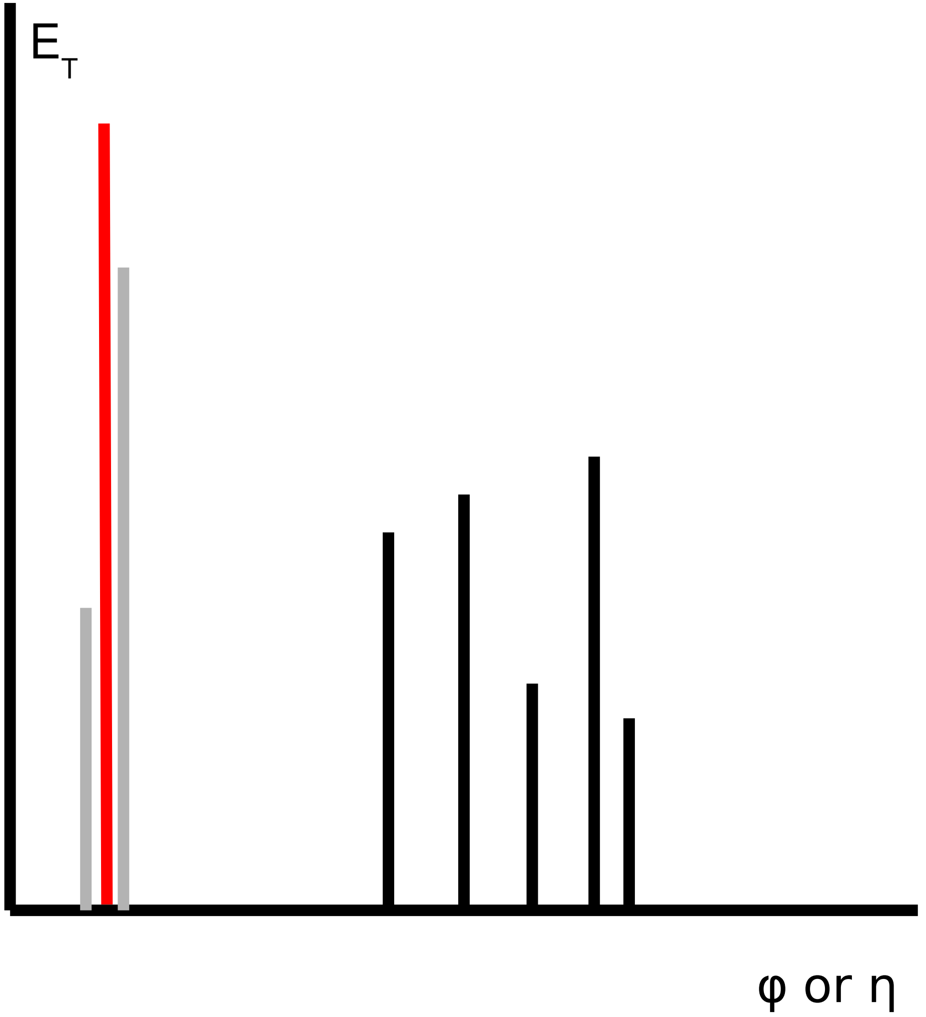
Kt

(Catani/Dokshitzer/Seymour/Webber - S.Ellis/Soper)



AntiKt

(Cacciari/Salam/Soyez)

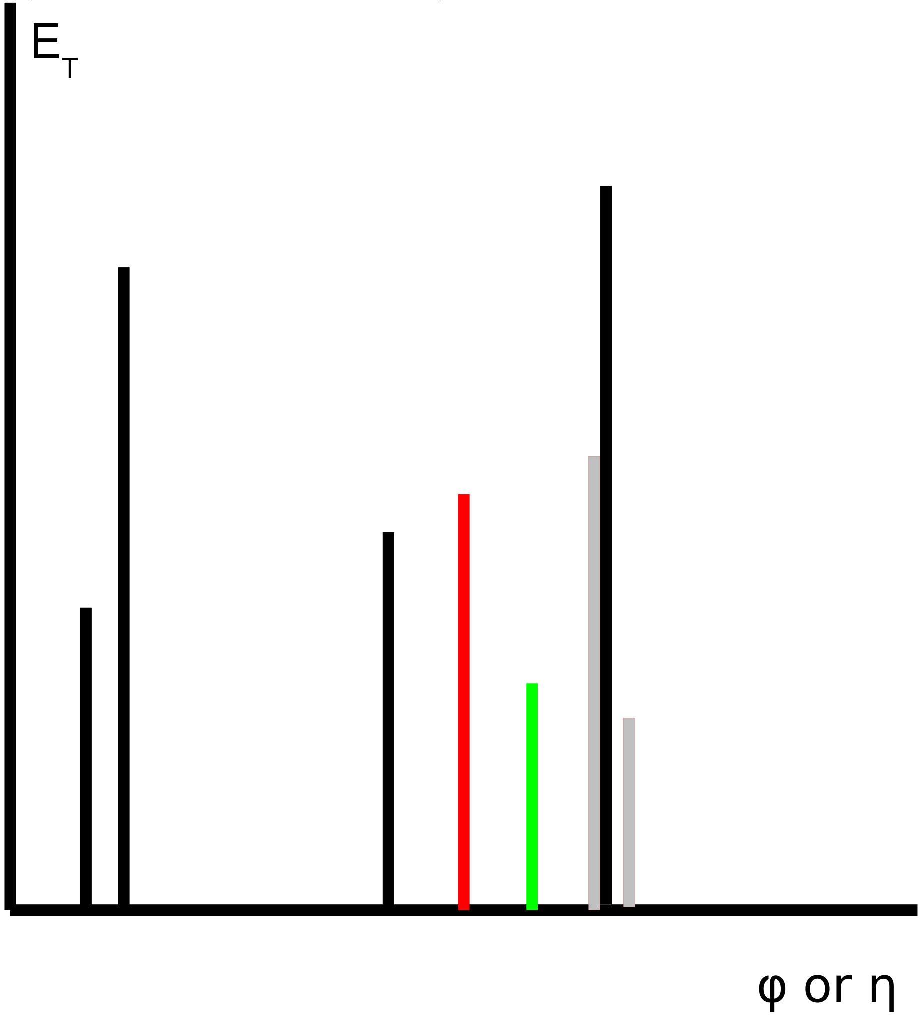


Clustering Algorithms

CTEQ-MCnet school 2008
Gavin Salam Lectures on Jets

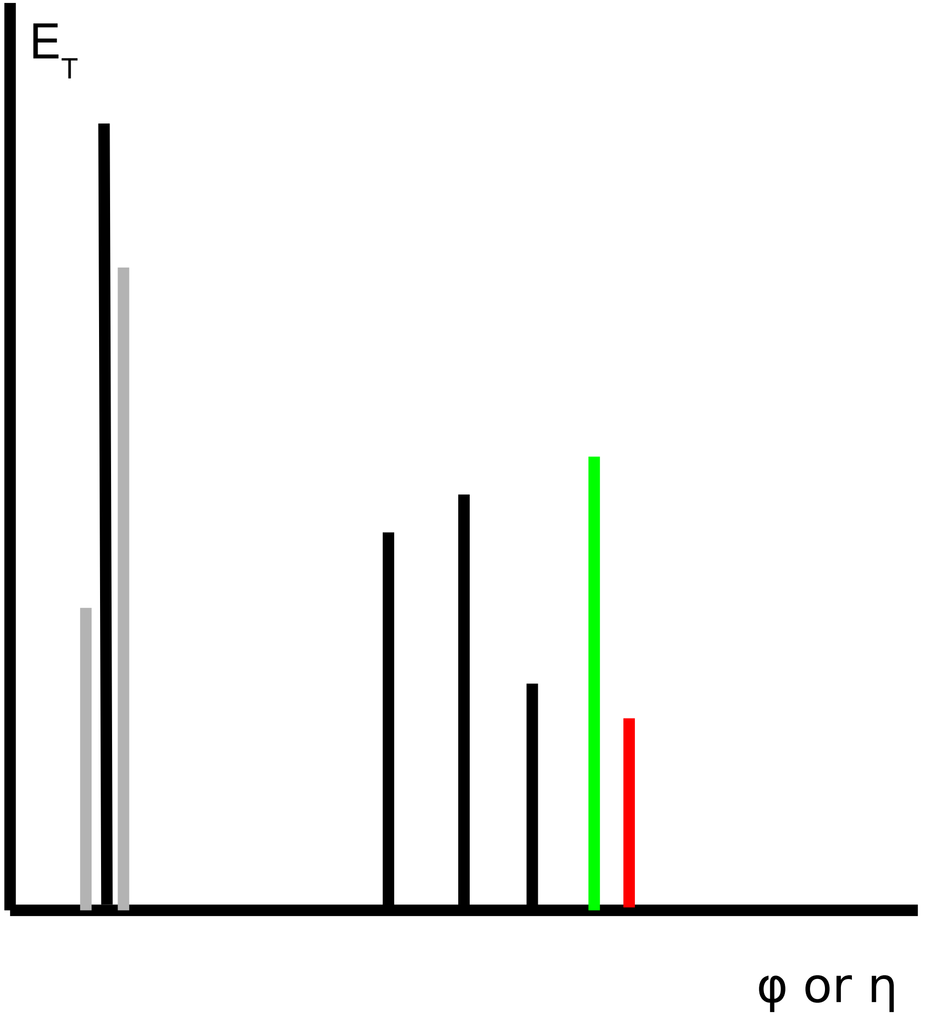
Kt

(Catani/Dokshitzer/Seymour/Webber - S.Ellis/Soper)



AntiKt

(Cacciari/Salam/Soyez)

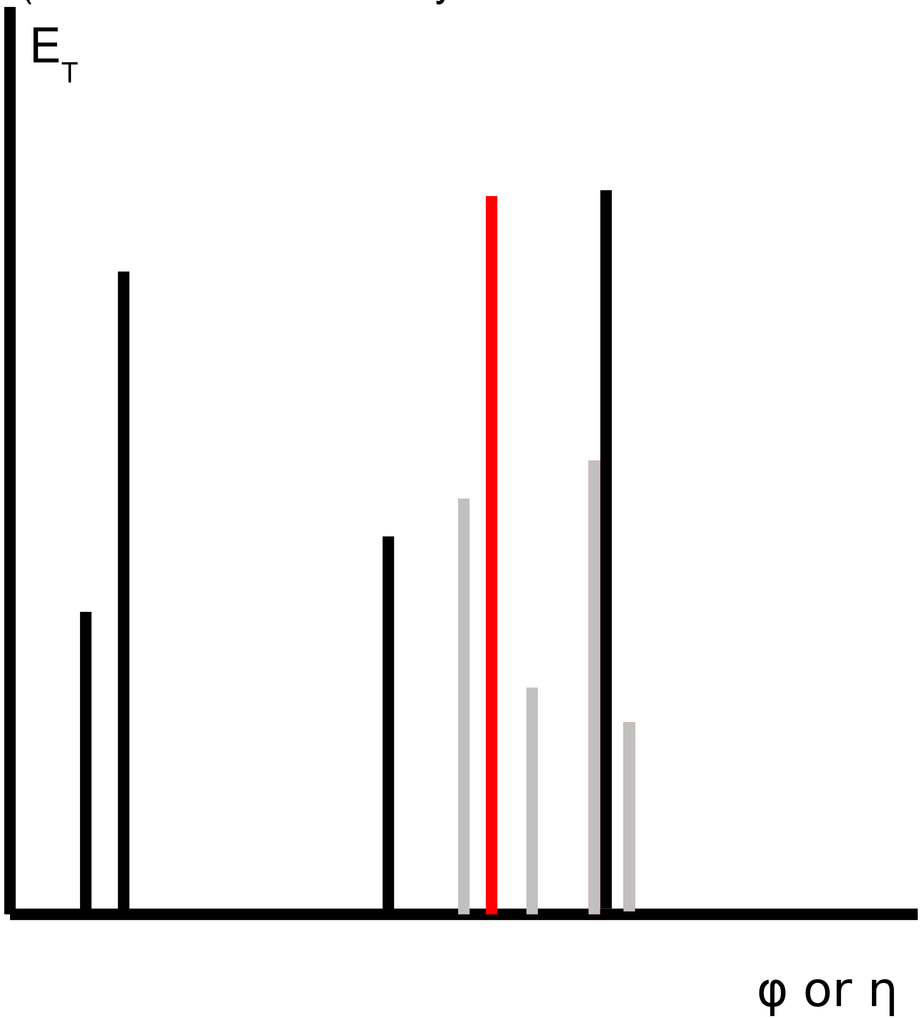


Clustering Algorithms

CTEQ-MCnet school 2008
Gavin Salam Lectures on Jets

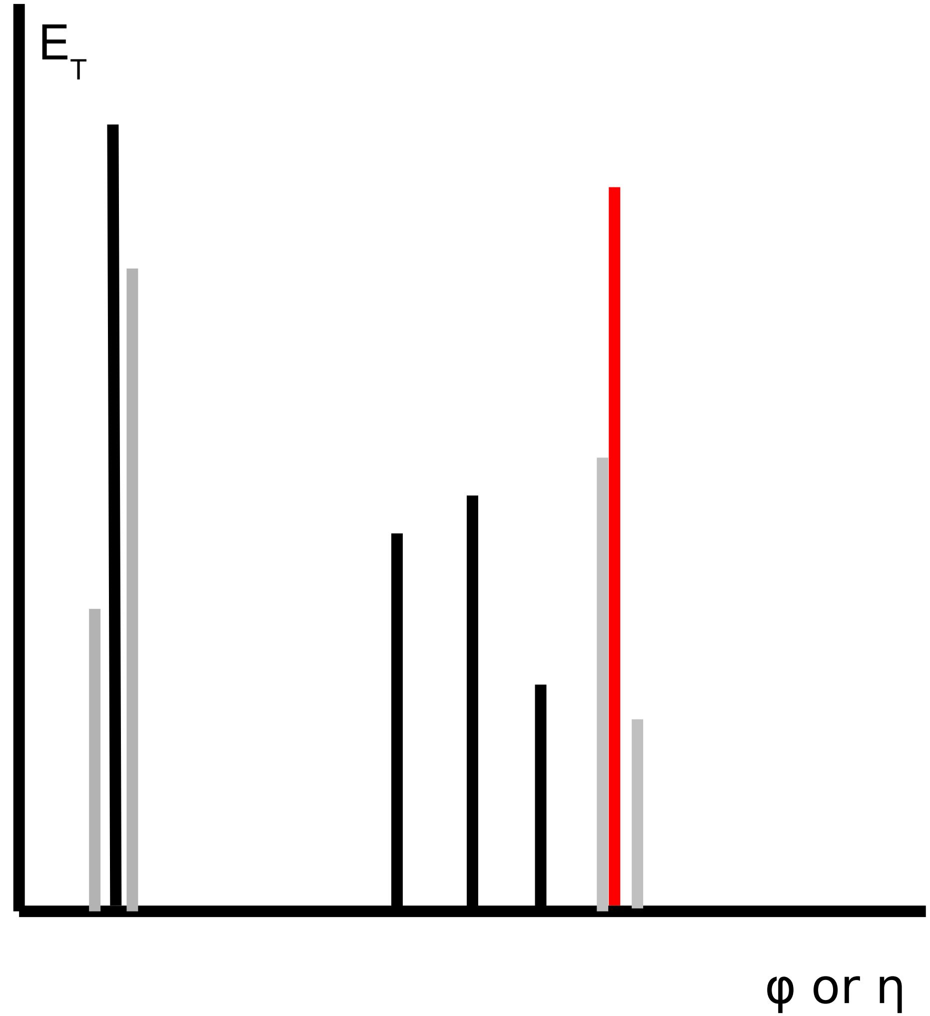
Kt

(Catani/Dokshitzer/Seymour/Webber - S.Ellis/Soper)



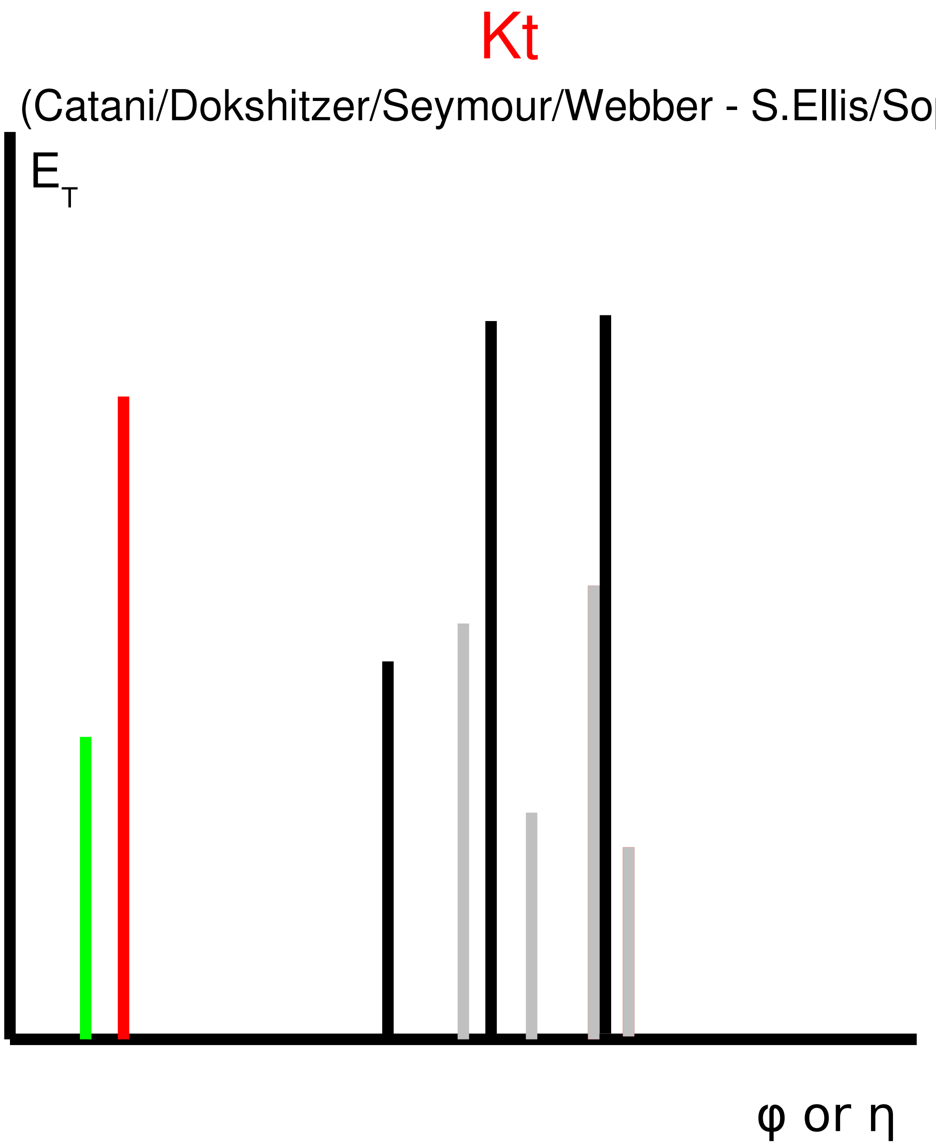
AntiKt

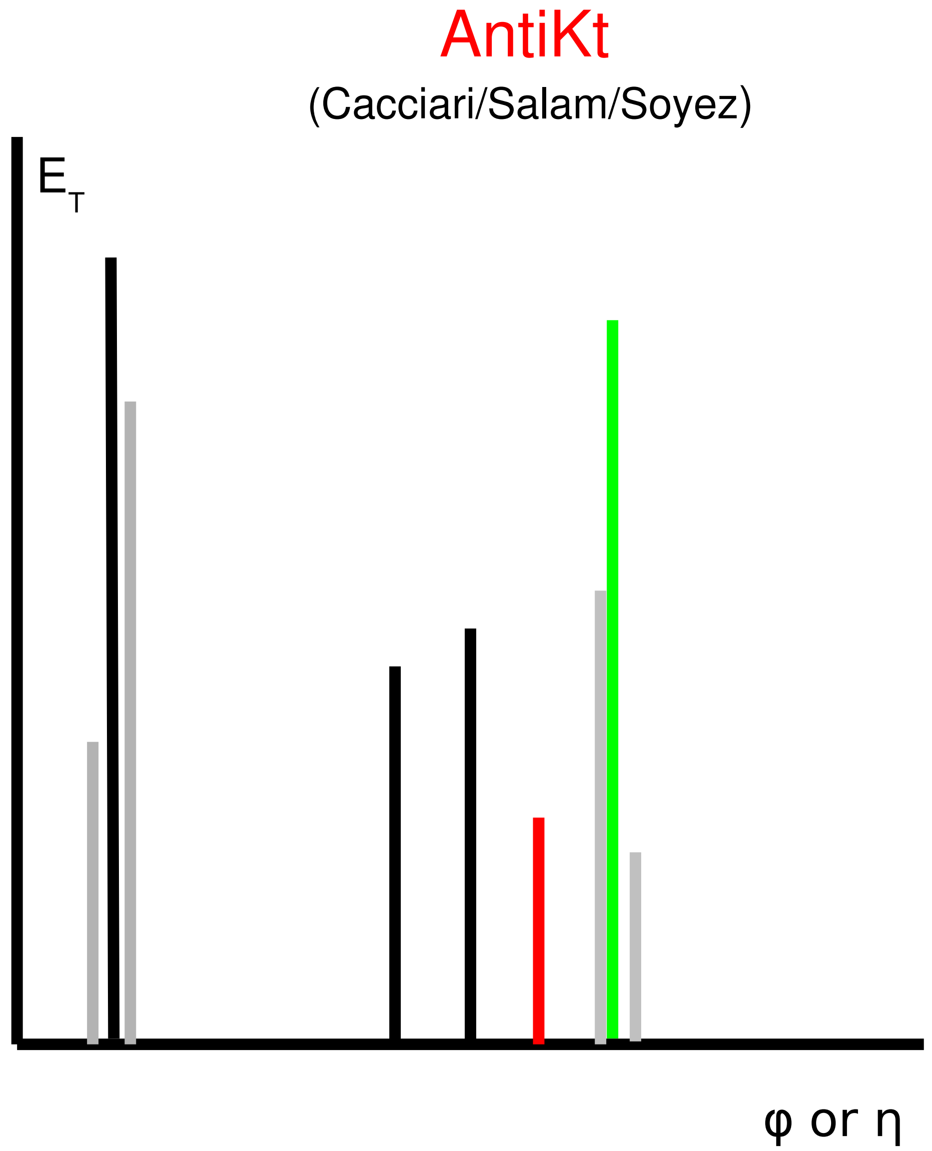
(Cacciari/Salam/Soyez)



Clustering Algorithms

CTEQ-MCnet school 2008
Gavin Salam Lectures on Jets



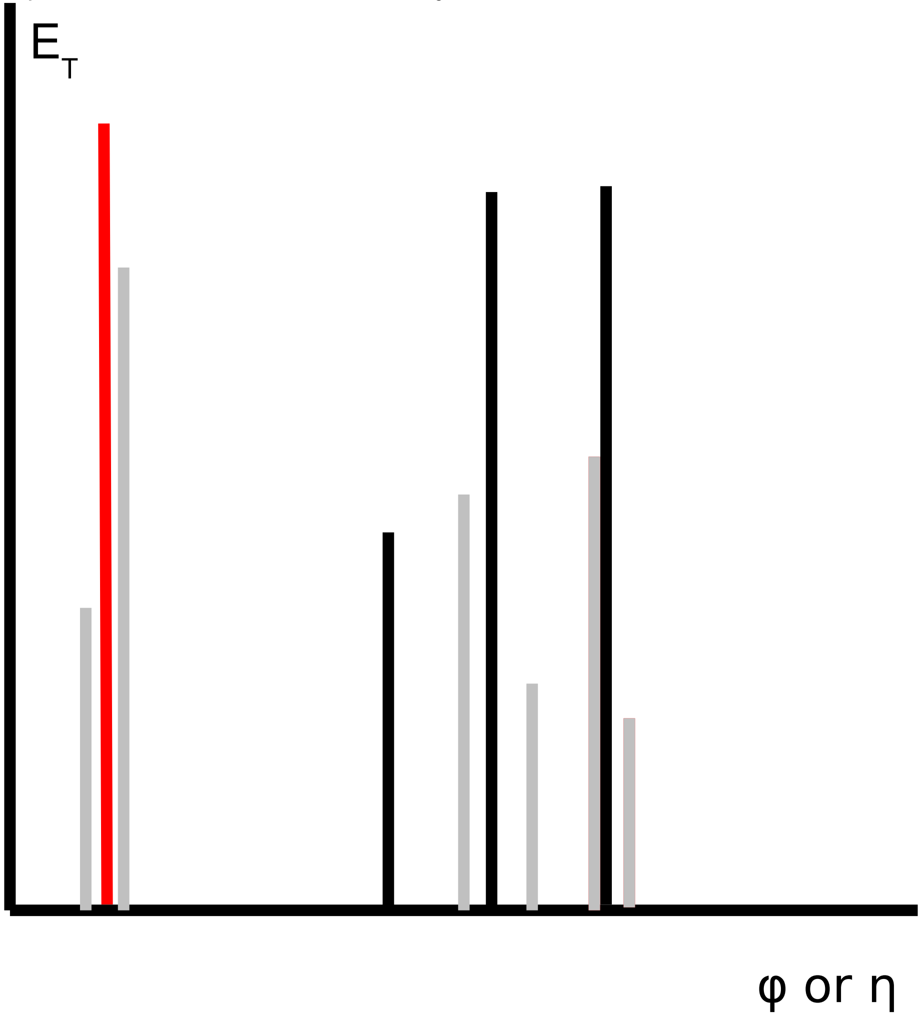


Clustering Algorithms

CTEQ-MCnet school 2008
Gavin Salam Lectures on Jets

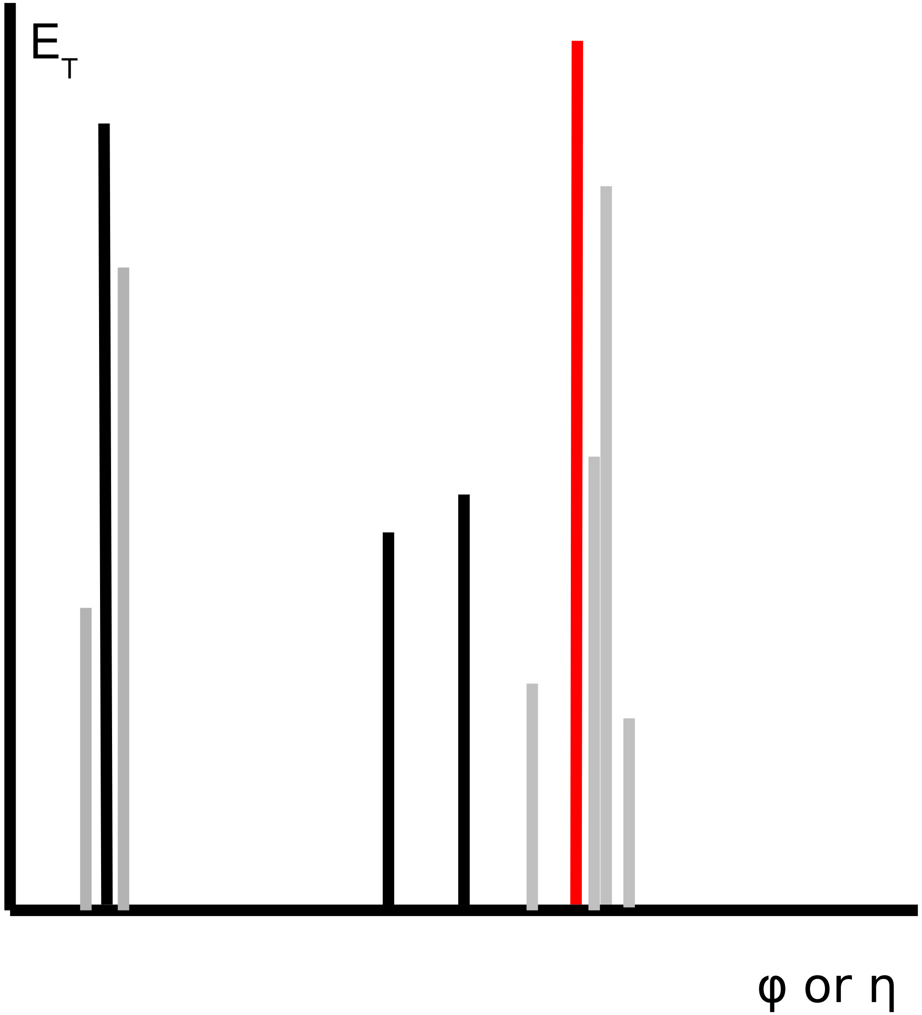
Kt

(Catani/Dokshitzer/Seymour/Webber - S.Ellis/Soper)



AntiKt

(Cacciari/Salam/Soyez)

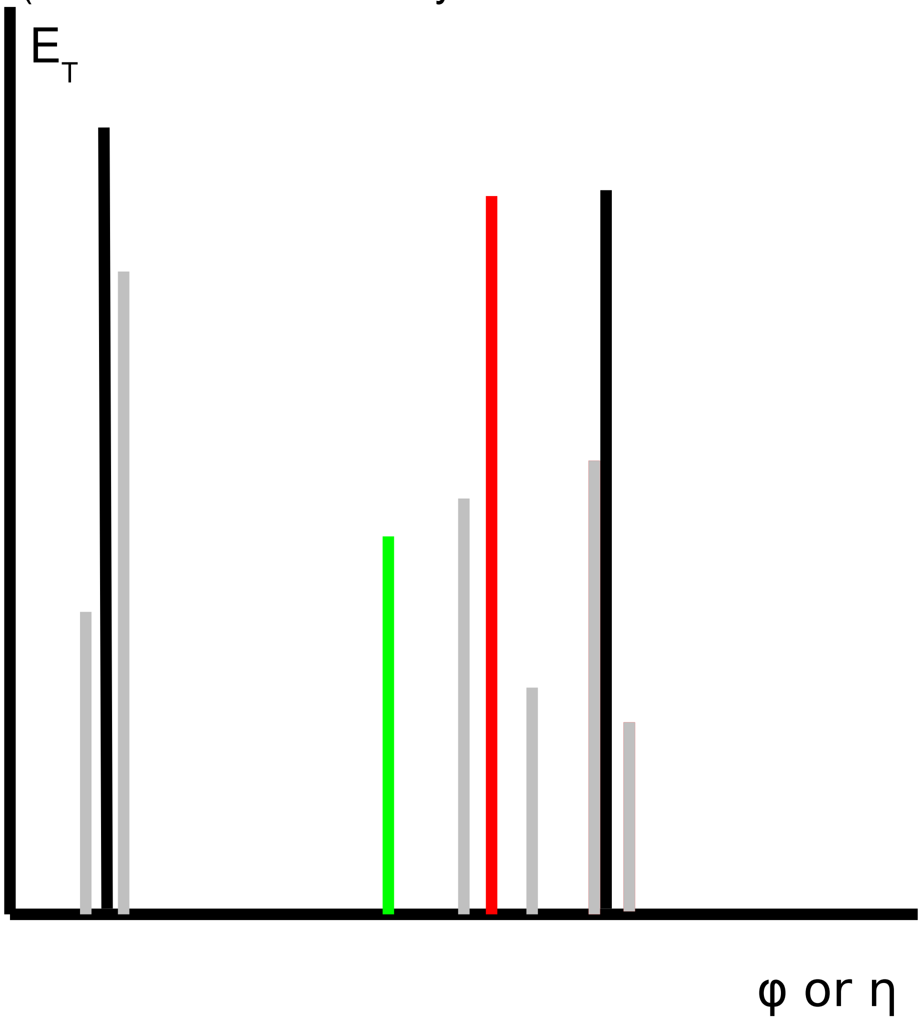


Clustering Algorithms

CTEQ-MCnet school 2008
Gavin Salam Lectures on Jets

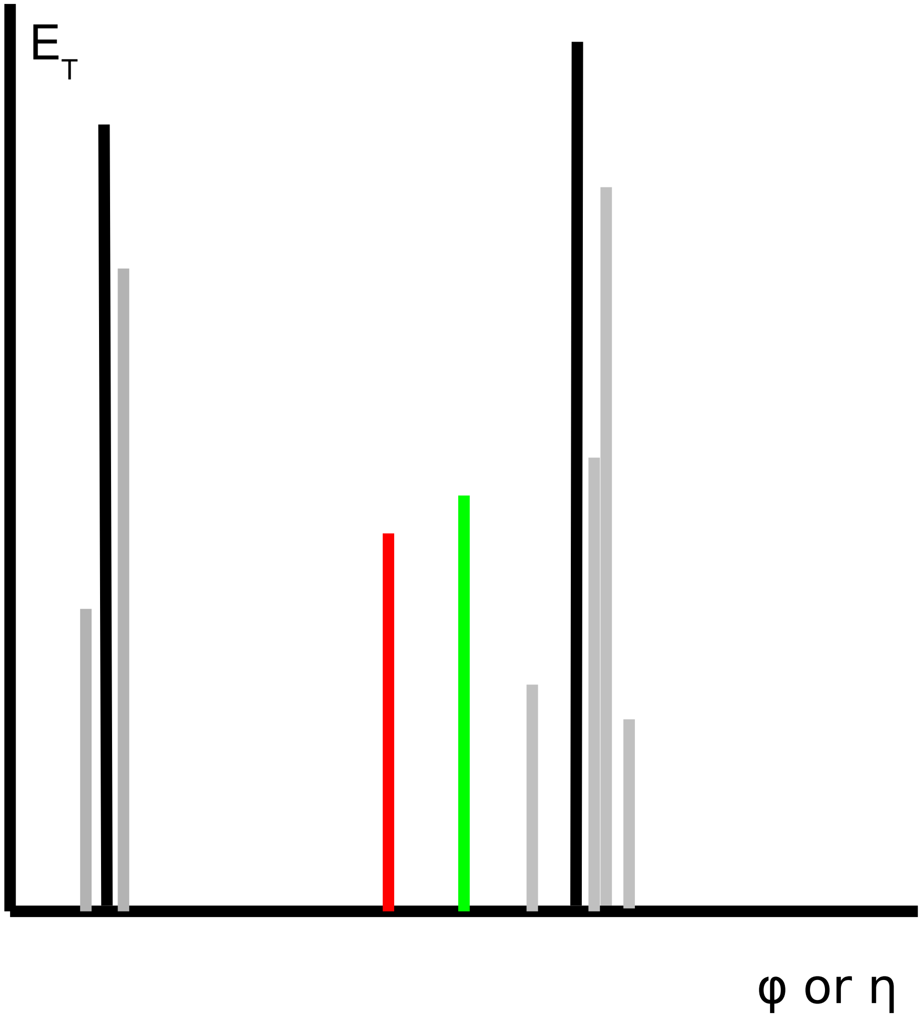
Kt

(Catani/Dokshitzer/Seymour/Webber - S.Ellis/Soper)



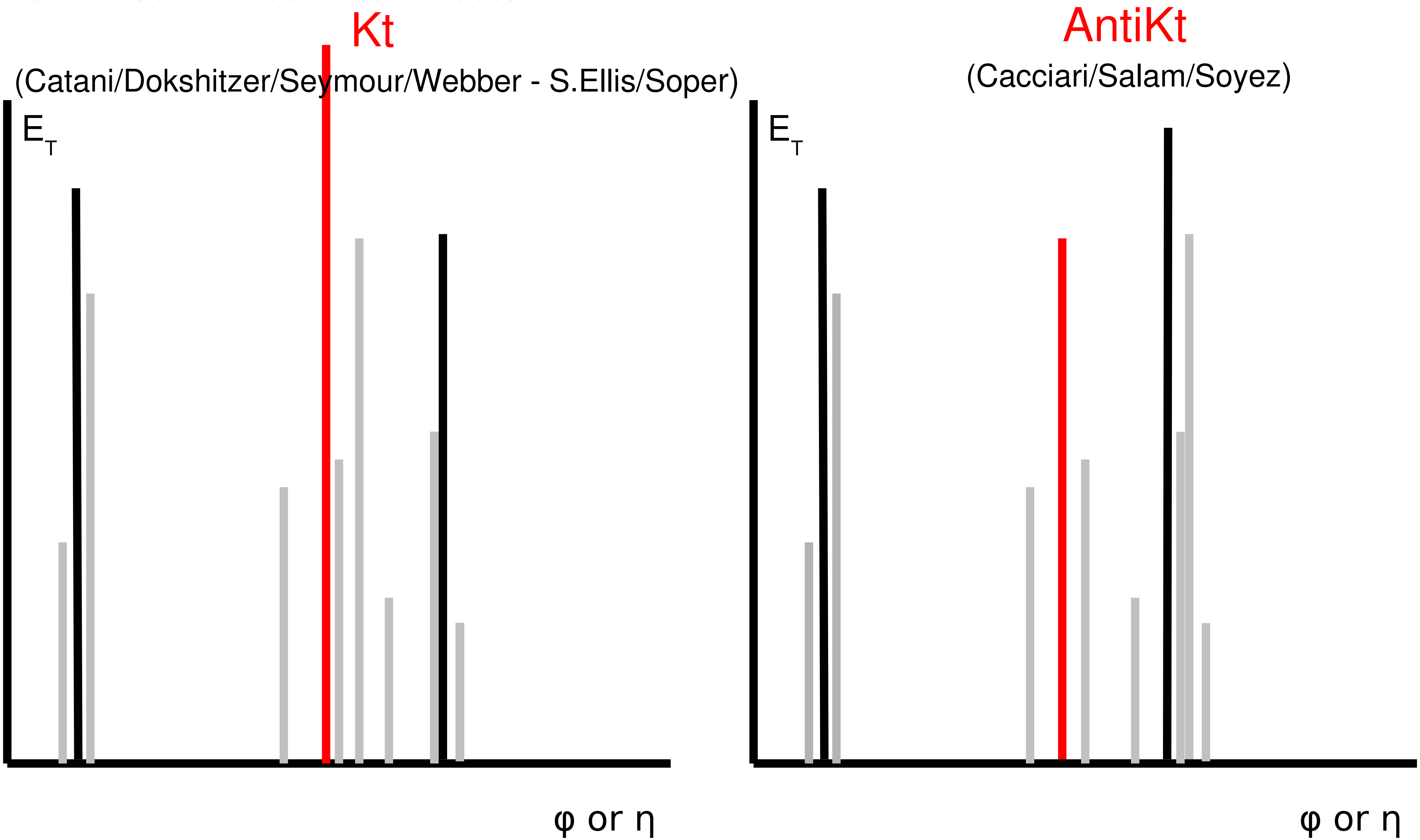
AntiKt

(Cacciari/Salam/Soyez)



Clustering Algorithms

CTEQ-MCnet school 2008
Gavin Salam Lectures on Jets



Clustering Algorithms

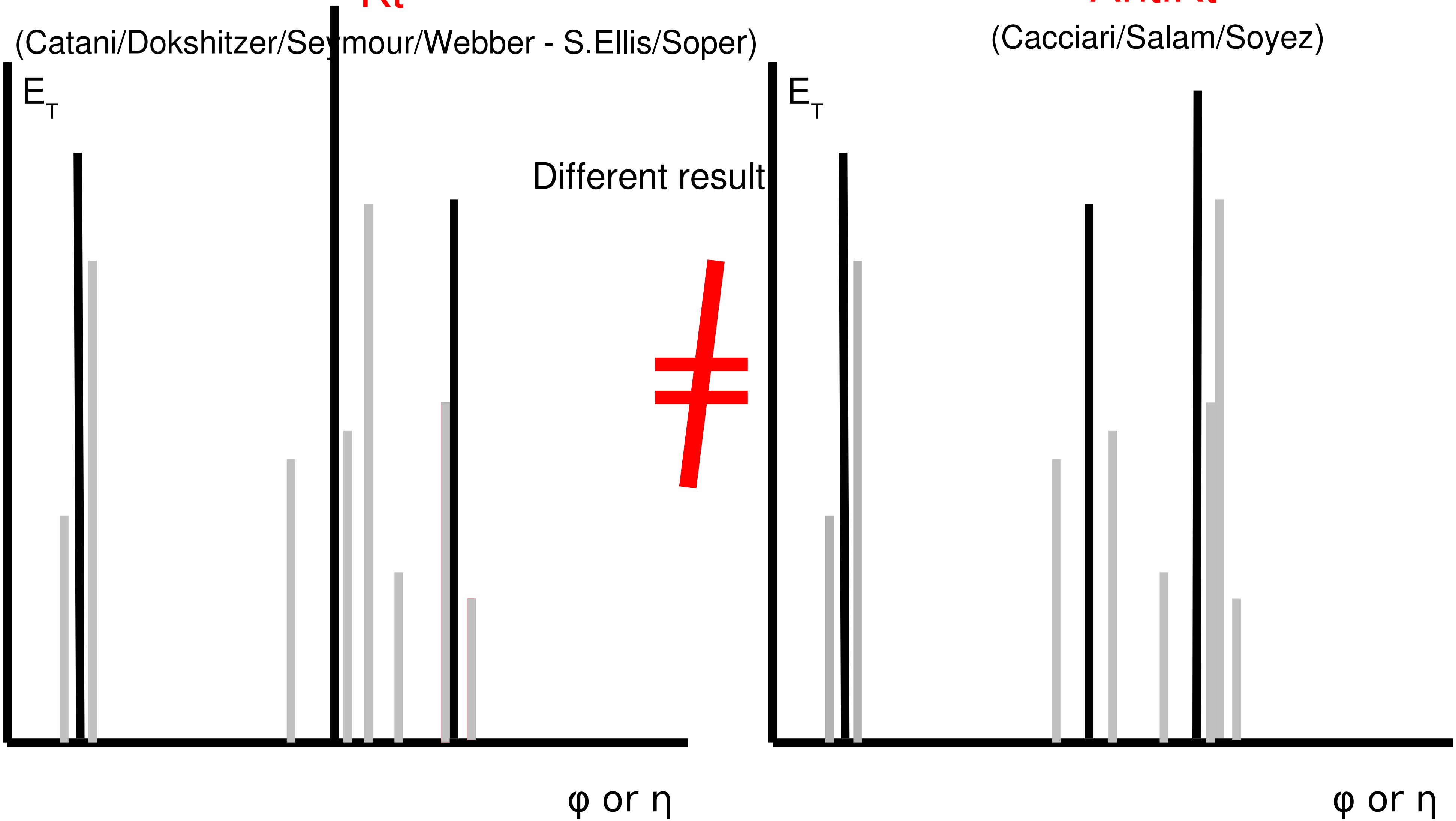
CTEQ-MCnet school 2008
Gavin Salam Lectures on Jets

Kt

(Catani/Dokshitzer/Seymour/Webber - S.Ellis/Soper)

AntiKt

(Cacciari/Salam/Soyez)



New Jet Algorithms

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p})\Delta R_{ij}^2/R^2$$

AntiKt (Cacciari/Salam/Soyez, '08 [+ Delsart unpublished]):

Kt and AntiKt have **different peculiarities** due to the exponent **p** in d_{ij} .

AntiKt Successively **merges objects** with **high relative Kt.**

$\mathbf{d_{ij}}$ is determined solely by the **Kt of the harder of i & j**, and **by** $\Delta R_{ij}/R$.

Soft stuff within R of a high Kt object will be merged with it.

If two hard jets are close the energy will be shared based on ΔR_{ij}.

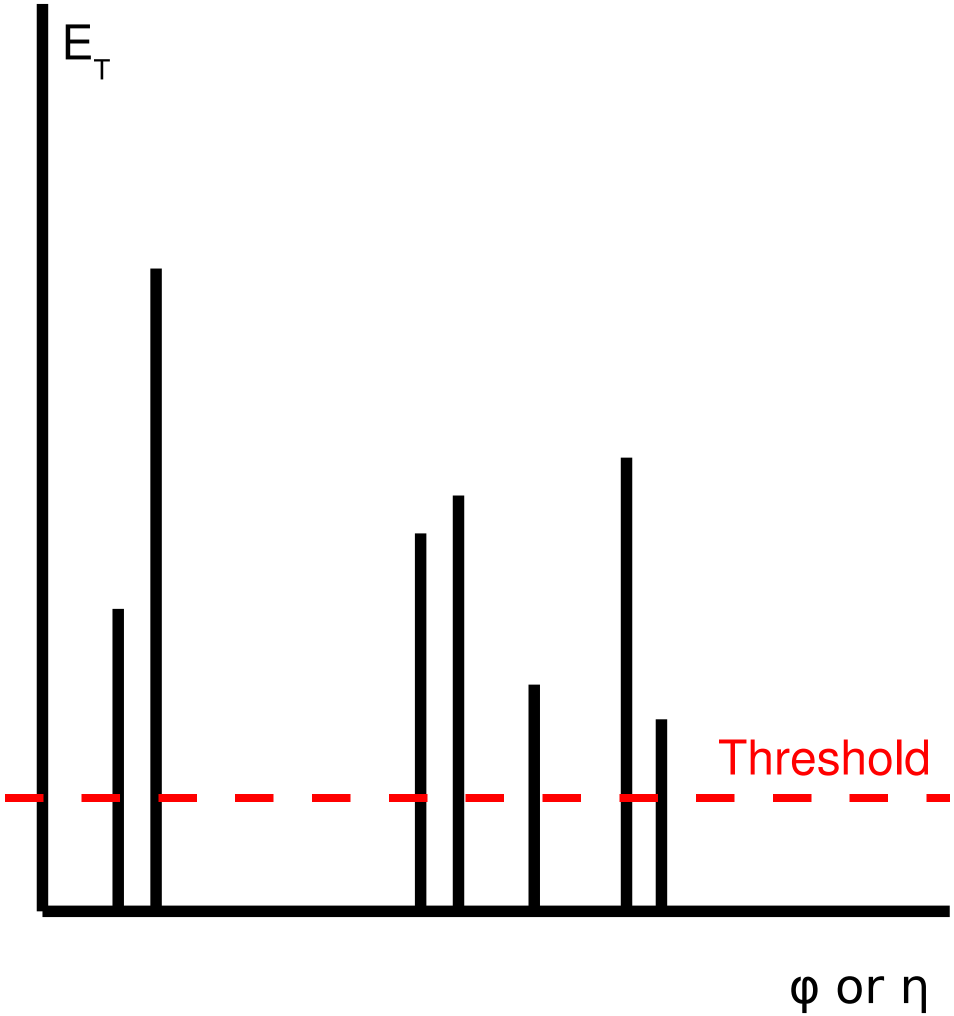
Shape of jet is unaffected by soft radiation (~ circle around the high Kt object).

Cambridge Aachen (Dokshitzer/Leder/Moretti/Webber,'97[Cambridge] - Wobisch/Wengler,'99[Aachen])

The Cambridge Aachen Algorithms is a clustering algorithms but with p =0. This means that it uses only the geometrical distance ΔR_{ij} for the recursive clustering. Even in this case the algorithm produce different peculiarities with respect to the Kt and to the AntiKt.

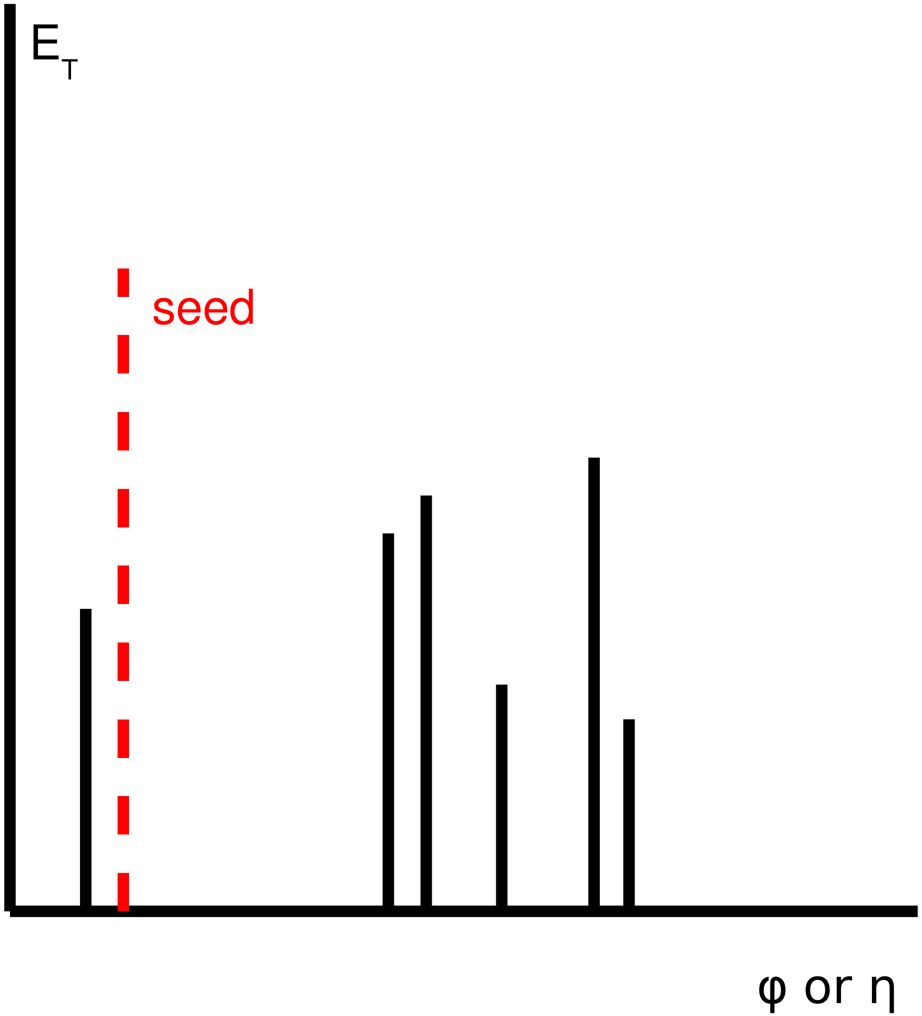
ATLAS Cone Algorithm

CTEQ-MCNet school 2008
G. Salam's Lectures on Jets



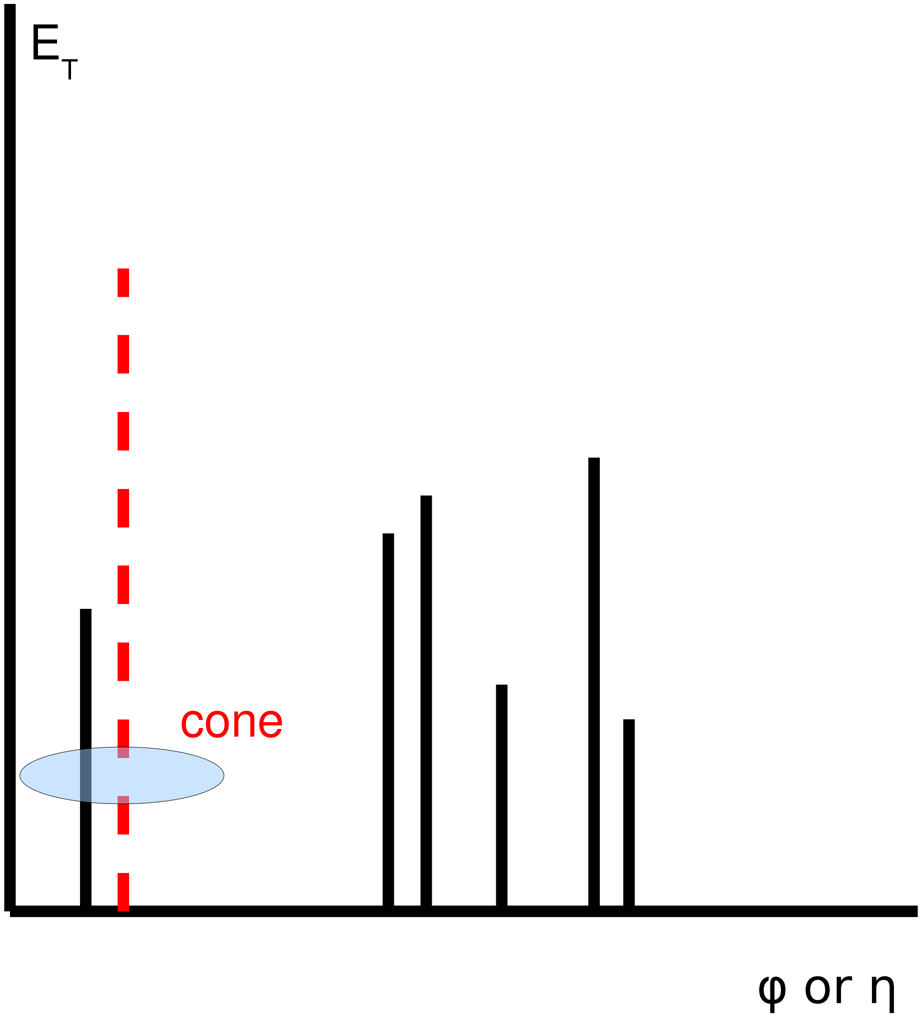
- Use every object above a certain threshold as possible seed
- Open a cone, merge the objects inside and iterate until stable cone.
- Add the stable cone to the list of protojets unless it's already there and remove the seed from the list of seeds.
- Until all seeds done.

ATLAS Cone Algorithm



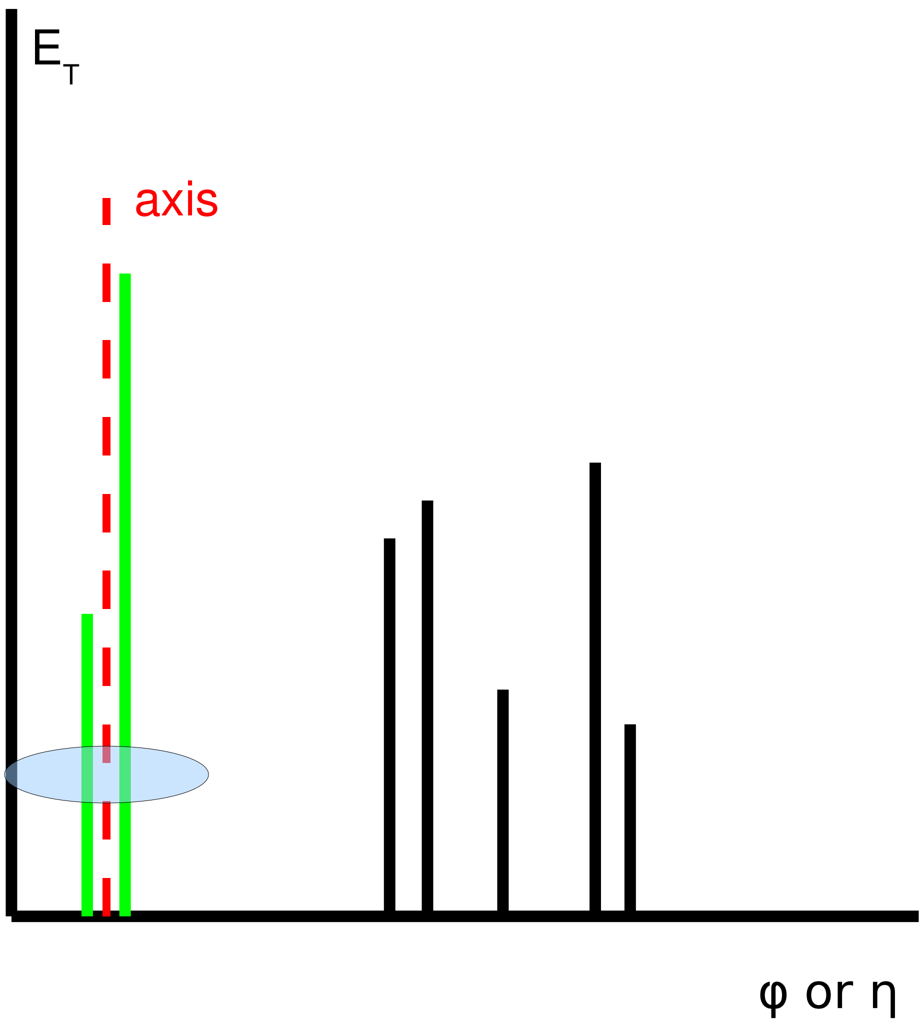
- Use every object above a certain threshold as possible seed
- Open a cone, merge the objects inside and iterate until stable cone.
- Add the stable cone to the list of protojets unless it's already there and remove the seed from the list of seeds.
- Until all seeds done.

ATLAS Cone Algorithm



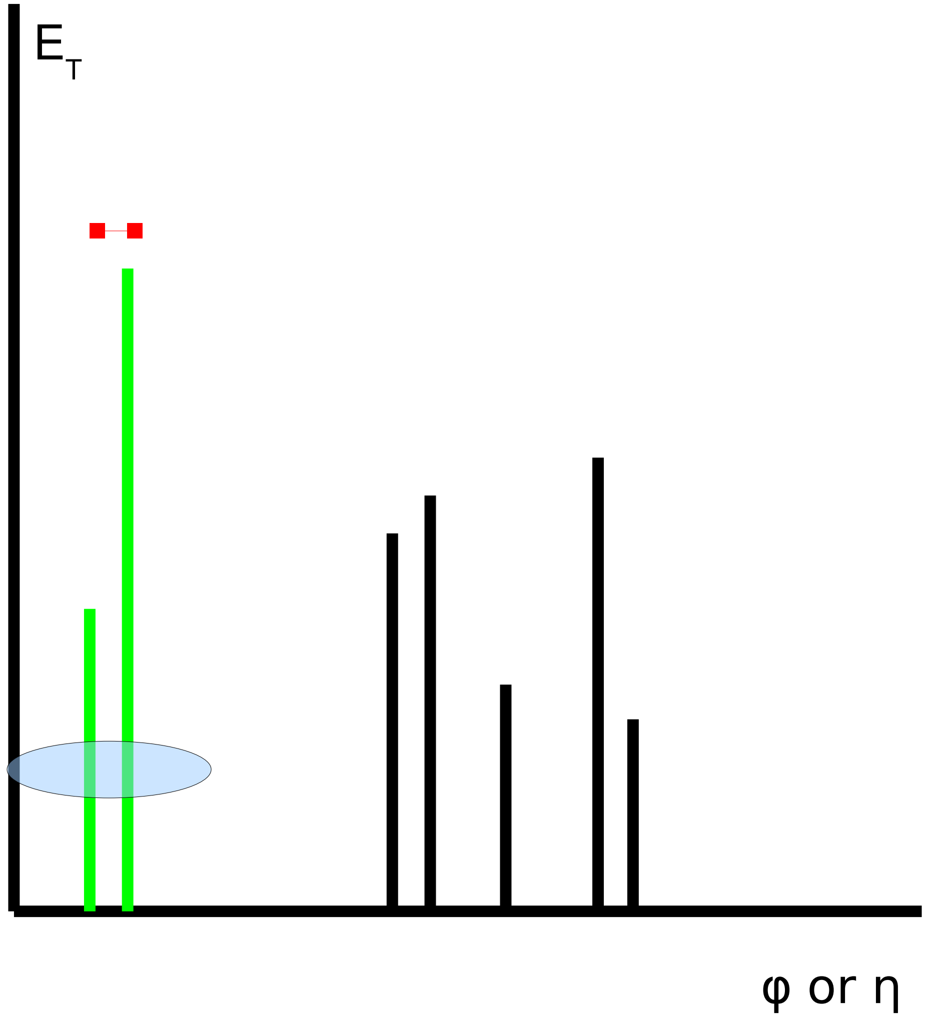
- Use every object above a certain threshold as possible seed
- Open a cone, merge the objects inside and iterate until stable cone.
- Add the stable cone to the list of protojets unless it's already there and remove the seed from the list of seeds.
- Until all seeds done.

ATLAS Cone Algorithm



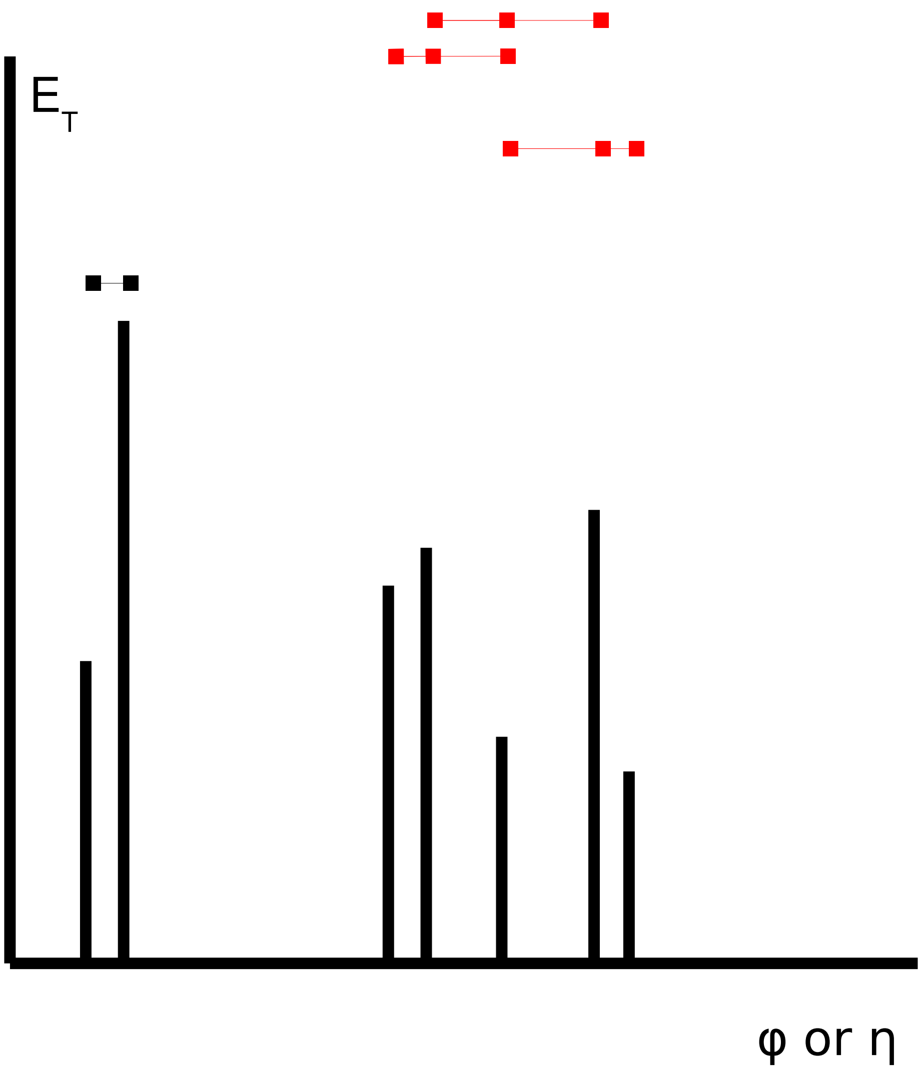
- Use every object above a certain threshold as possible seed
- Open a cone, merge the objects inside and iterate until stable cone.
- Add the stable cone to the list of protojets unless it's already there and remove the seed from the list of seeds.
- Until all seeds done.

ATLAS Cone Algorithm



- Use every object above a certain threshold as possible seed
- Open a cone, merge the objects inside and iterate until stable cone.
- Add the stable cone to the list of protojets unless it's already there and remove the seed from the list of seeds.
- Until all seeds done.

ATLAS Cone Algorithm

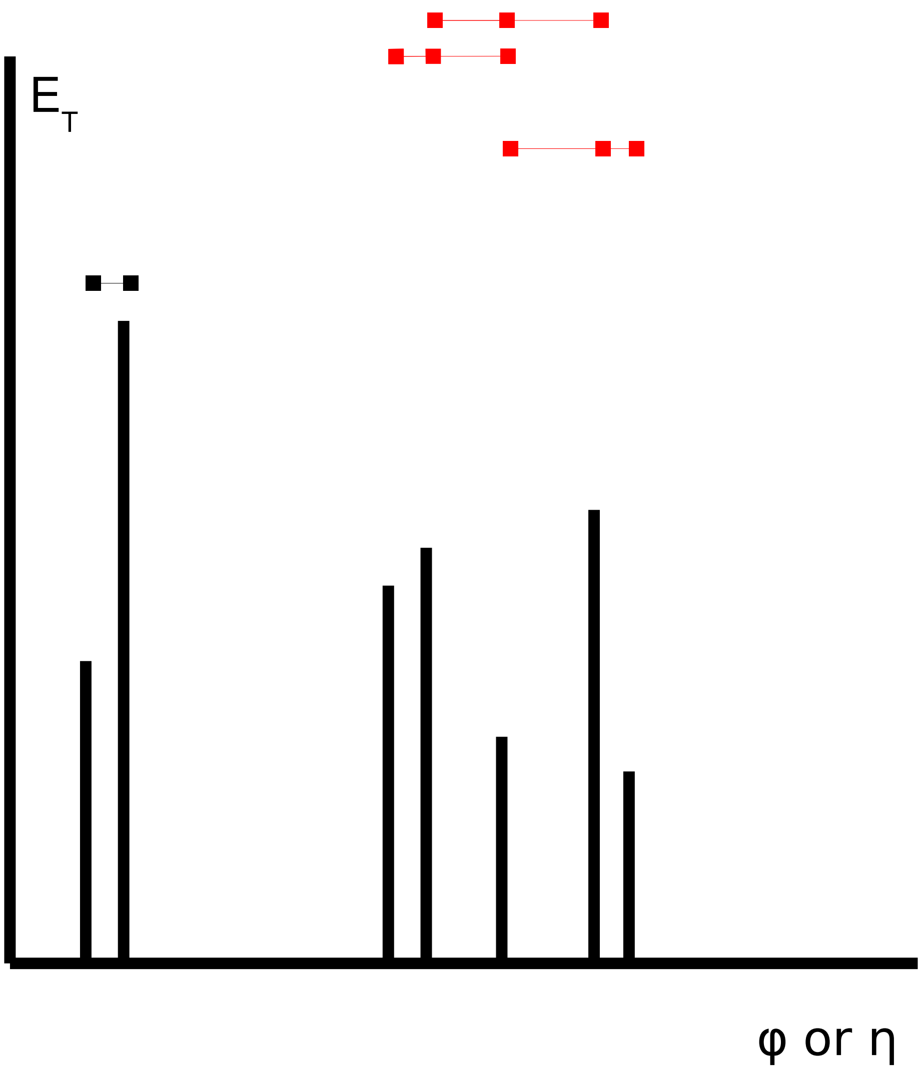


- Use every object above a certain threshold as possible seed
- Open a cone, merge the objects inside and iterate until stable cone.
- Add the stable cone to the list of protojets unless it's already there and remove the seed from the list of seeds.
- Until all seeds done.

Note: protojets overlap. Certain objects appear in many protojets. Must resolve the overlaps.

- Use a split–merge procedure.

ATLAS Cone Algorithm



Split and merge (SM) procedure

but common to most cone algorithms with SM.

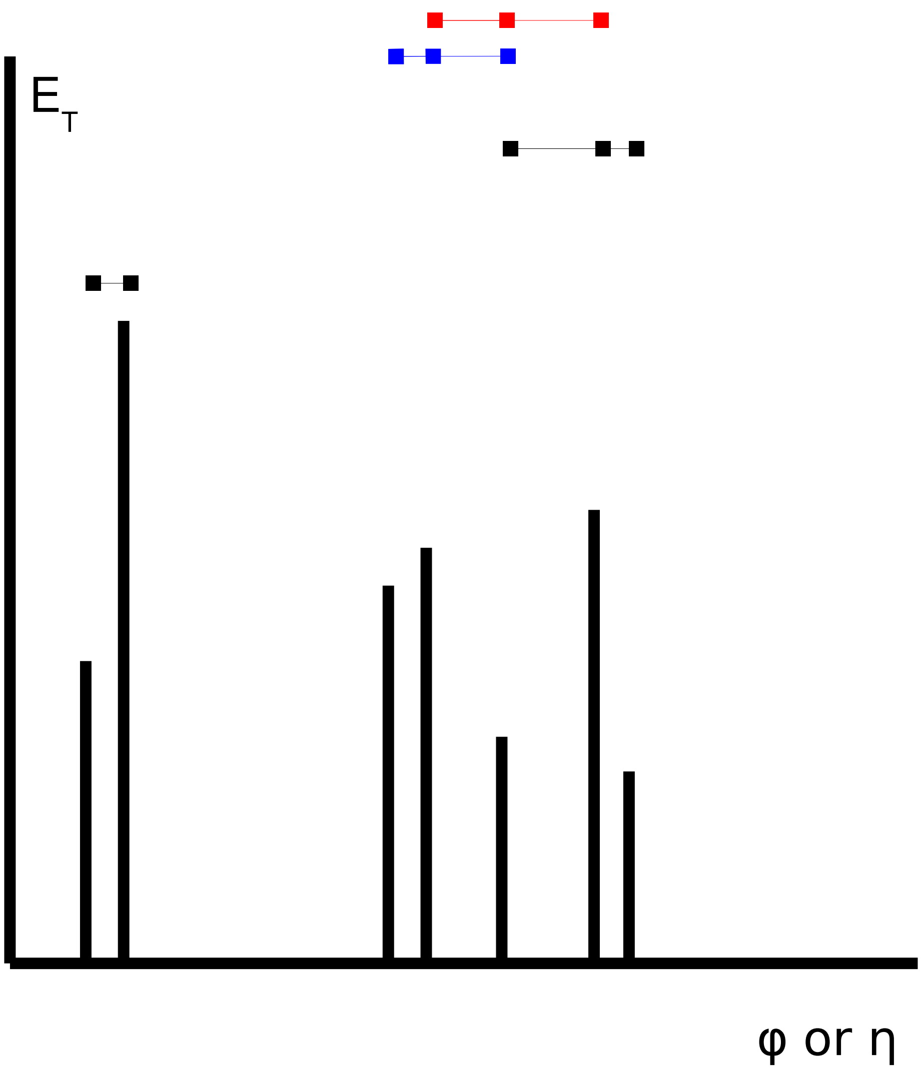
- Introduce overlap threshold f.
- Identify hardest protojet (PJ), p1. Find hardest PJ that shares constituents with it, p2.
- Calculated overlap fraction (O):

$$O = P_{T,shared} / P_{T,2}$$

if O < f , split along axis at center of two PJs
if O > f merge the two PJs to one PJ.

- If there is no overlap, PJ → jet.
- Repeat. . .

ATLAS Cone Algorithm



Split and merge (SM) procedure

but common to most cone algorithms with SM.

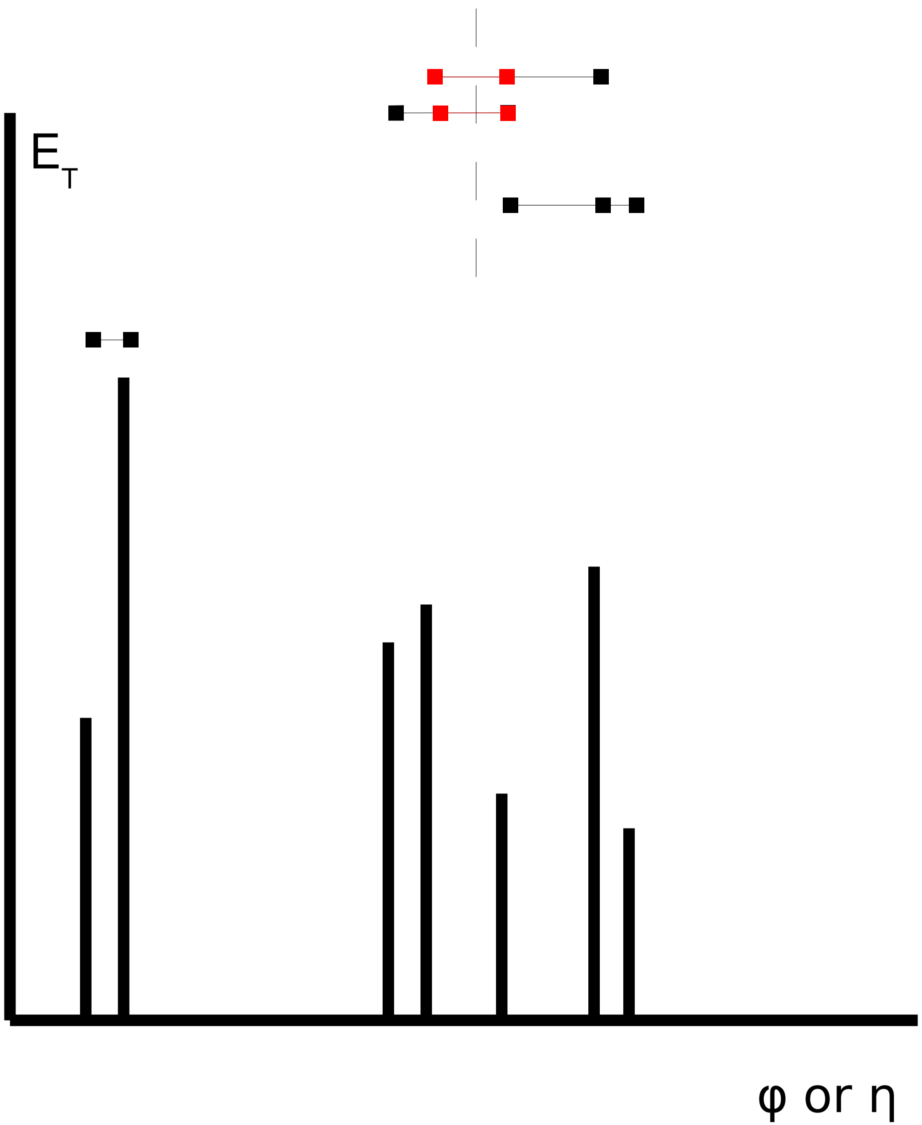
- Introduce overlap threshold f.
- Identify hardest protojet (PJ), p1.
  Find hardest PJ that shares constituents with it, p2.
- Calculated overlap fraction (O):

$$O = P_{T,shared} / P_{T,2}$$

if O < f , split along axis at center of two PJs
if O > f merge the two PJs to one PJ.

- If there is no overlap, PJ → jet.
- Repeat. . .

ATLAS Cone Algorithm



Split and merge (SM) procedure

but common to most cone algorithms with SM.

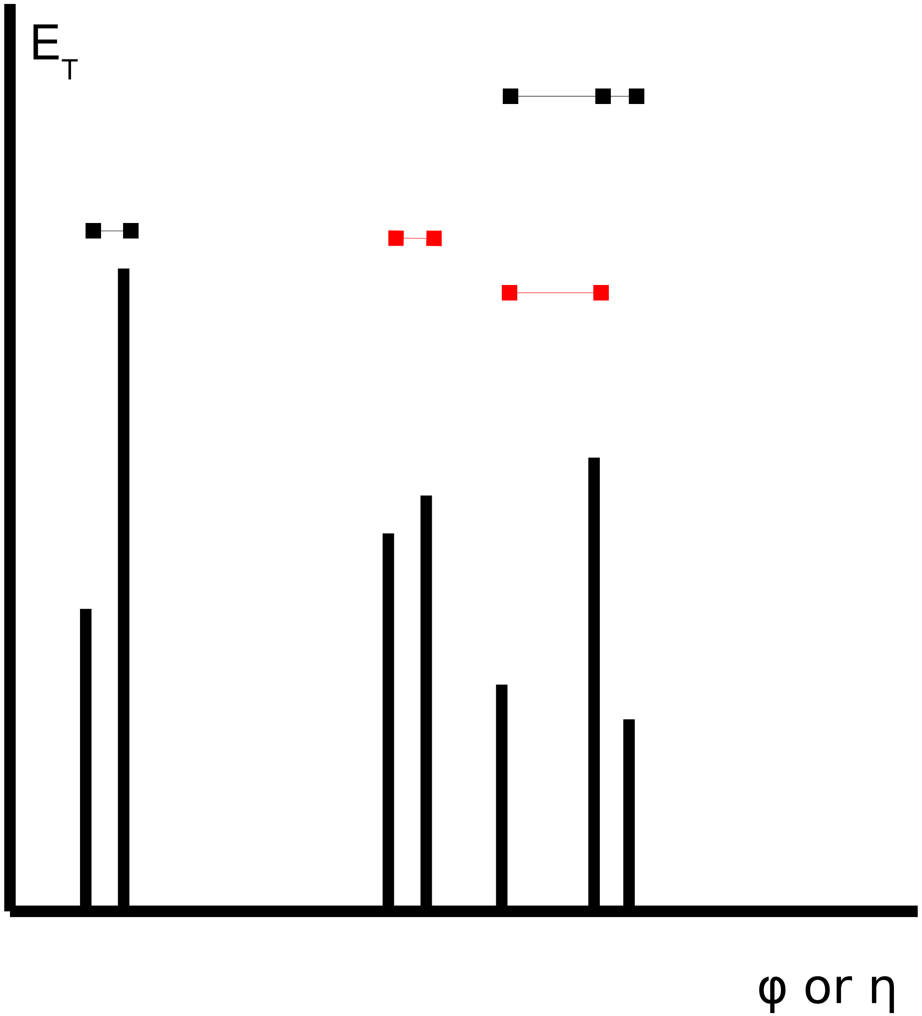
- Introduce overlap threshold f.
- Identify hardest protojet (PJ), p1. Find hardest PJ that shares constituents with it, p2.
- Calculated overlap fraction (O):

$$O = P_{T,shared} / P_{T,2}$$

if $O < f$, split along axis at center of two PJs
if $O > f$ merge the two PJs to one PJ.

- If there is no overlap, PJ → jet.
- Repeat. . .

ATLAS Cone Algorithm



Split and merge (SM) procedure

but common to most cone algorithms with SM.

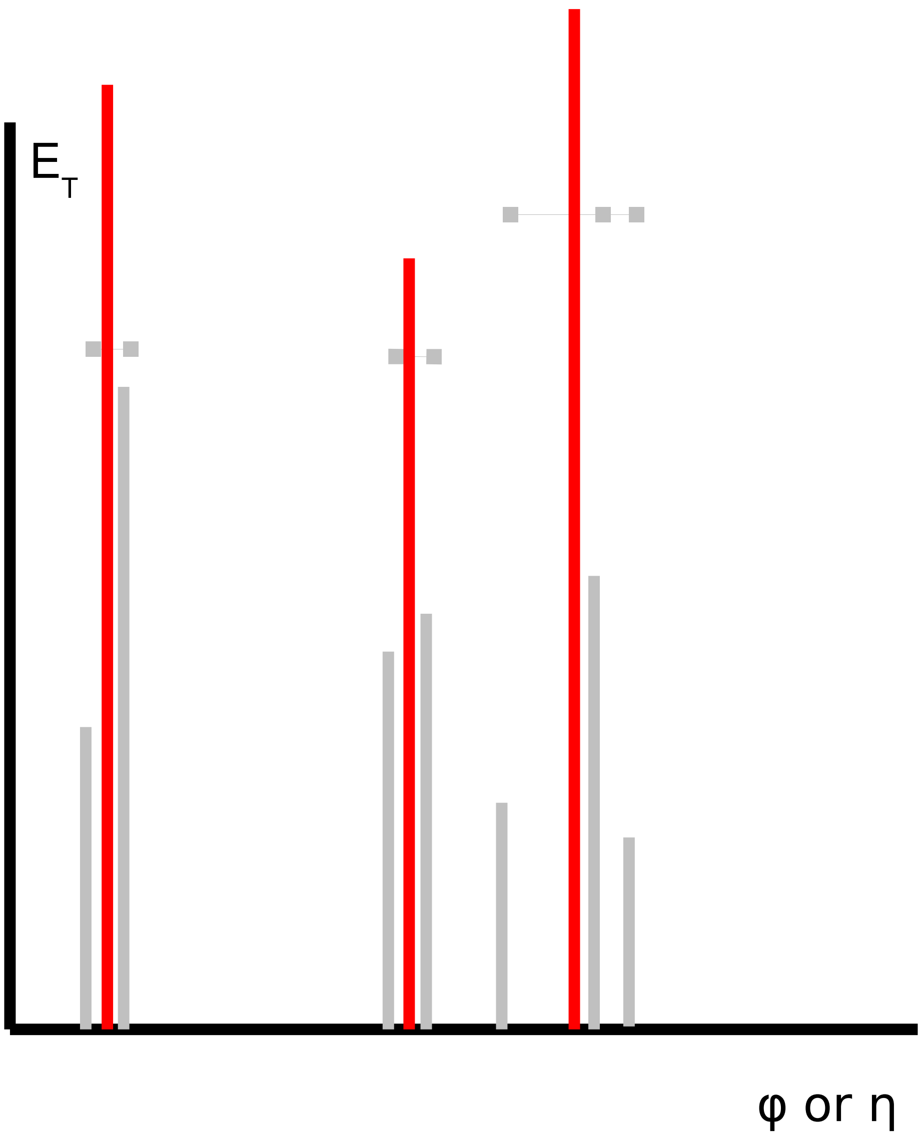
- Introduce overlap threshold f.
- Identify hardest protojet (PJ), p1. Find hardest PJ that shares constituents with it, p2.
- Calculated overlap fraction (O):

$$O = P_{T,shared} / P_{T,2}$$

if O < f , split along axis at center of two PJs
if O > f merge the two PJs to one PJ.

- If there is no overlap, PJ → jet.
- Repeat. . .

ATLAS Cone Algorithm



Split and merge (SM) procedure

but common to most cone algorithms with SM.

- Introduce overlap threshold f.
- Identify hardest protojet (PJ), p1. Find hardest PJ that shares constituents with it, p2.
- Calculated overlap fraction (O):

$$O = P_{T,shared} / P_{T,2}$$

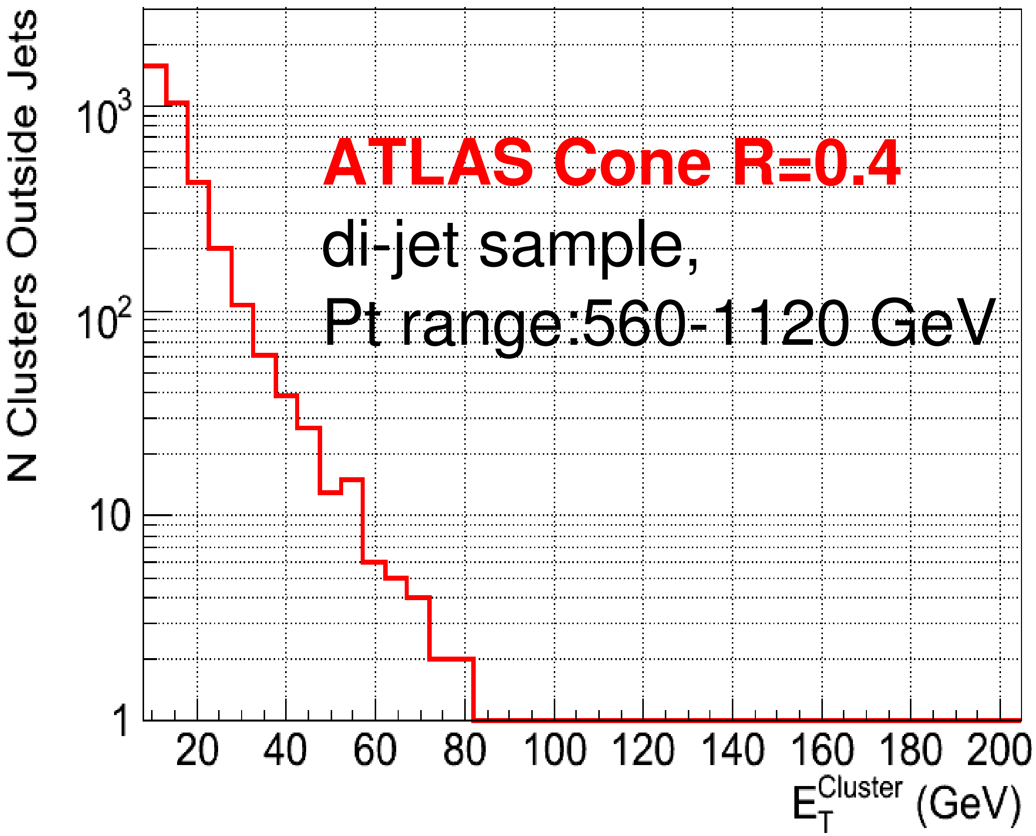
if O < f , split along axis at center of two PJs
if O > f merge the two PJs to one PJ.

- If there is no overlap, PJ → jet.
- Repeat. . .

ATLAS Cone: Dark towers

Studied the clusters not included in any jet.

Selected clusters with E_T > 10 GeV. All these clusters should belong to a jet.



For the **Kt, AntiKt, and SISCone** this is **true** (No cluster found with E_T> 10 GeV).

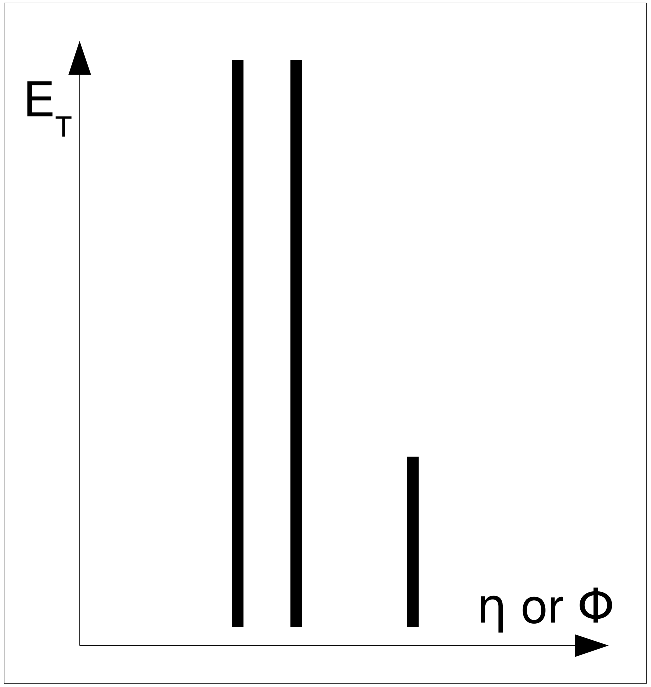
For the **Cone4** we have energetic clusters (up to E_T~70 GeV).

~3500 clusters in ~24000 events

Similar for ATLAS Cone R=0.7.

If we use the ATLAS Cone, we do not associate the energy in these clusters. **Why?**

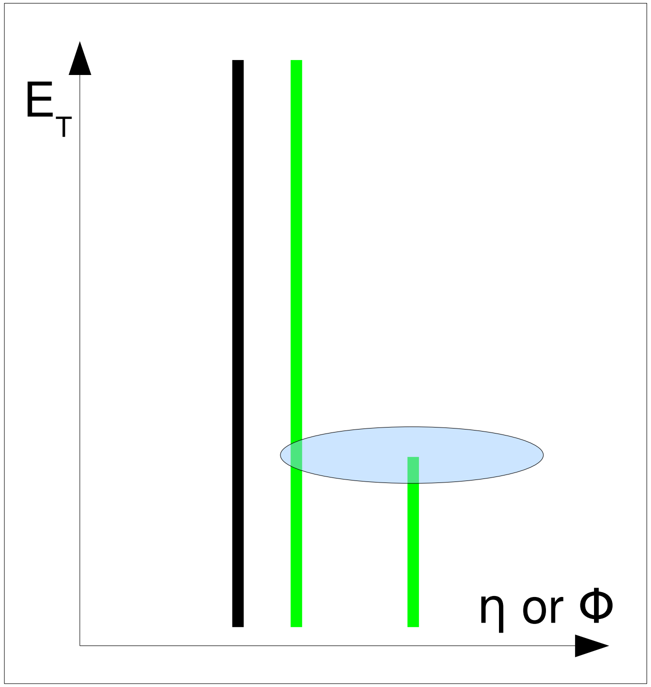
ATLAS Cone: Dark towers



The iteration of the Atlas Cone

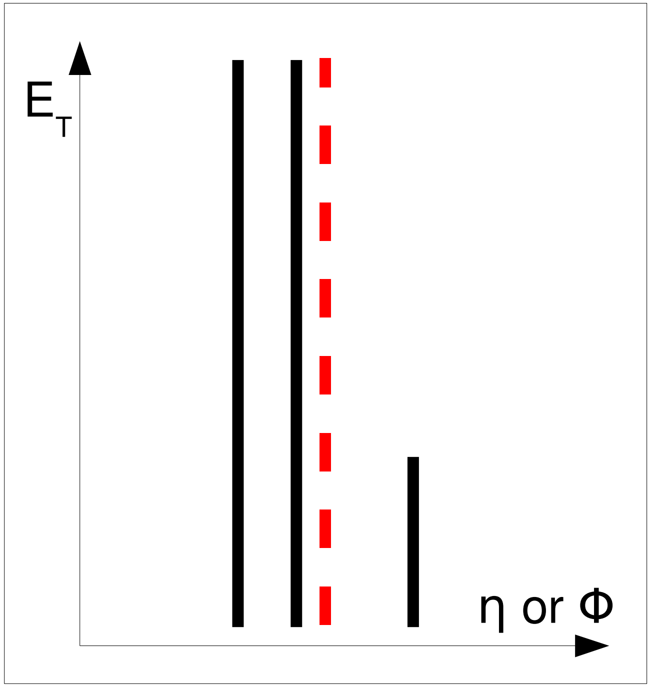
Animation Idea:
thanks to Gavin Salam

ATLAS Cone: Dark towers



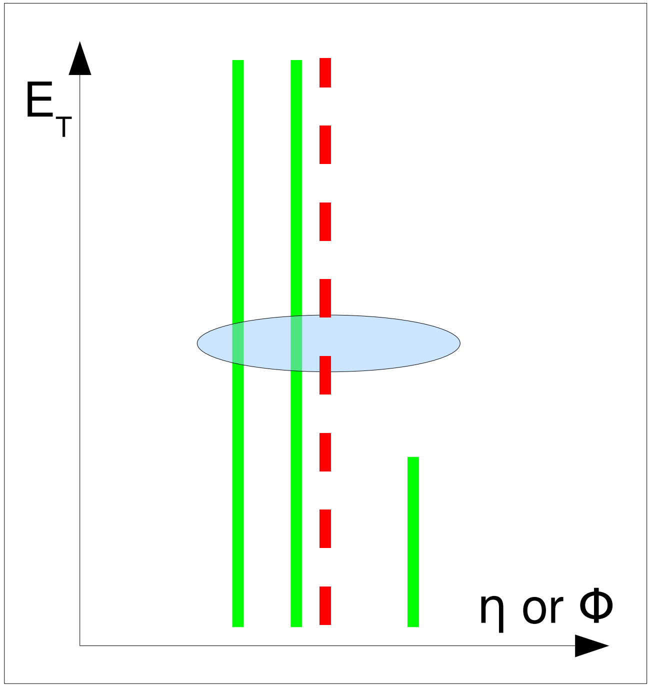
The iteration of the Atlas Cone

ATLAS Cone: Dark towers



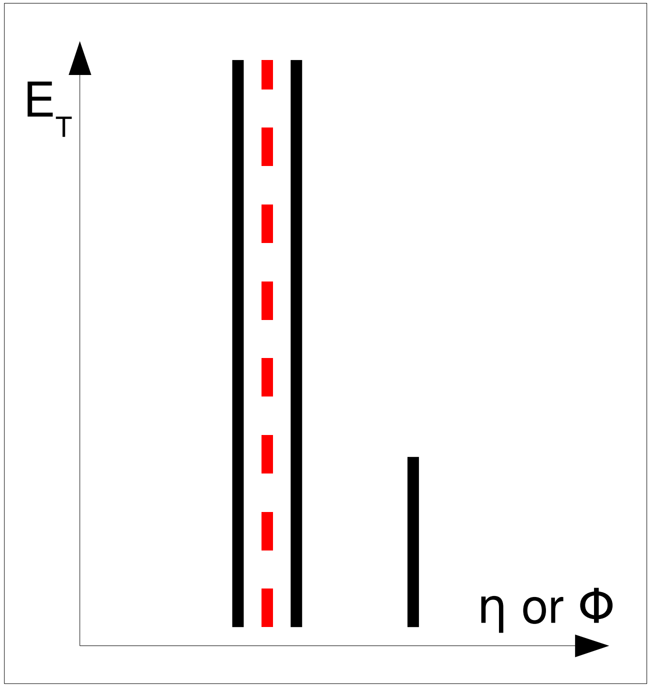
The iteration of the Atlas Cone

ATLAS Cone: Dark towers



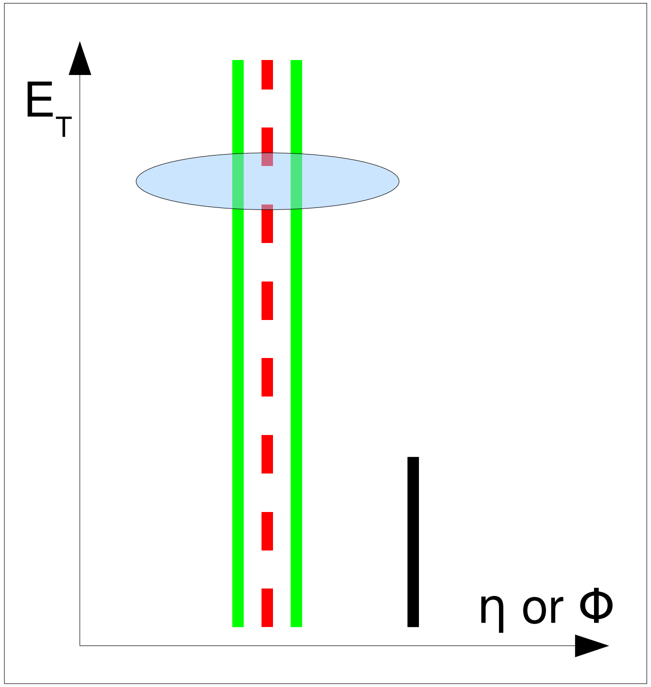
The iteration of the Atlas Cone

ATLAS Cone: Dark towers



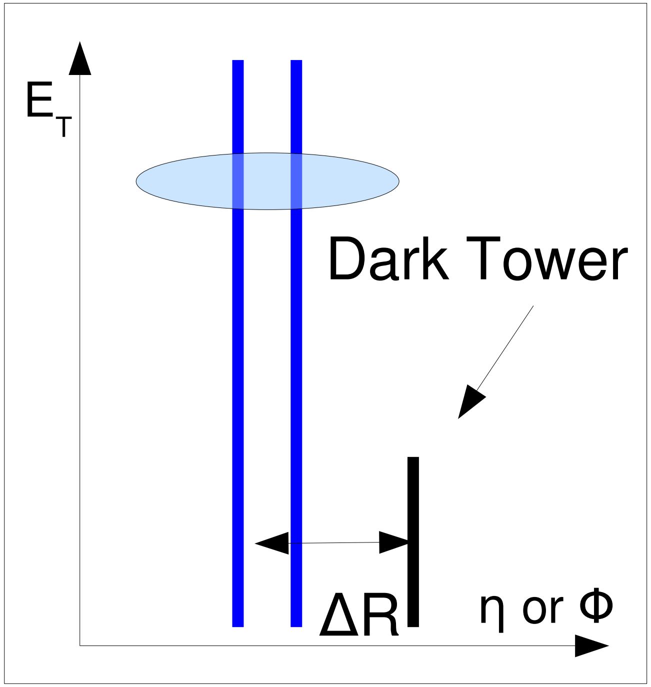
The iteration of the Atlas Cone

ATLAS Cone: Dark towers



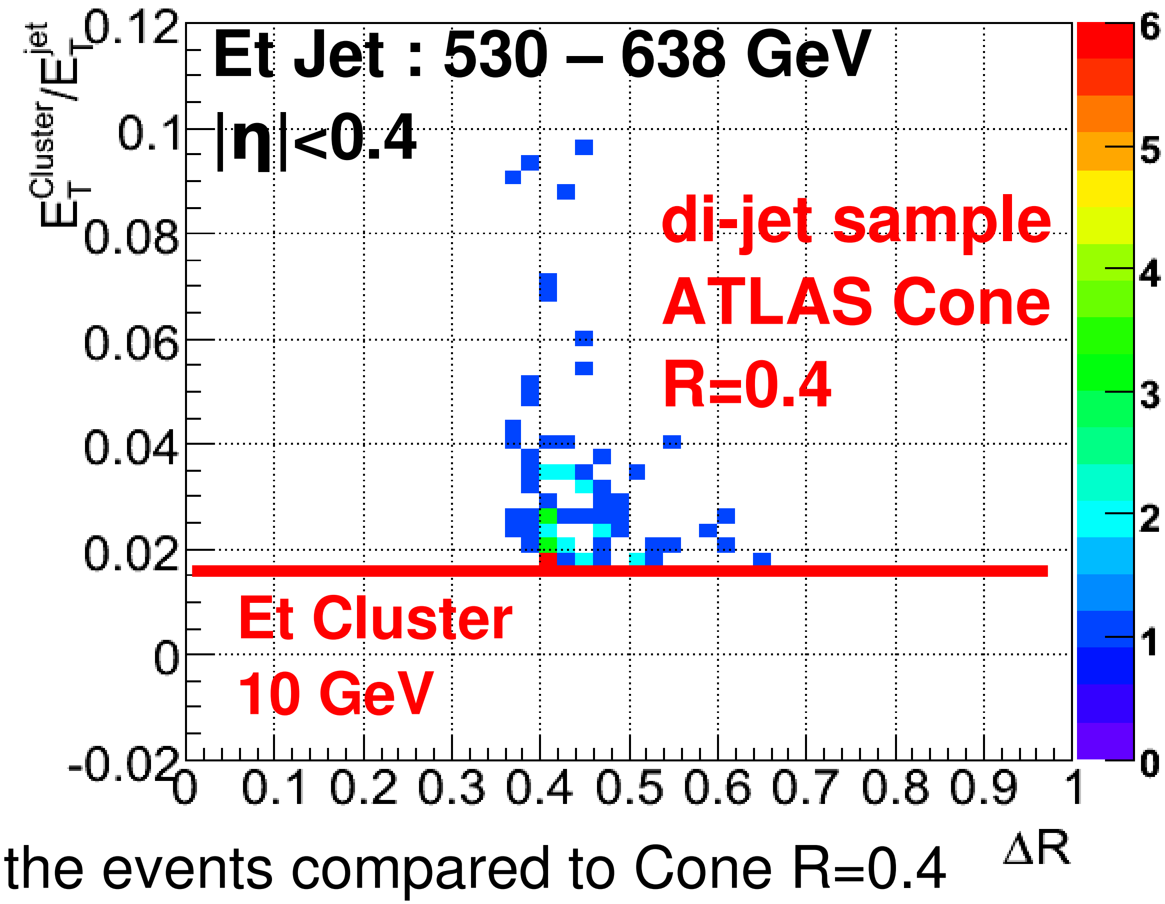
The iteration of the Atlas Cone

ATLAS Cone: Dark towers



Dark towers are close to jets.
Associate every cluster to the closest jet in ΔR.

Comparison done @EM Scale
(cluster outside jets not calibrated)



ATLAS Cone R=0.4:

Jx Sample	Dark tower events
J6	14.8%
J5	5.6%
J4	1.0%
J3	0.4%

ATLAS Cone R=0.7: ~ half of the events compared to Cone R=0.4

Theoretical requirements

We want to **connect** what we **measure** back to the **fundamental** degrees of freedom of our **model** (measurements of PDF, Top mass , W mass, Higgs? ...). (Usually) this requires **comparison** to the state-of-art theory, so our jet finding procedure had better be something the theory/model can replicate.

Infrared Safety? Adding an arbitrarily soft gluon should not change the jet configuration.

Collinear Safety? Splitting a parton in a collinear pair should not change the jet configuration.

Infrared/collinear-safe jet algorithms needed for higher order calculations.

CTEQ school-MCnet 2008
Gavin Salam Lectures on Jets

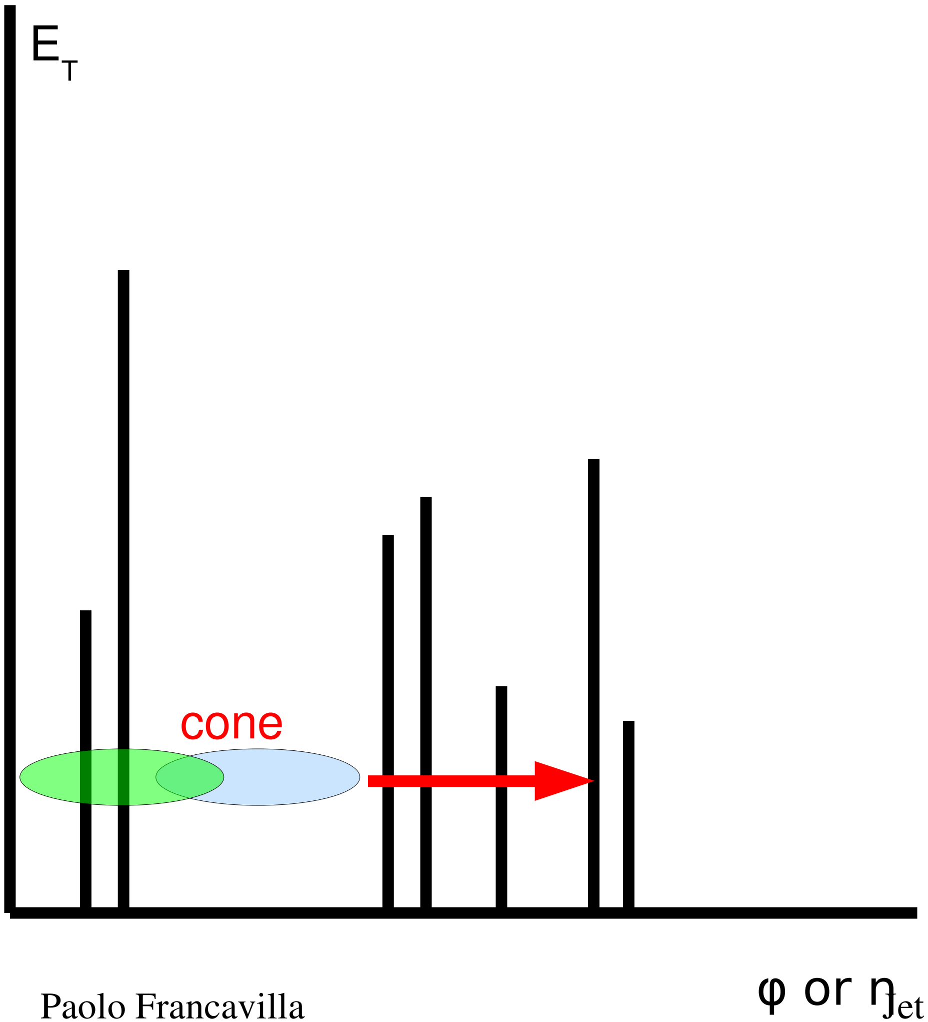
Among consequences of IR unsafety:

	Last meaningful order			
	JetClu, ATLAS cone [IC-SM]	MidPoint [IC_{mp}-SM]	CMS it. cone [IC-PR]	Known at
Inclusive jets	LO	NLO	NLO	NLO ($\rightarrow$ NNLO)
$W/Z + 1$ jet	LO	NLO	NLO	NLO
3 jets	**none**	LO	LO	NLO [nlojet++]
$W/Z + 2$ jets	**none**	LO	LO	NLO [MCFM]
m_{jet} in $2j + X$	**none**	**none**	**none**	LO

New Jet Algorithms

SISCone (Salam/Soyez,'07):

Used a sliding window idea to find **all possible stable cones,** independently of any seeds, **then run usual split–merge procedure**
This gives an infrared safe cone algorithm



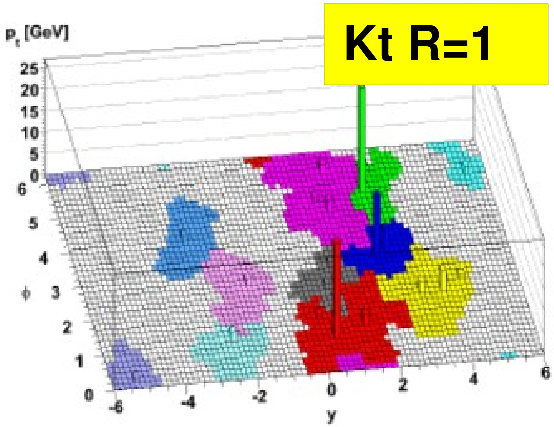
Jet Area

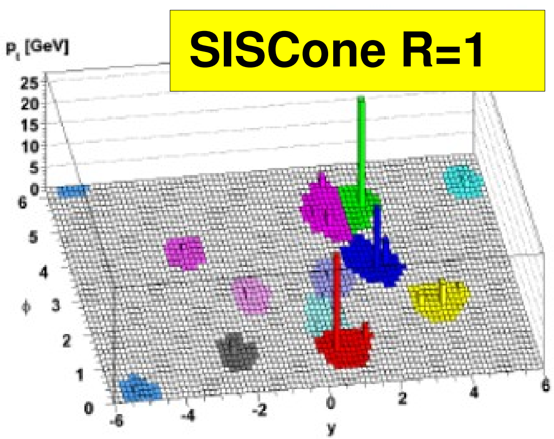
In association with the new algorithms, new methods of calculating the jet area were developed. (Cacciari/Salam/Soyez,'08, JHEP 0804:005,2008)

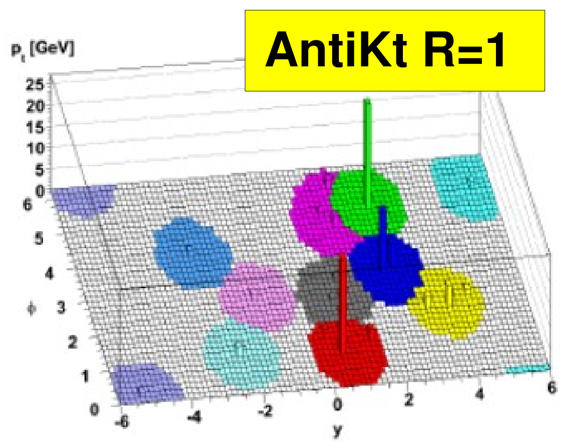
The determination of jet areas is crucial to the subtraction of the underlying event and pile-up from a jet.

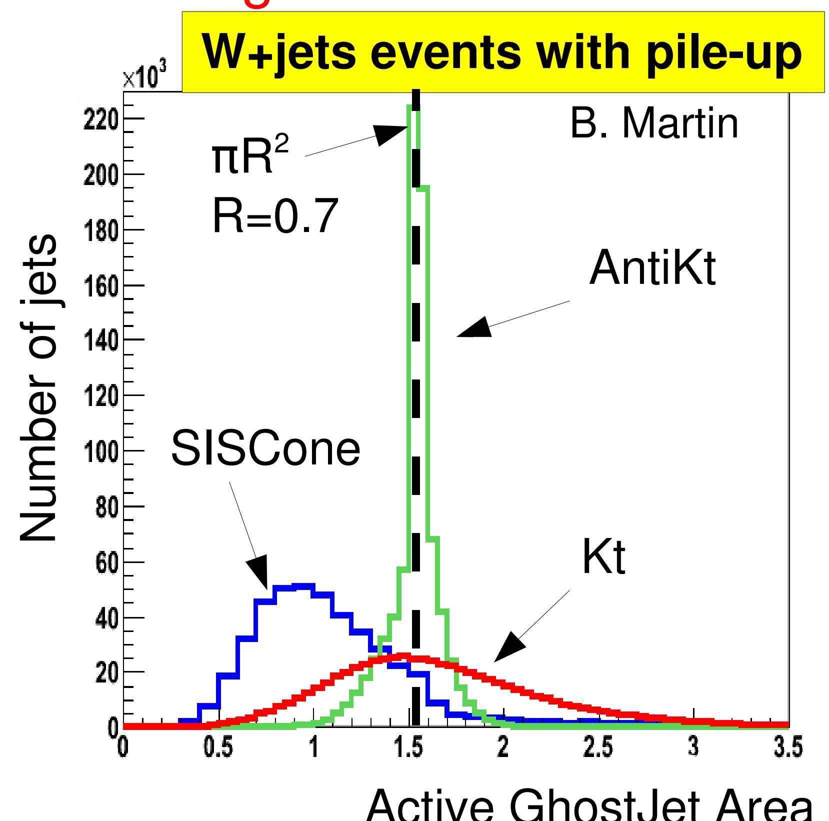
Jet Area:

– Kt algorithm has the largest one
– SISCone algorithm has the smallest one
– AntiKt algorithm has the most regular one





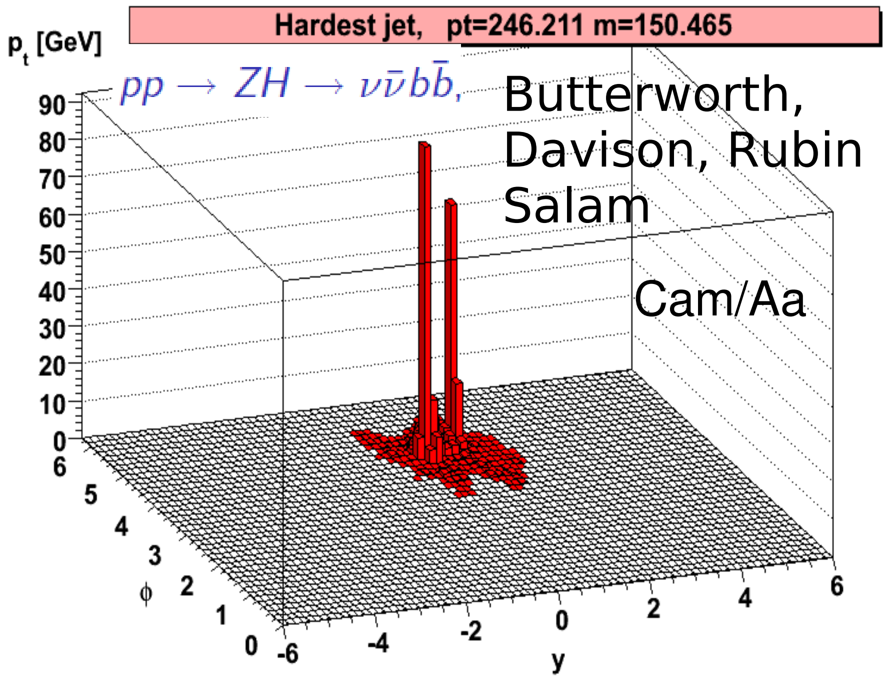




Regular/circular area gives a clear definition of the calorimetric region of interest of the jet.

Jet Substructure

When a W, Z,H or a top → a single jet
Not unusual at LHC: m_W, $m_t \ll$ 14 TeV



Can we identify hadronically decaying EW bosons when they're produced at high pt?

Colour singlet heavy objects,
two body decay (W,Z,H...)
Colour singlet heavy objects,
three body decay (Neutralino,..)
Coloured heavy objects (top,...)

QCD jets (quark gluon separation)

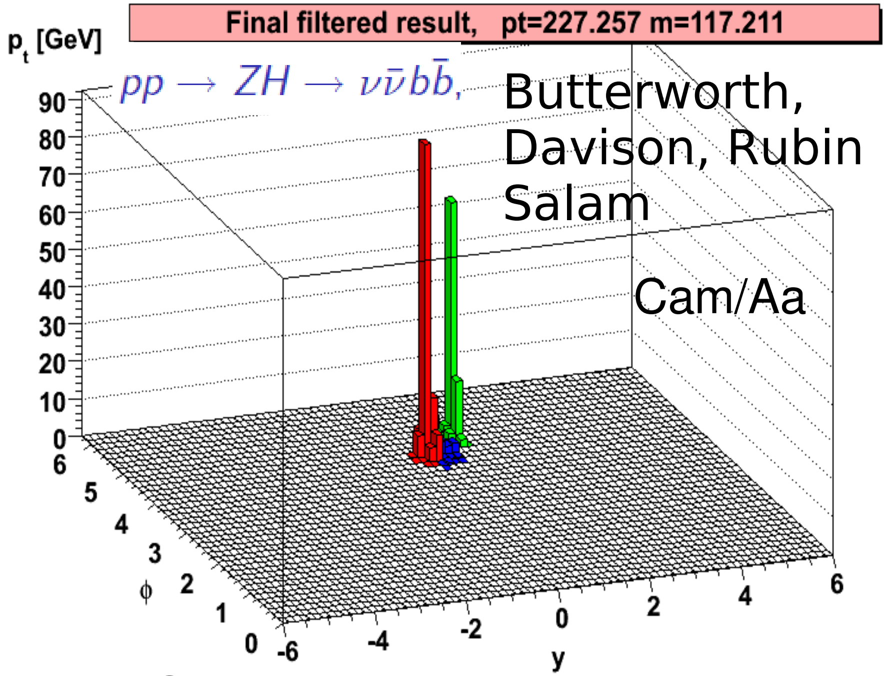
Butterworth+ WW (kT y scales)
Butterworth+ W,Z,H in SUSY cascades
Butterworth+ W/Z+higgs
ttH Atlas CSC results
Butterworth+ neutralinos
ATLAS public note (SUSY09)
Kaplan+ top tagging

Schwartz talk on top-tagging
Almeida+ and Almeida+ top event shapes
Top ID using kT y scales Les Houches report
Seymour+ clustering
Dokshitzer+ clustering
S. Ellis+ jet pruning, using top as an example
Krohn+ variable R

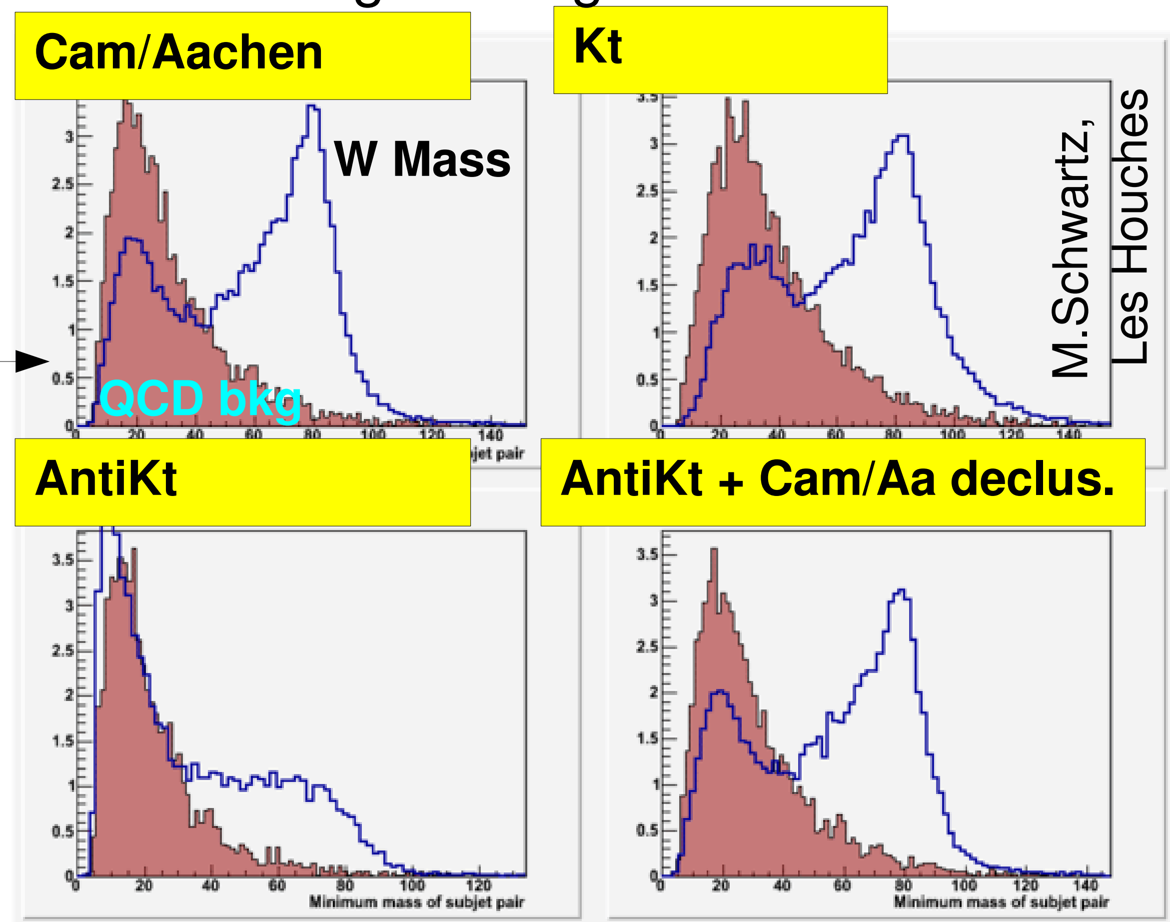
List of references: Les Houches 2009

Jet Substructure

The subjets structures can be used as a possible new (light) Higgs discovery ch.



Similar procedures can be used for other objects: i.e boosted tops.
For the clustering algorithms the declustering is straight-forward.



Use Case:

Reconstructing the W mass in boosted top:

Anti-Kt declustering does not provide strong discrimination, but finding the jets with Anti-Kt, and then declustering with C/A works well.

Experimental requirements

See Sebastian's talk

ATLAS events will be complex: many hard jets, underlying event, pile-up, high multiplicities of outgoing particles.

Fast even for very high multiplicities (trigger, off-line analysis,etc...).

Good at picking out the “hard” physics, robustness under pile-up.

Small sensitivity to non-perturbative effects (fragmentation and underlying event)

The jet algorithm should facilitate experimental reconstruction and calibration:

Jet energy scale and resolution, spatial resolution, efficiency and purity.

Small systematic uncertainty on the jet energy scale, quickly obtainable in the first data.

We need flexibility to optimise the most suitable algorithm for a specific analysis:

1)Exotics physics: high P_T di-jets to search for new short distance interactions→ **large R**

2)Top Physics: high jet multiplicity→ **small R**

3)Higgs: Low energy jets, forward jet tagging, central jet veto, b-jets,...

The jet algorithms should be able to resolve the sub-jet structure.

Reconstruction of high P_T top quarks, W/Z and Higgs bosons in highly boosted jets.

Conclusions

A lot of recent progress in jet algorithms

Jets are not just smeared partons, they have a complex substructure.

We should take the best technology with us, in order to compare our measurements with the state-of -art theory.

The IR/Collinear safe jet algorithms give sense to pQCD predictions.

Use of various algorithms and parameter tuning for a specific analysis is necessary to get the most out of the jet algorithm:

- Try out different combinations of algorithm & R
- Check Variations of alg. & R don't change extracted physical quantities
  - Flexibility is needed.

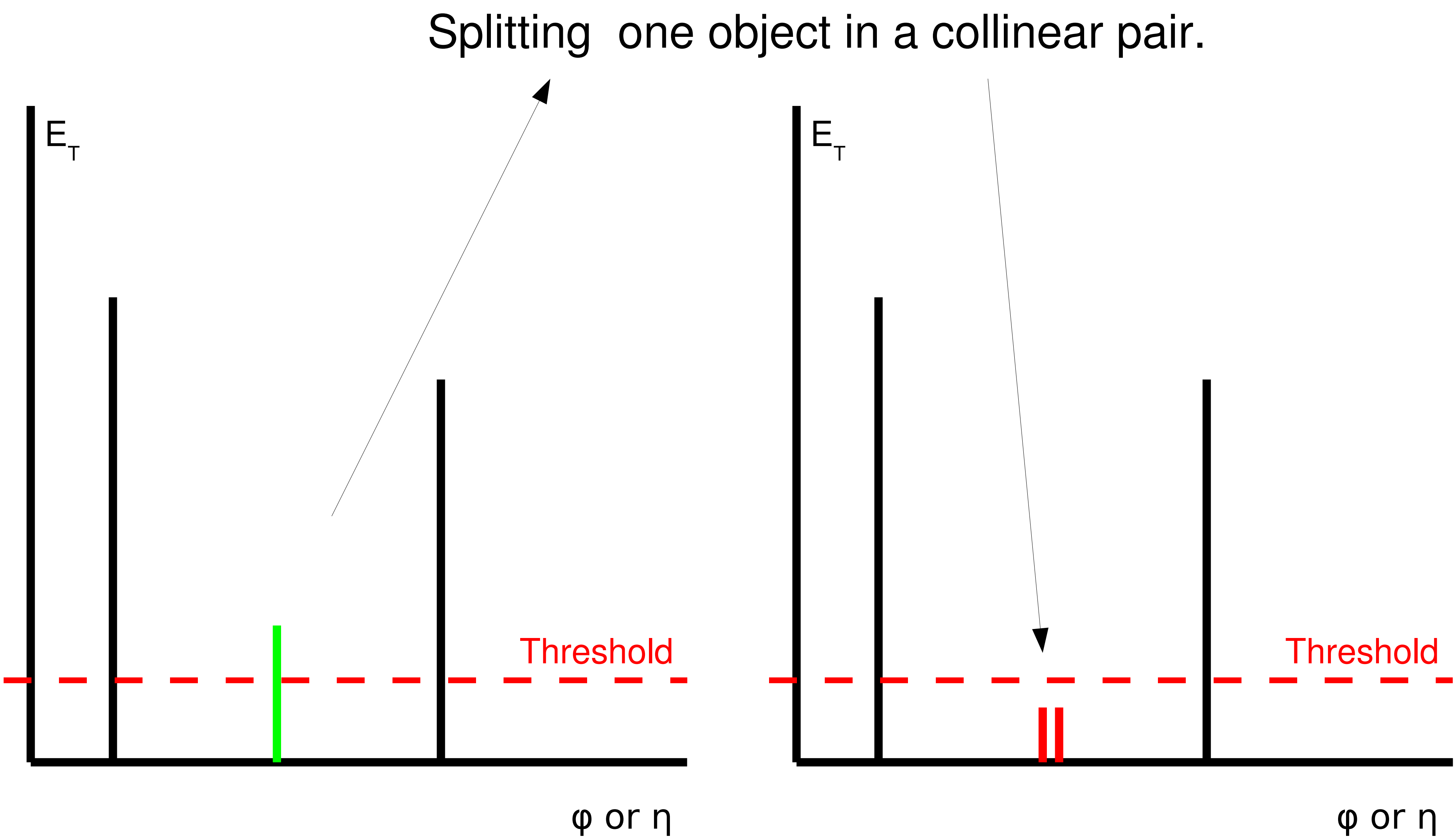
Experimental requirements are the crucial points (see Sebastian's talk).

Acknowledgements

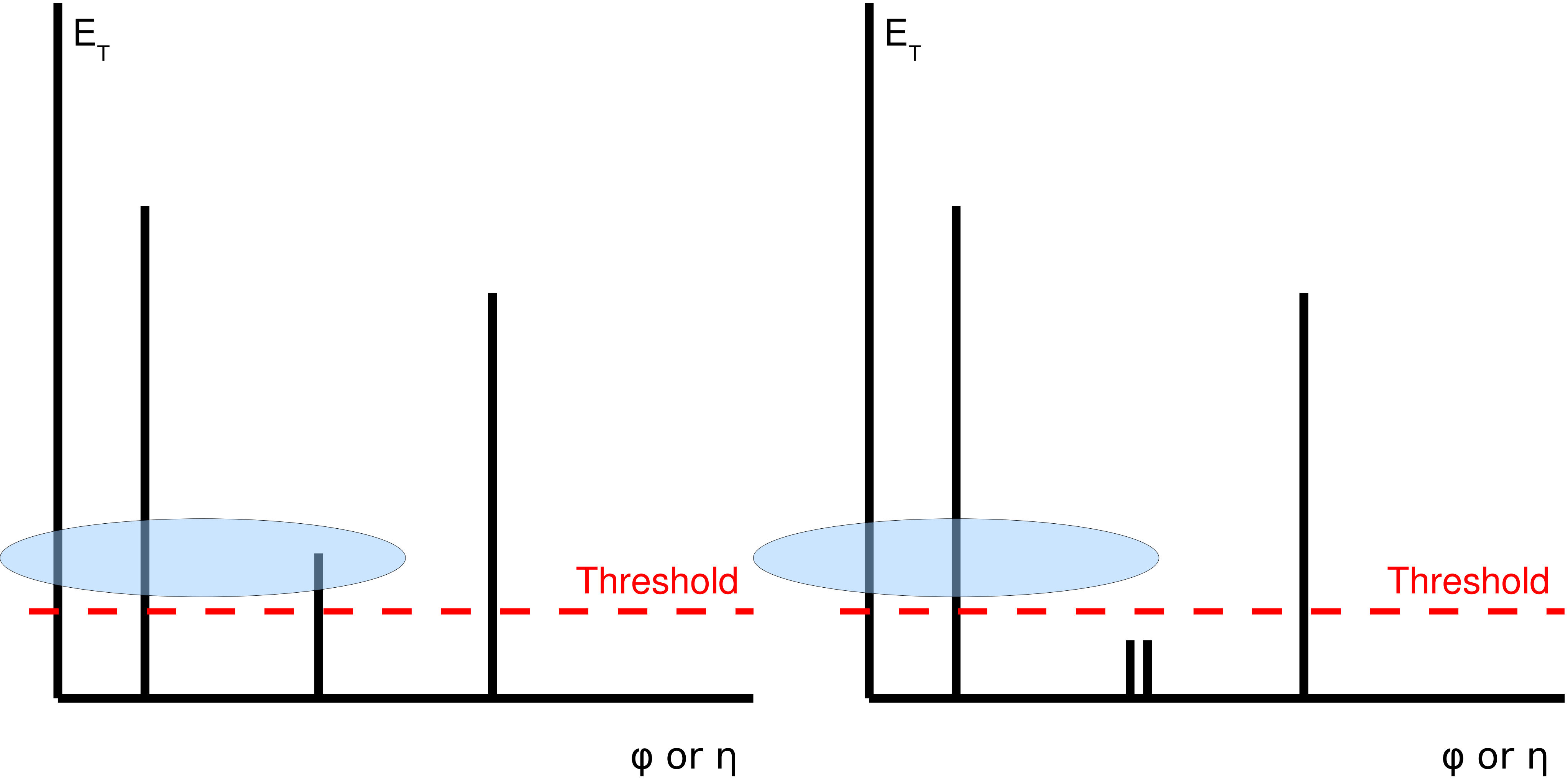
J. Butterworth, M. Cacciari, M. Campanelli, T. Carli, M. Cascella, P.A. Delsart, V. Giangiobbe, J. Huston, I. Jen-La Plante, P. Loch, K. Lohwasser, S. Majewski, B. Martin, D. Miller, J. Proudfoot, C. Roda, G. Salam, M. Seymour, I. Vivarelli, Z. Zenon

BACKUP

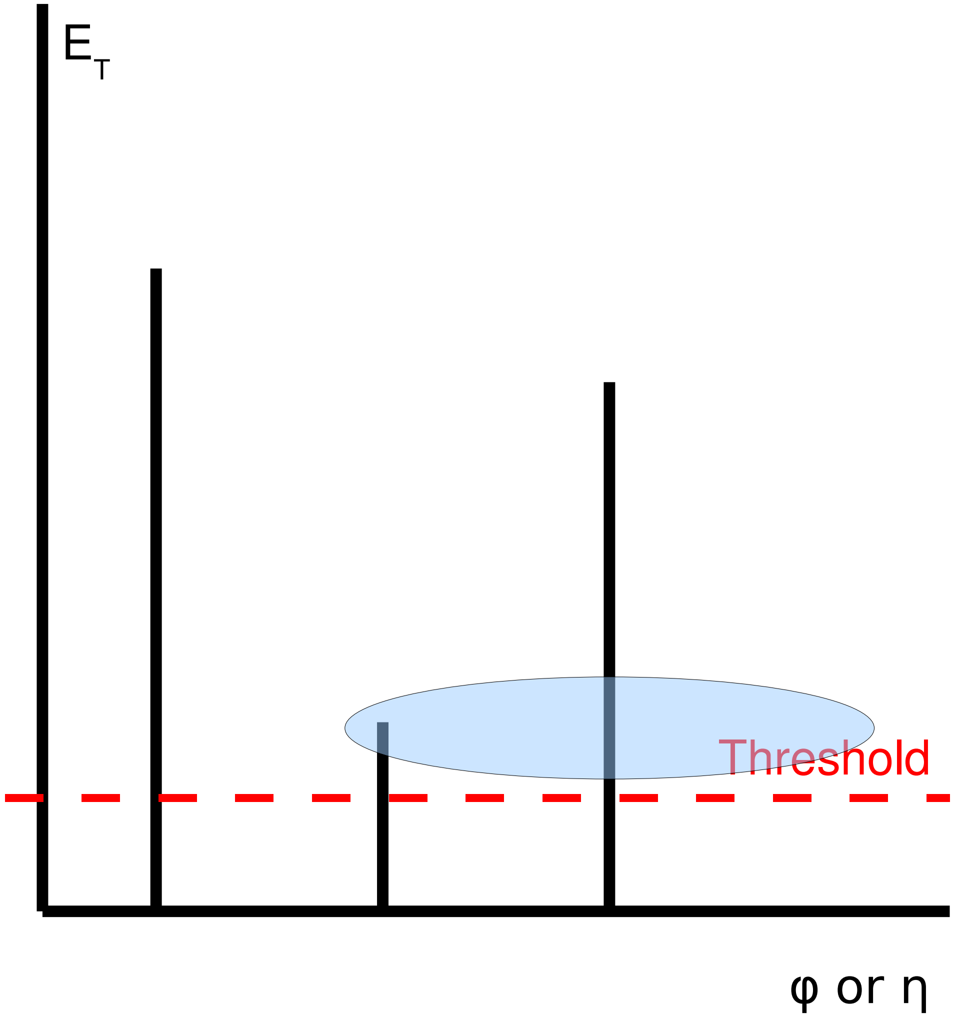
ATLAS Cone: A special case

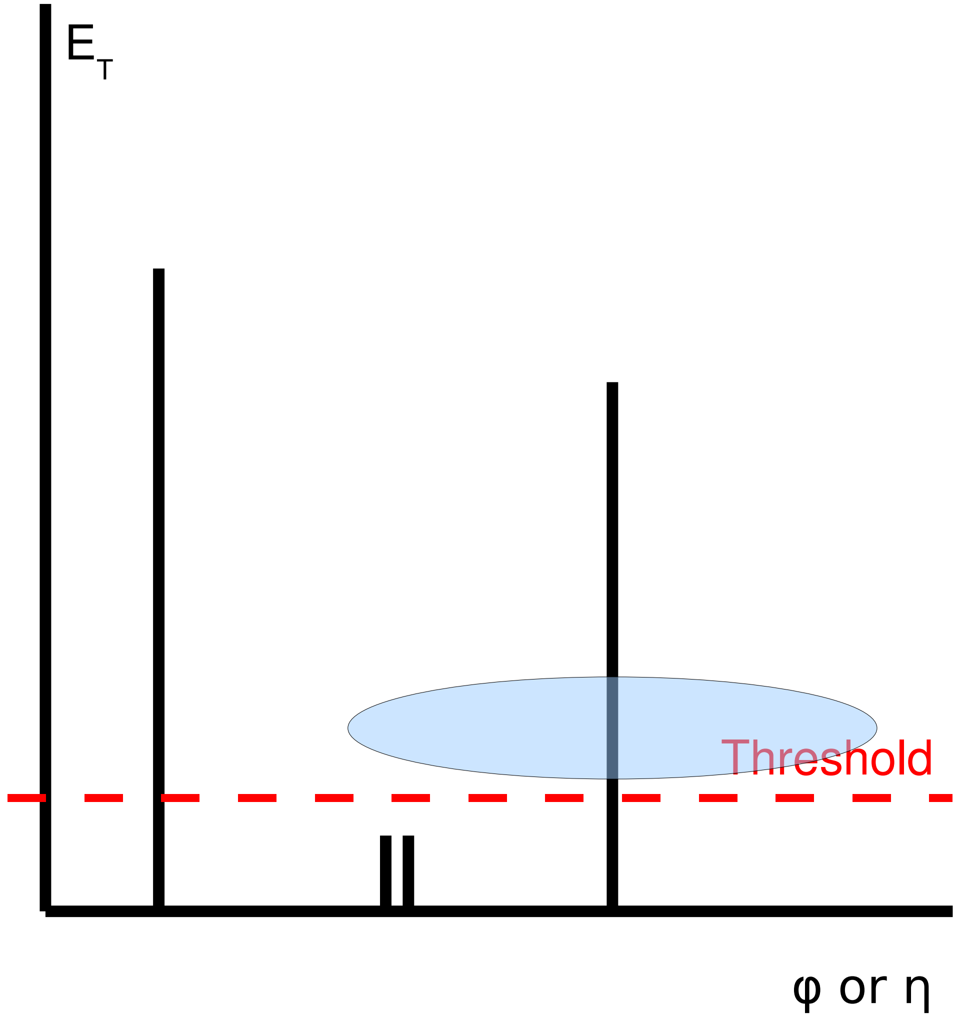


ATLAS Cone: A special case

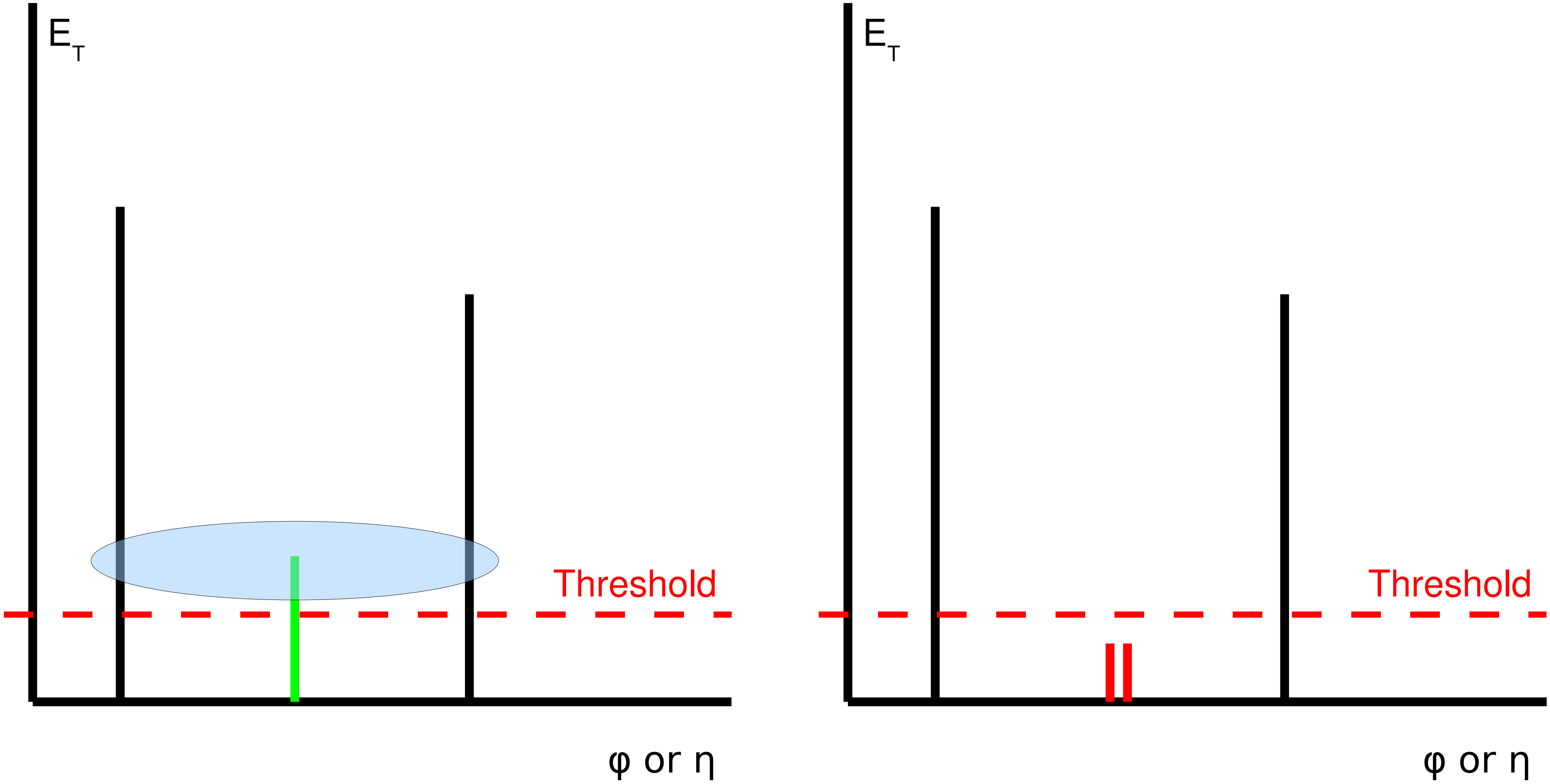


ATLAS Cone: A special case

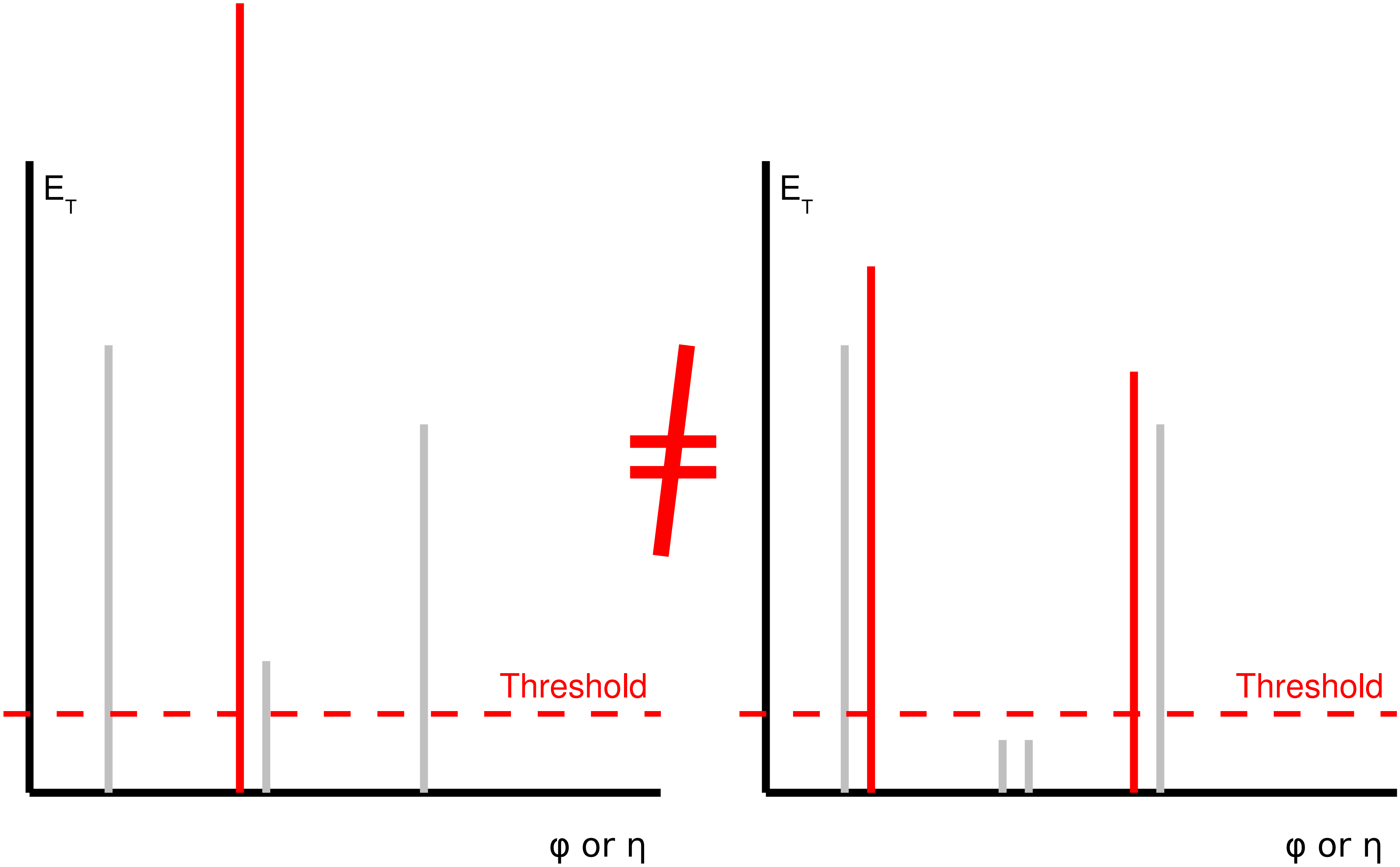




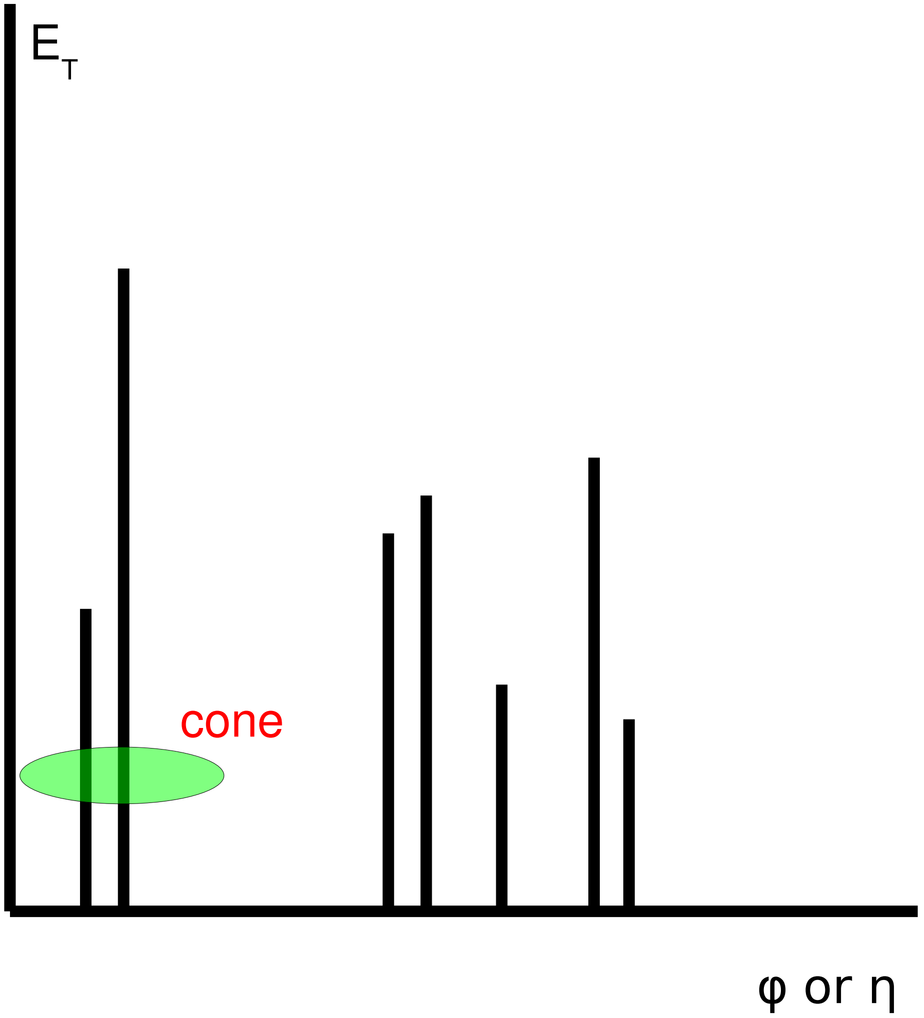
ATLAS Cone: A special case



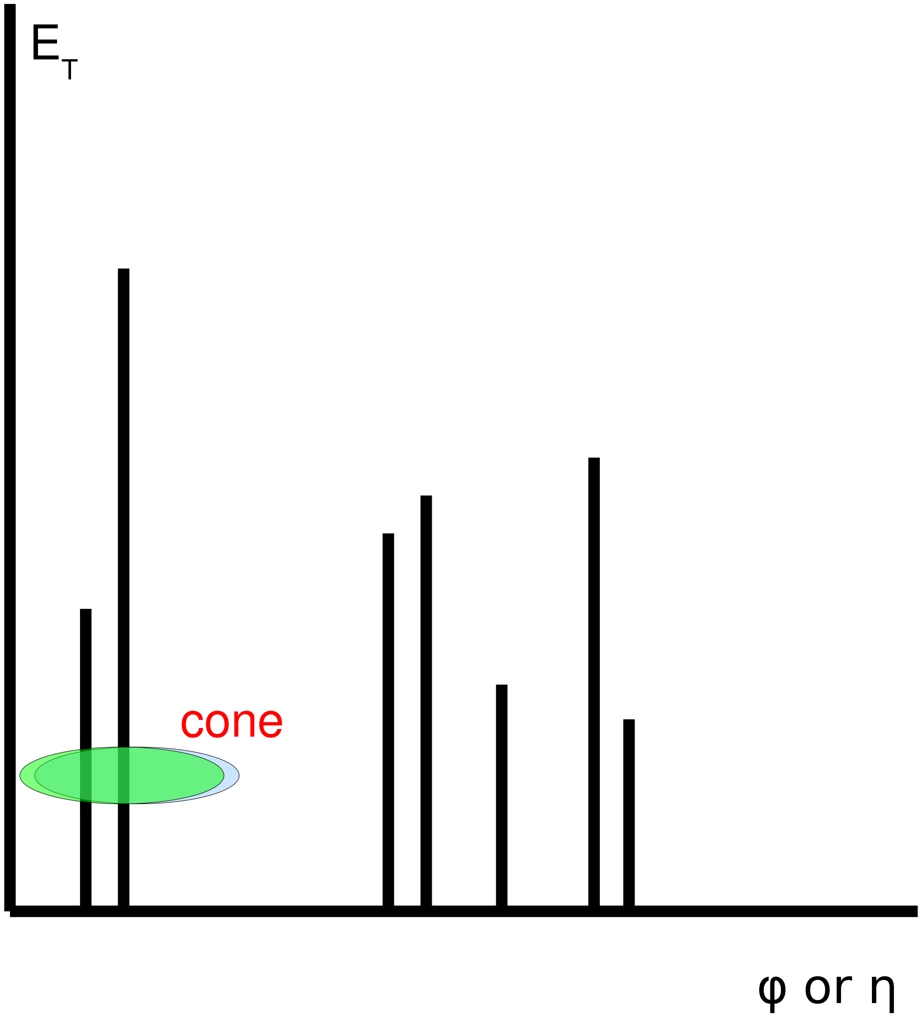
ATLAS Cone: A special case



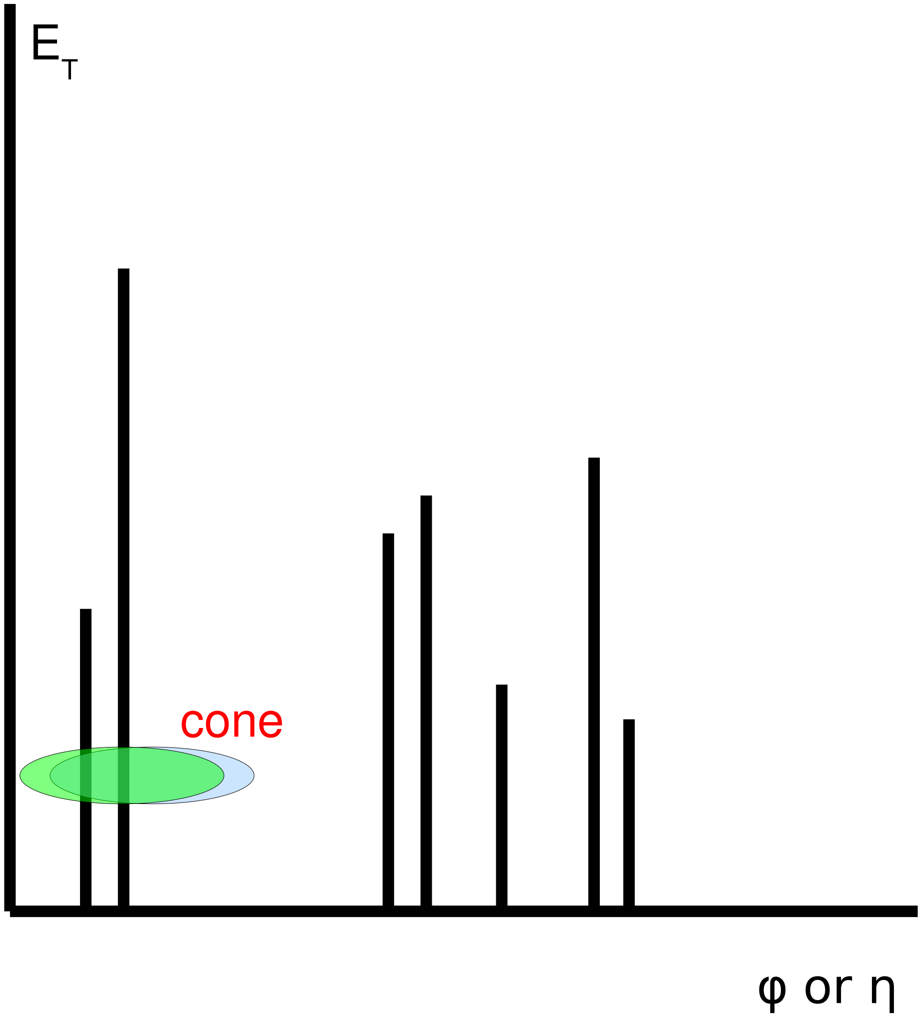
SISCone Algorithm



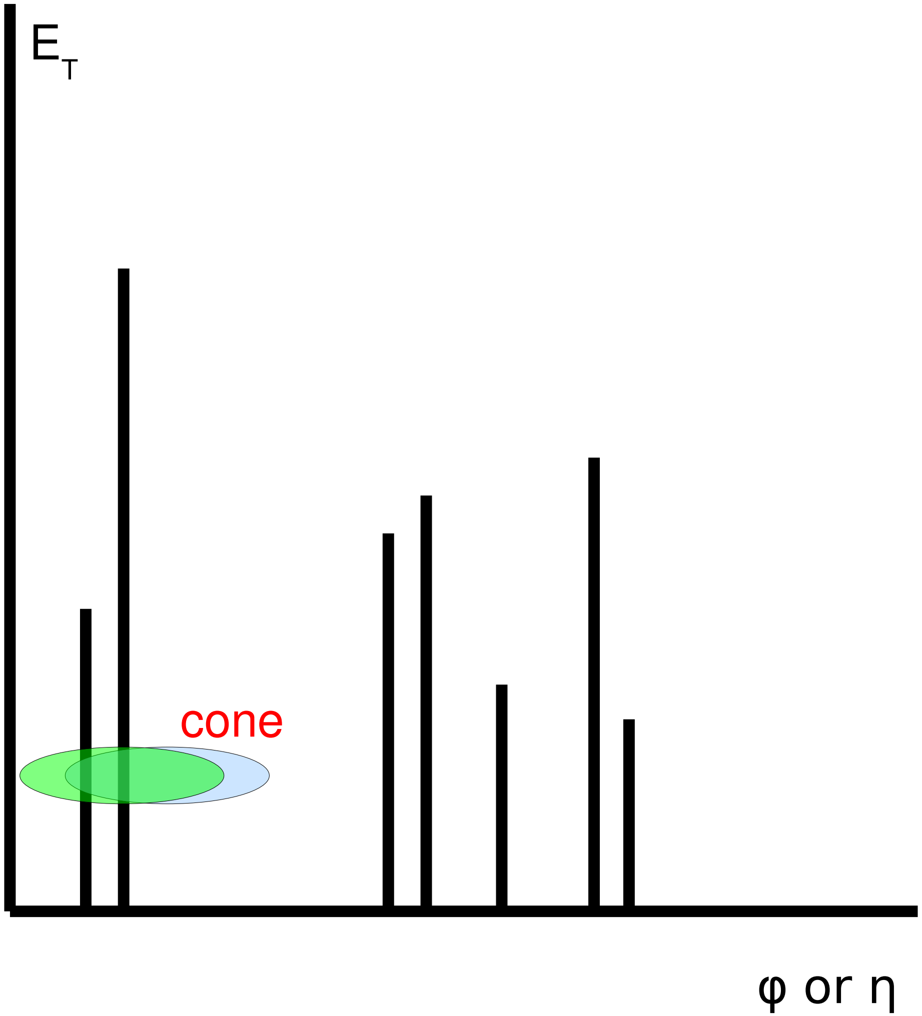
SISCone Algorithm



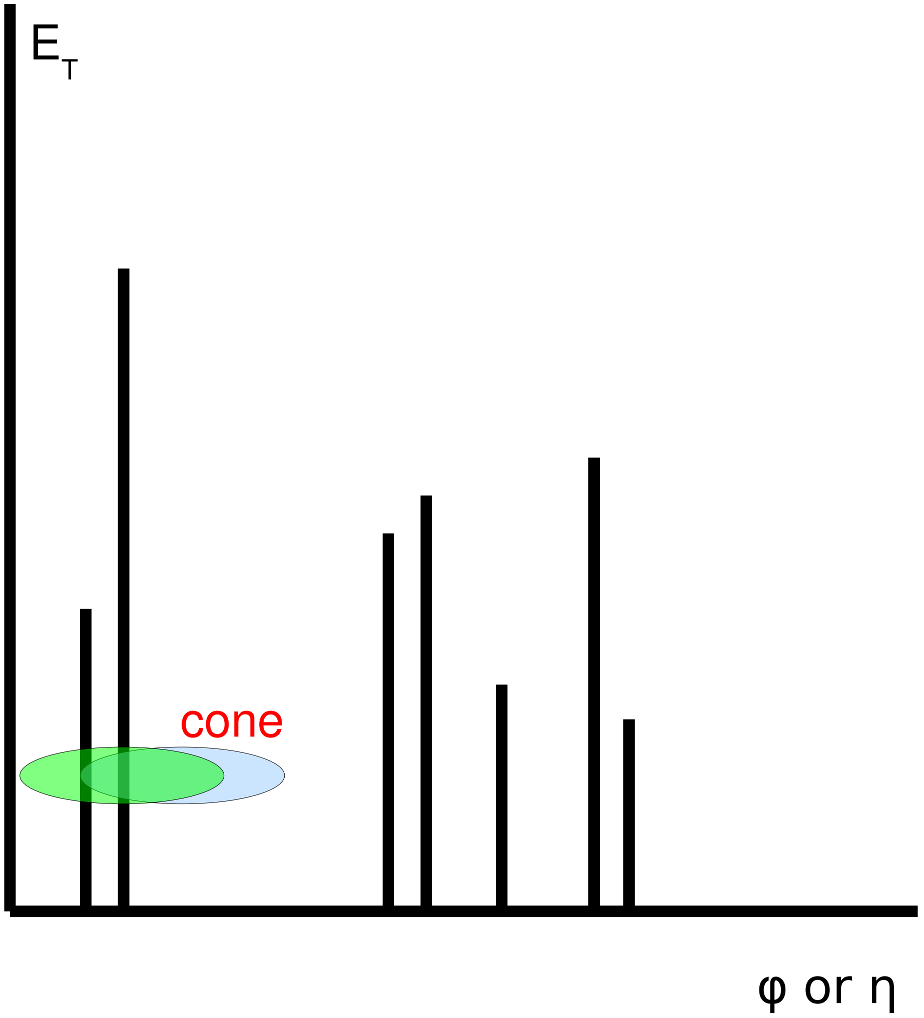
SISCone Algorithm



SISCone Algorithm



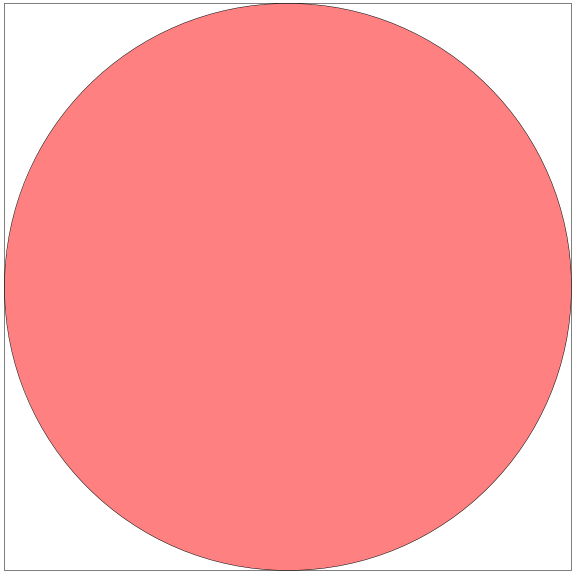
SISCone Algorithm



Calculating the Jet Area - Ghosts

In association with the new algorithms, new methods of calculating the jet area were developed.

Starting point: Monte-Carlo estimation of the area of circle

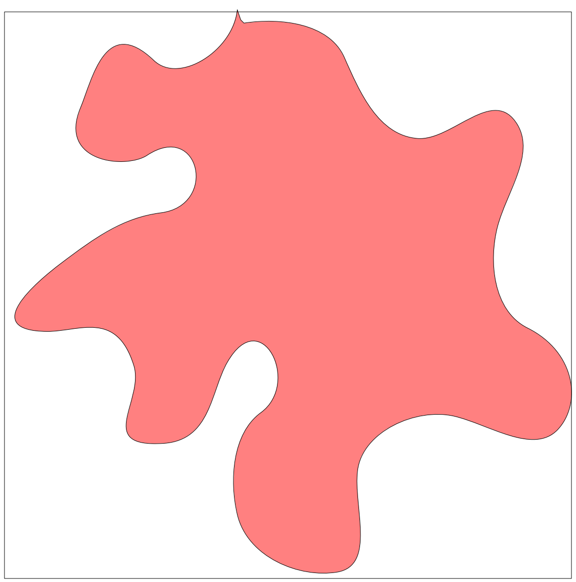


- extract randomly distributed points in the square;
- check if the point is inside the circle (success);
- calculate the the number of successes over the total number of tries

Calculating the Jet Area - Ghosts

In association with the new algorithms, new methods of calculating the jet area were developed.

This strategy works even for other shapes.
We can use the same procedure to estimate the area of a jet.
BUT a jet is a clustering of objects (i.e. particles).

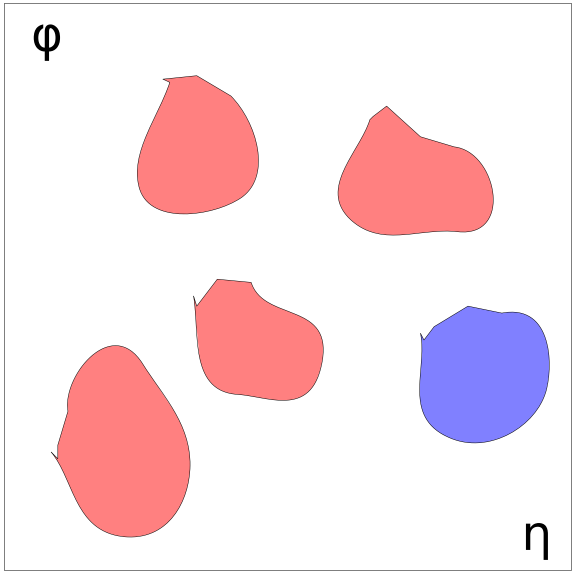


Adding extra “points” could change the clustering.

We must take care that the addition of these randomly distributed points do not change the real content of the jets.
We need a technique that is
robust under the addition of these random points.

Calculating the Jet Area - Ghosts

We know that an infra-red safe jet algorithm is “stable” under the addition of soft particles in the final state.
So the idea is to add randomly distributed points with a really really low Pt (GHOSTS).
Running the jet algorithm, we know the number of ghosts for every jet and we can use this information to estimate the area.



We can eventually produce low Pt (<< eV) jets without real particles inside, just ghosts.
(they are not real jets, ghost jets).