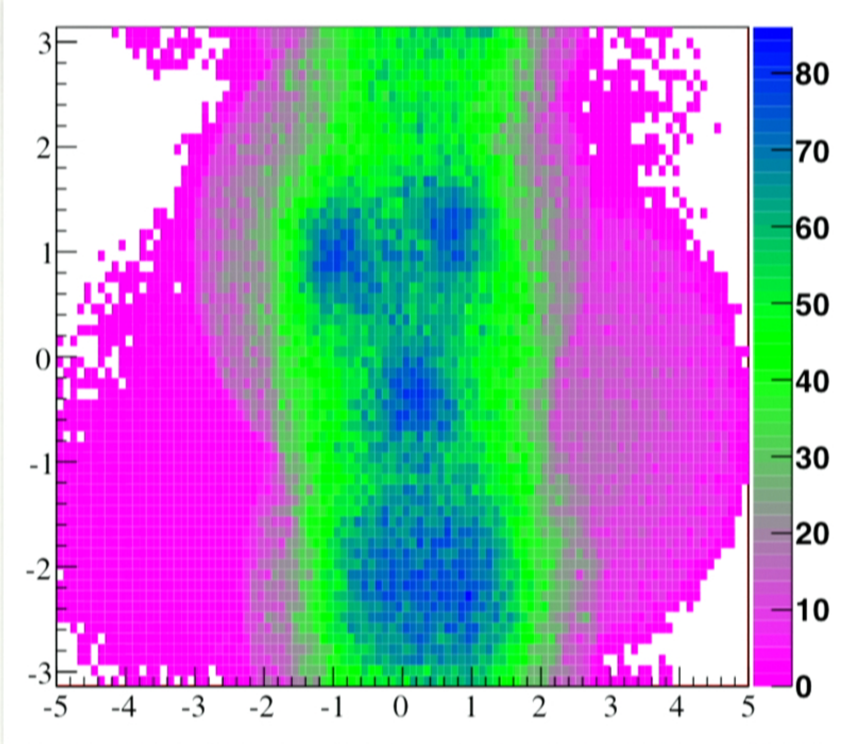
Title: Qanti-kt: Nondeterministic jet clustering

Date: Feb 19, 2013 01:00 PM

URL: http://www.pirsa.org/13020135

Abstract: The interpretation of events with jets is often ambiguous, especially for the sort of highly complex events one encounters at the LHC. One often finds that an event interpreted as signal-like using one choice of jet algorithm and radius parameter is no longer signal-like with another, even if the two are very similar. Here we present an extension of the Qjets procedure designed to account for this ambiguity and assign each plausible interpretation of an event a weight, so that events which are unambiguously signal-like carry more influence on one's results than events which are only marginally so. This procedure can be used with any existing analysis employing a sequential recombination algorithm like anti-kT and we will show that through its use the statistical power of an analysis often increases. In particular, we will see that a up to a 20\% improvement in statistical significance can be realized for a Higgs-like analysis searching for a resonance recoiling against an associated vector boson.



Qanti-k_T: Nondeterministic jet clustering

David Krohn (Harvard)

PI 2/19/13

Based on work with S. Ellis, A. Hornig, T. Roy, and M. Schwartz
and work in progress with D. Kahawala

Takeaway

- Many jet algorithms have a good motivation
  - Typically, we use anti-k_T, k_T, or C/A to chose the best one
- However, these can give very different interpretations of the same event
- By considering many algorithms at once we can get weighted interpretations of a jet
  - Weighted is better than Unweighted -> better statistics

Review of Jets & Jet Substructure

Types of Algorithms

- There are two main classes of jet algorithm
- Sequential recombinations
  - Combine four-momenta one by one
- Cone algorithms
  - Stamp out jets as with a cookie cutter

Focus on these

Sequential Recombination

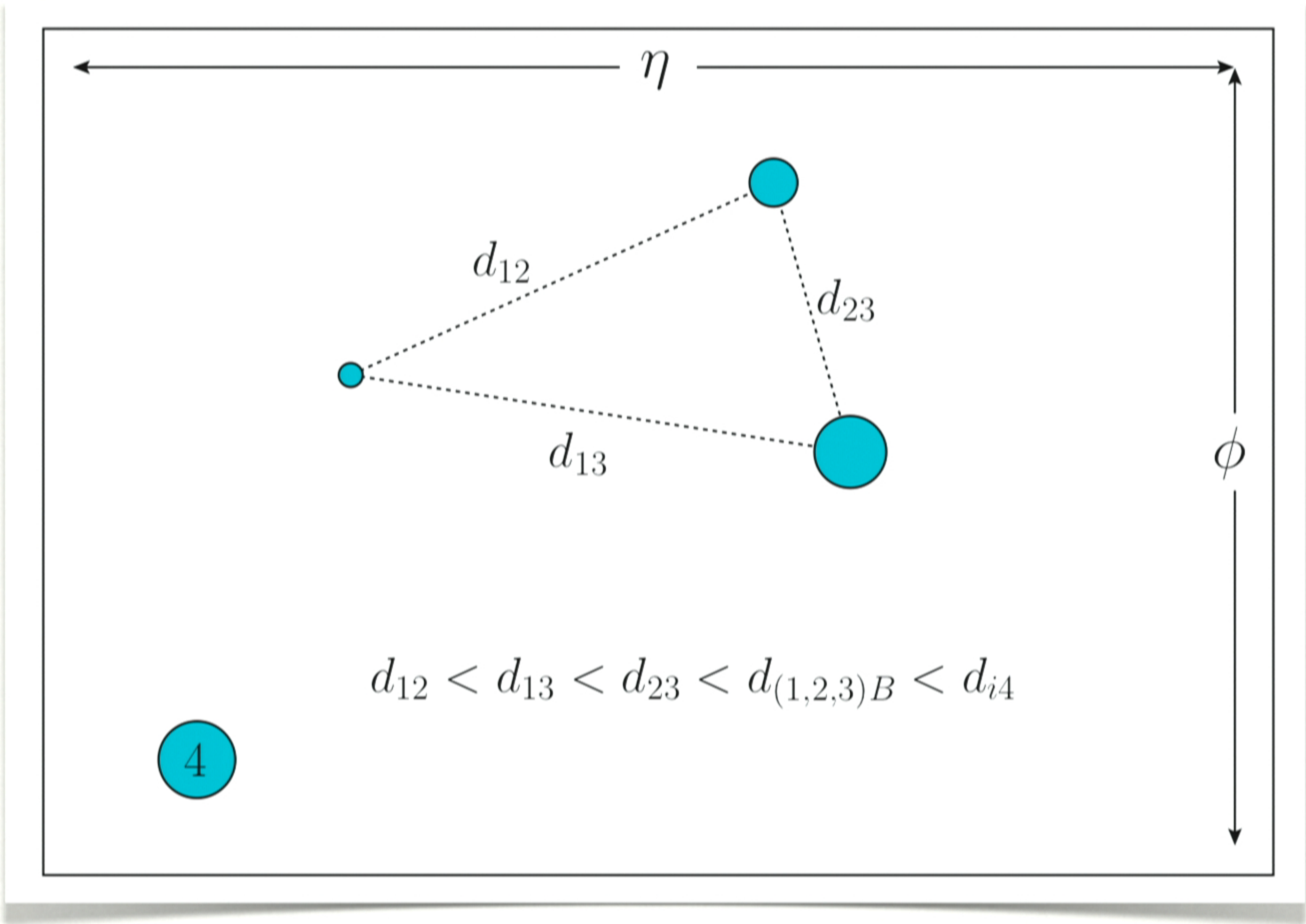
- Define a distance measure between every pair of four-momenta in an event (jet-jet distances)

$$d_{ij}$$

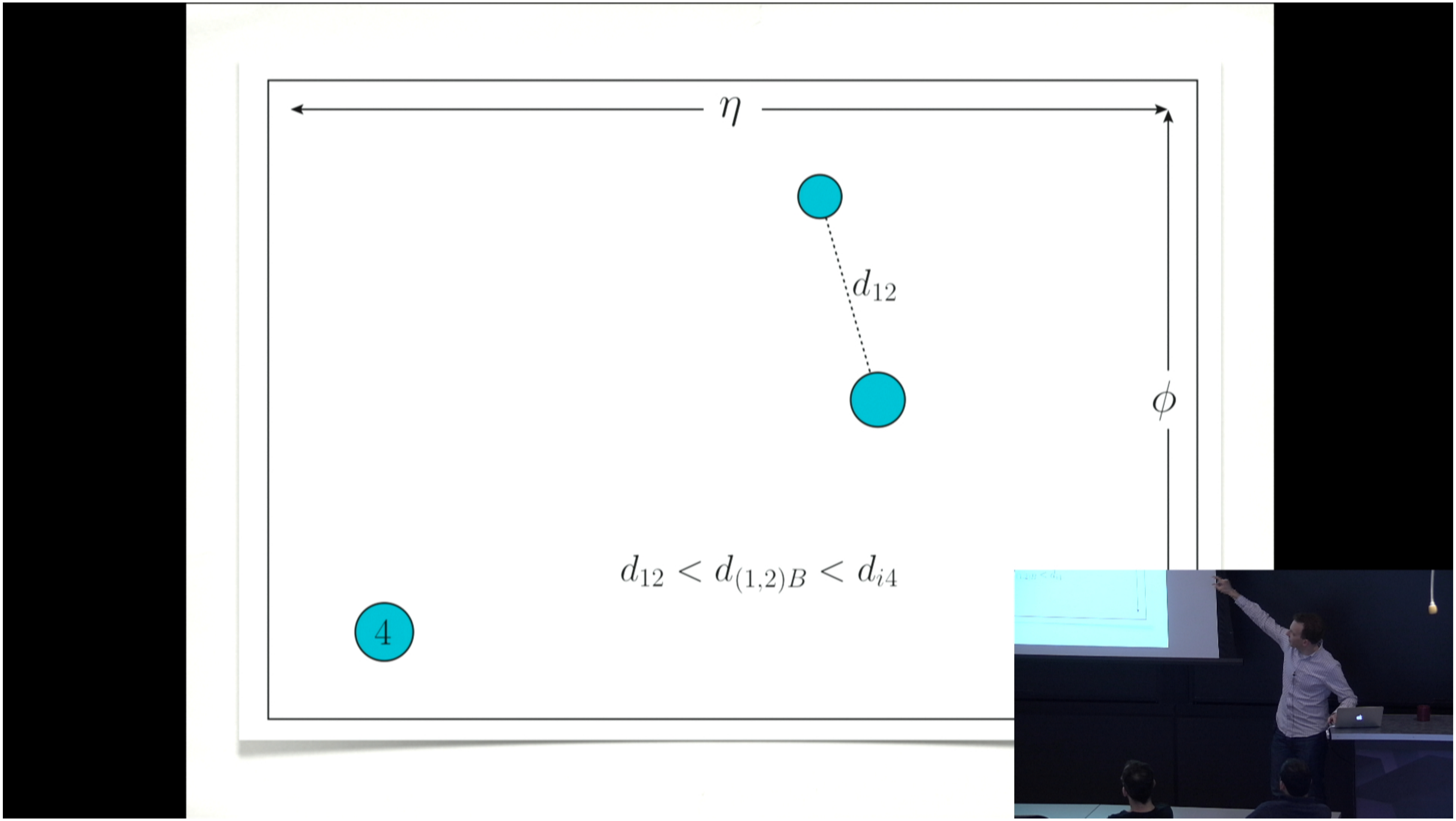
- Define a distance measure for each four-momenta individually (jet-beam distances)

$$d_{iB}$$

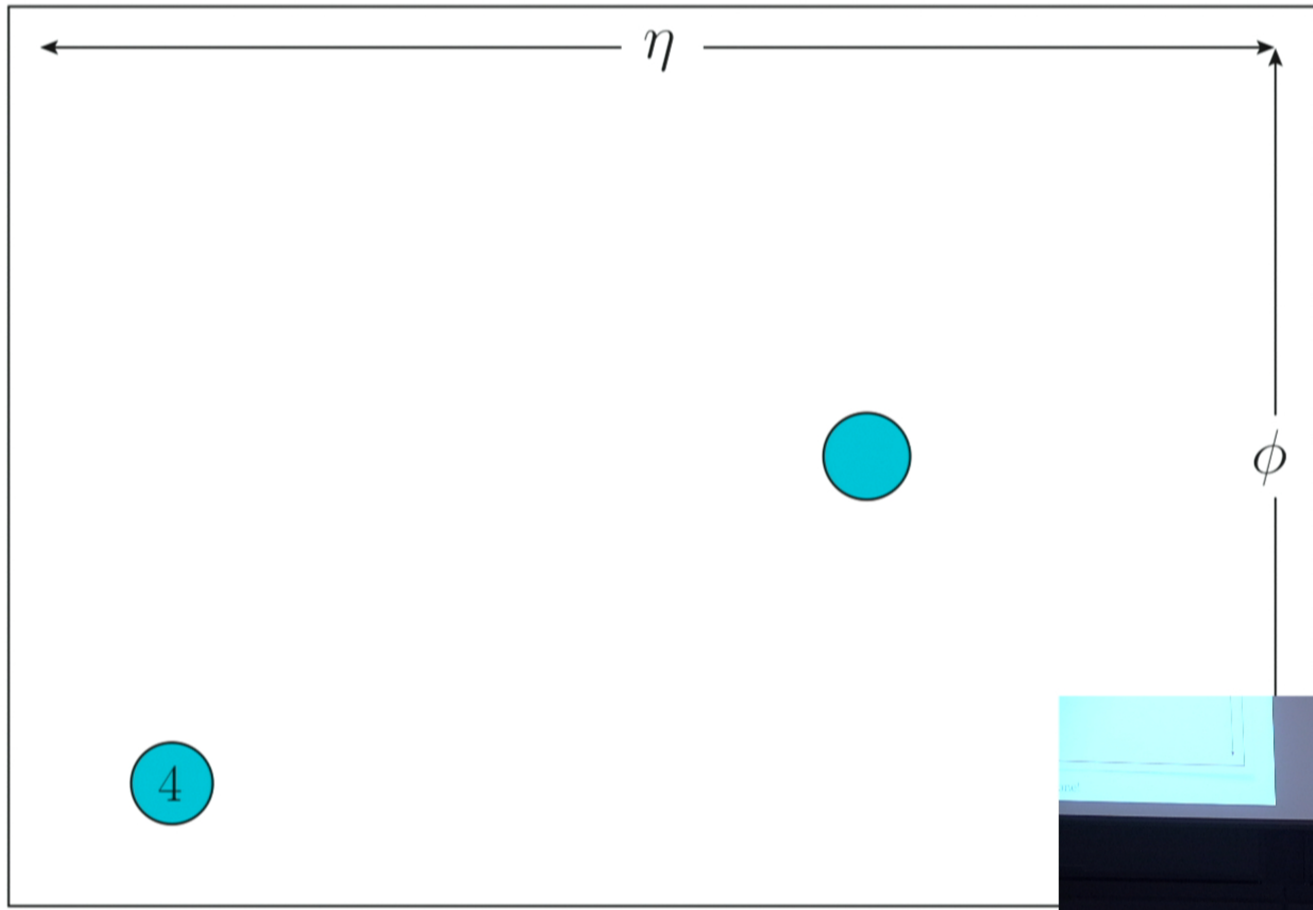
- If smallest distance at any stage in clustering is jet-jet, add together corresponding four-momenta
  - Otherwise take jet with smallest jet-beam distance and set it aside
- Repeat till all jets are set aside
- In this way, jets are constructed by pairwise recombinations - get a tree-like sequence at the end.



η
d_{12}
d_{23}
d_{13}
ϕ
$d_{12} < d_{13} < d_{23} < d_{(1,2,3)B} < d_{i4}$
4



η
d_{12}
φ
$d_{12} < d_{(1,2)B} < d_{i4}$
4



Done!

Standard Recombination Algorithms

- k_T algorithm

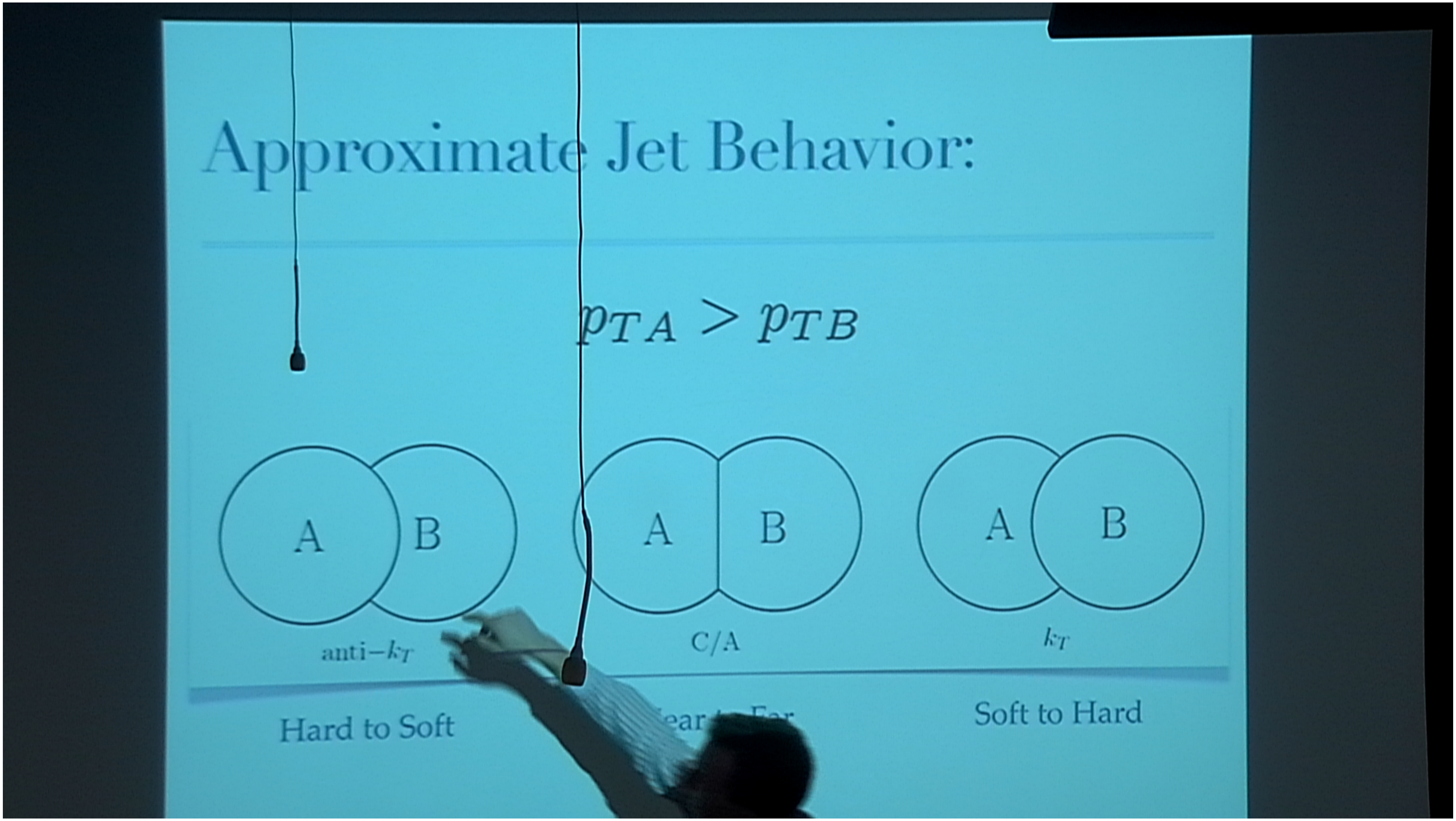
$$d_{ij} = \min(p_{Ti}^2, p_{Tj}^2)\left(\frac{\Delta R}{R_0}\right)^2, \; d_{iB} = p_{Ti}^2$$

- C/A algorithm

$$d_{ij} = \left(\frac{\Delta R}{R_0}\right)^2, \; d_{iB} = 1$$

- anti-k_T algorithm

$$d_{ij} = \min(p_{Ti}^{-2}, p_{Tj}^{-2})\left(\frac{\Delta R}{R_0}\right)^2, \; d_{iB} = p_{Ti}^{-2}$$


Approximate Jet Behavior:
$p_{TA} > p_{TB}$
A
B
anti−k_T
A
B
C/A
A
B
k_T
Hard to Soft
Soft to Hard

Tradeoffs

- k_T & C/A
  - Pro: Cluster near to far (both) & soft to hard (k_T). Allows us to use parton shower heuristics to understand behavior.
  - Con: Jets can have perverse shapes, weird areas
- anti-k_T
  - Pro: Jets are cone-like. Area relatively well defined.
  - Con: The ordering of the shower has little or no physical significance.

Tradeoffs

- k_T & C/A
  - Pro: Cluster near to far (both) & soft to hard (k_T). Allows us to use parton shower heuristics to understand behavior.
  - Con: Jets can have perverse shapes, weird areas
- anti-k_T
  - Pro: Jets are cone-like. Area relatively well defined.
  - Con: The ordering of the shower has little or no physical significance.

Kinematics of Boosted Particles

- The cone containing the decay products of a particle scales as

$$R \sim \frac{2m_X}{p_T}$$

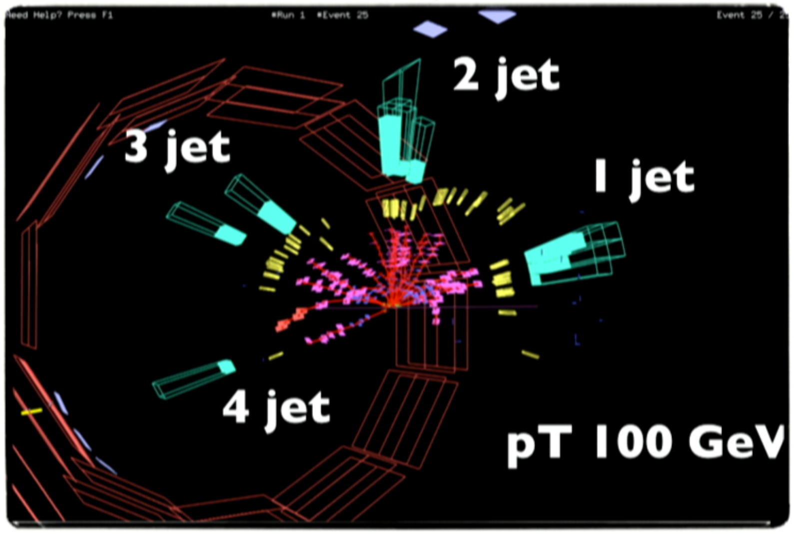
- At LHC energies, even the heaviest particles we know of (Top, W, Z, Higgs) become can become collimated.
- When this happens we say that they're "boosted".
- So we find that EW scale particles are clustered as a single jet as soon as their p_T exceeds a few hundred GeV.

Kinematics of Boosted Particles

- The cone containing the decay products of a particle scales as

$$R \sim \frac{2m_X}{p_T}$$

- At LHC energies, even the heaviest particles we know of (Top, W, Z, Higgs) become can become collimated.
- When this happens we say that they're "boosted".
- So we find that EW scale particles are clustered as a single jet as soon as their p_T exceeds a few hundred GeV.



Unboosted t-tbar pair

All three decay products of the top go into one jet

Boosted t-tbar pair

Figure source: http://www.pha.jhu.edu/groups/particle-theory/seminars/talks/F08/Yumiceva.pdf

Boosted Collider Physics

- This can be a problem!
- Most new physics models include heavy states at the TeV scale
  - If these decay down to W/Z/t, what do we do if everything's collimated?
  - Traditional answer: use the leptonic decays to avoid this mess.
- Modern answer: look inside the jet and make use of QCD to see if the jet came from a boosted heavy object.

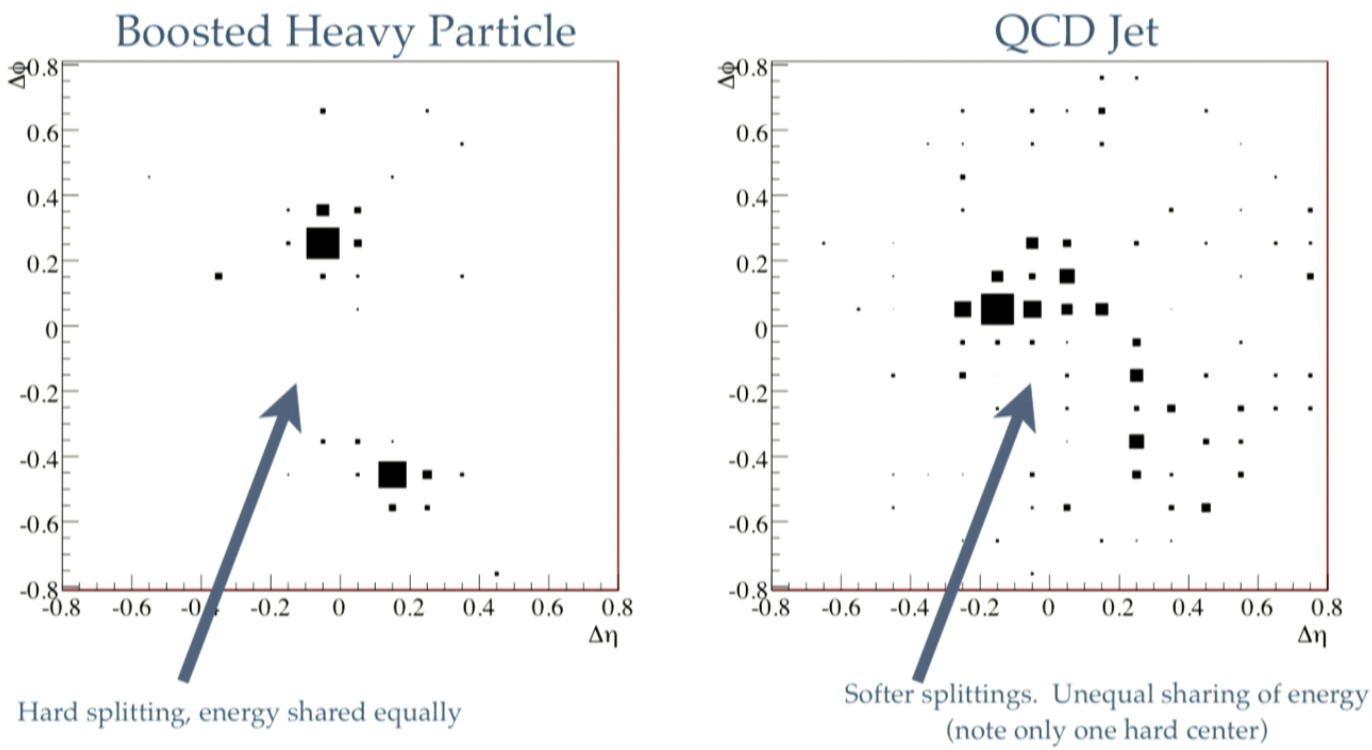
Tools

- QCD jets look really different than the jets of boosted heavy objects.
- QCD has soft/collinear singularities.
  - If we start with a high energy gluon/quark, it wants to emit soft/collinear gluons:

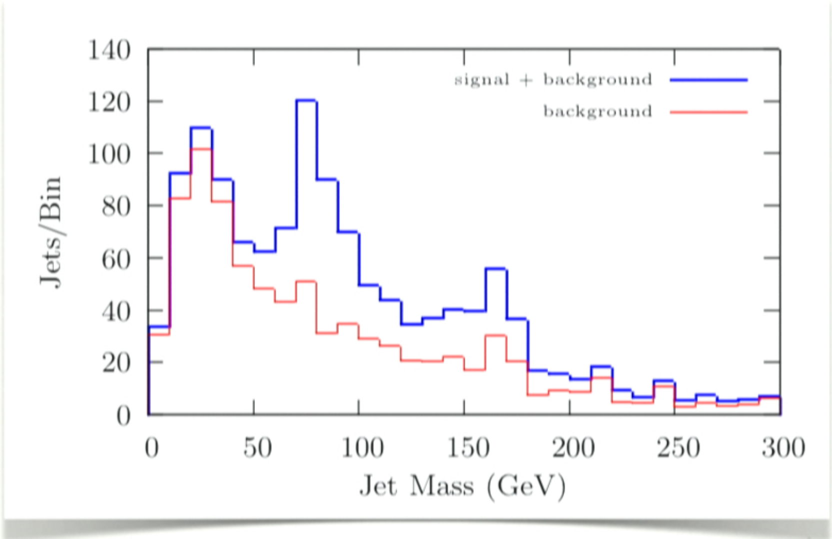
$$P_{q\to qg}(z) = C_F\,\frac{1+z^2}{1-z},$$
$$P_{g\to gg}(z) = C_A\left[\frac{1-z}{z} + \frac{z}{1-z} + z(1-z)\right]$$
$$P_{g\to q\bar{q}}(z) = T_R\left[z^2 + (1-z)^2\right],$$

  - Here P(z) measures how much a particle wants to emit another with energy fraction “z” (Altarelli-Parisi splitting fcns.).

- However, a high energy heavy particle (W/Z/t/h) just decays - it has no singularity.



- Moreover, QCD jets have a continuum mass distribution, while the jets of boosted heavy particles have a fixed mass.



- These will form our main tools.

  1. Jet radiation distribution

  2. Jet mass

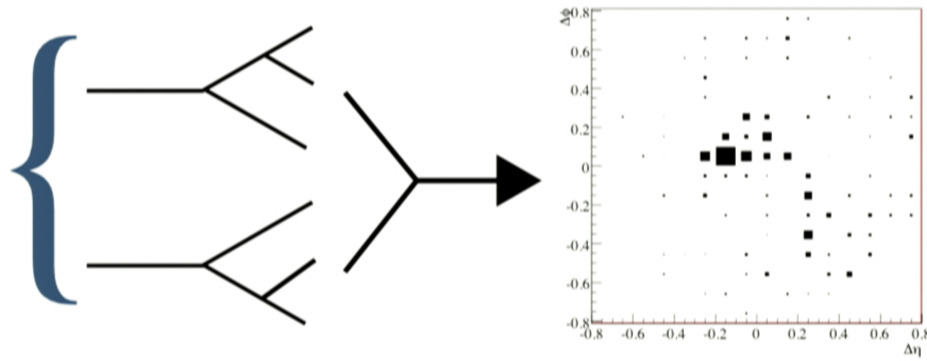
Figure source: Using jet mass to discover vector quarks at the LHC, W. Skiba, D. Tucker-Smith, [hep-ph/0701247] Phys.Rev. D75 (2007) 115010

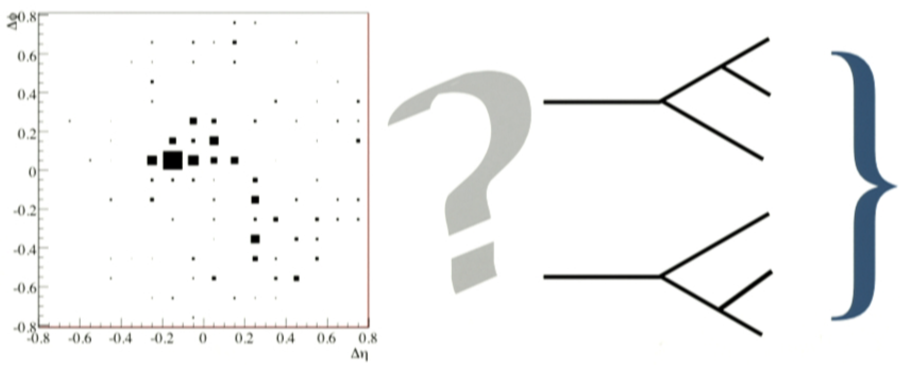
Two Basic Approaches to Substructure

1. Consider only the two-dimensional distribution of energy in a jet
   - Examples: Trimming & Filtering, N-Subjettiness, Jet substructure w/o trees
2. Try to associate a tree structure with a jet
   - Allows one to use heuristic pictures of parton shower & decay chains.
   - Examples: Pruning, energy sharing variables, mass drop
   - However, the current procedure for constructing a tree is not ideal.

Mapping Jets to Trees

The energy distribution for a particular tree is unambiguous

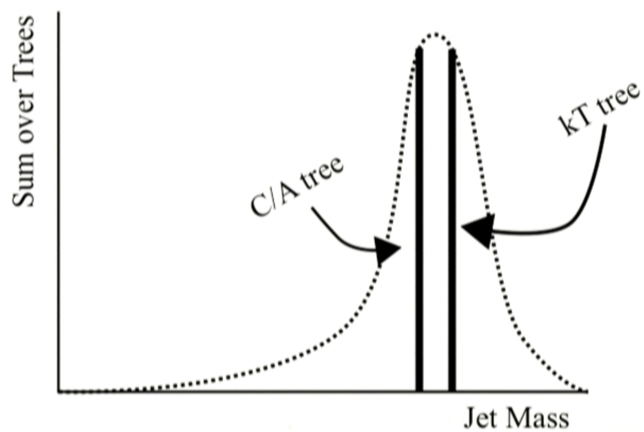




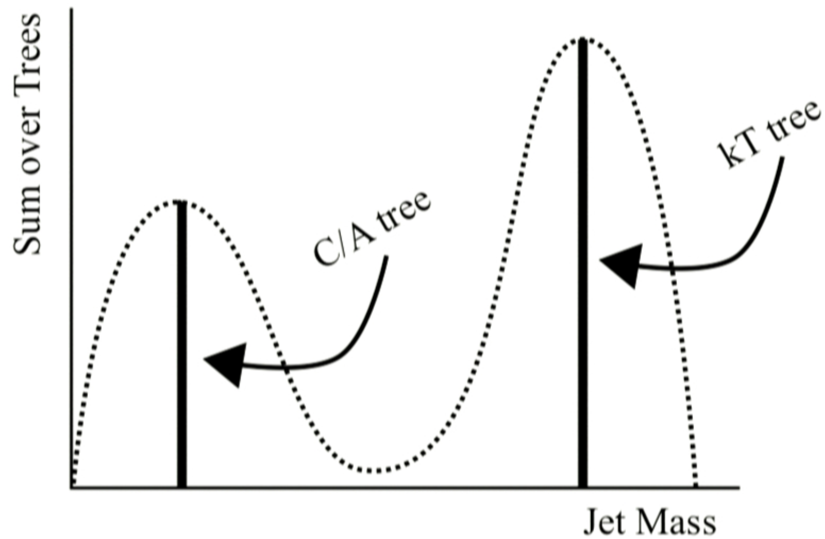
But, more than one tree can correspond to the same energy distribution

Unnecessary Choices

- How do we assign a particular tree to an energy distribution?
- Standard answer: Use a well motivated algorithm like C/A or kT
- Ideally, since both are well motivated algorithms they'll give the same answer:



- However, sometimes the answers are very different.

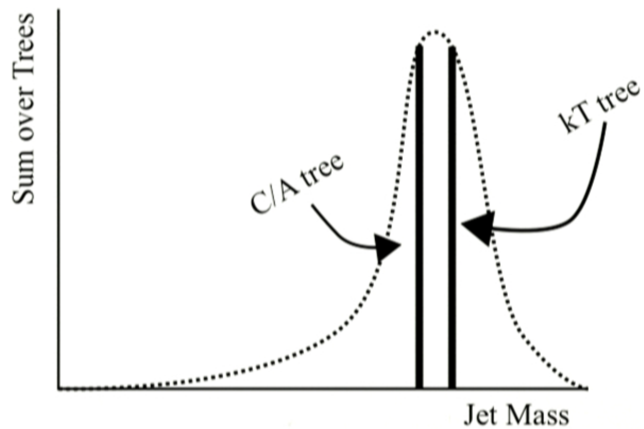


- Considering only the kT or C/A tree introduces an element of randomness into this process, resulting in unnecessary fluctuations in the final state observable.

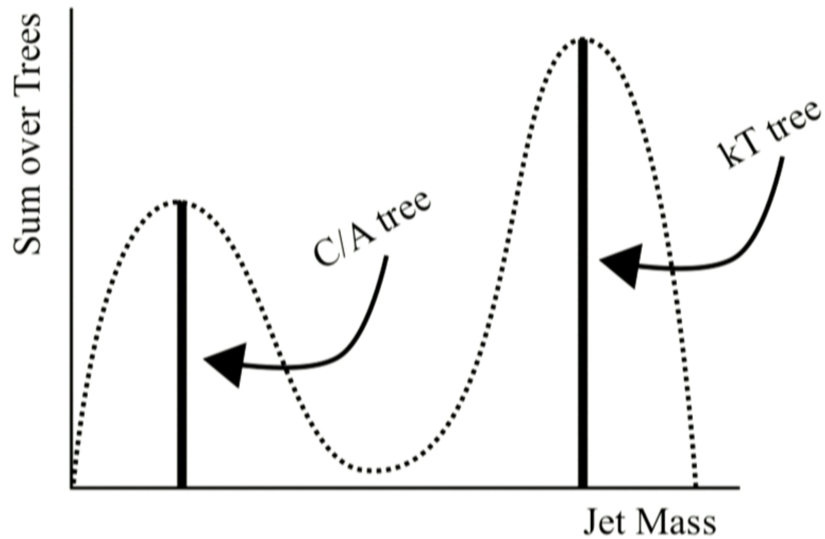
- Intuitively it makes sense that defining an observable in a way which reflects the ambiguity of this clustering should yield better results.

Unnecessary Choices

- How do we assign a particular tree to an energy distribution?
- Standard answer: Use a well motivated algorithm like C/A or kT
- Ideally, since both are well motivated algorithms they'll give the same answer:



- However, sometimes the answers are very different.

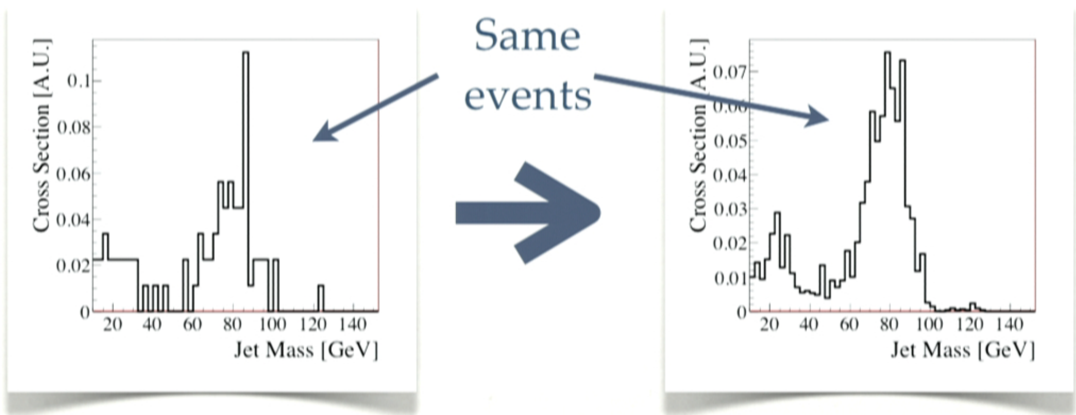


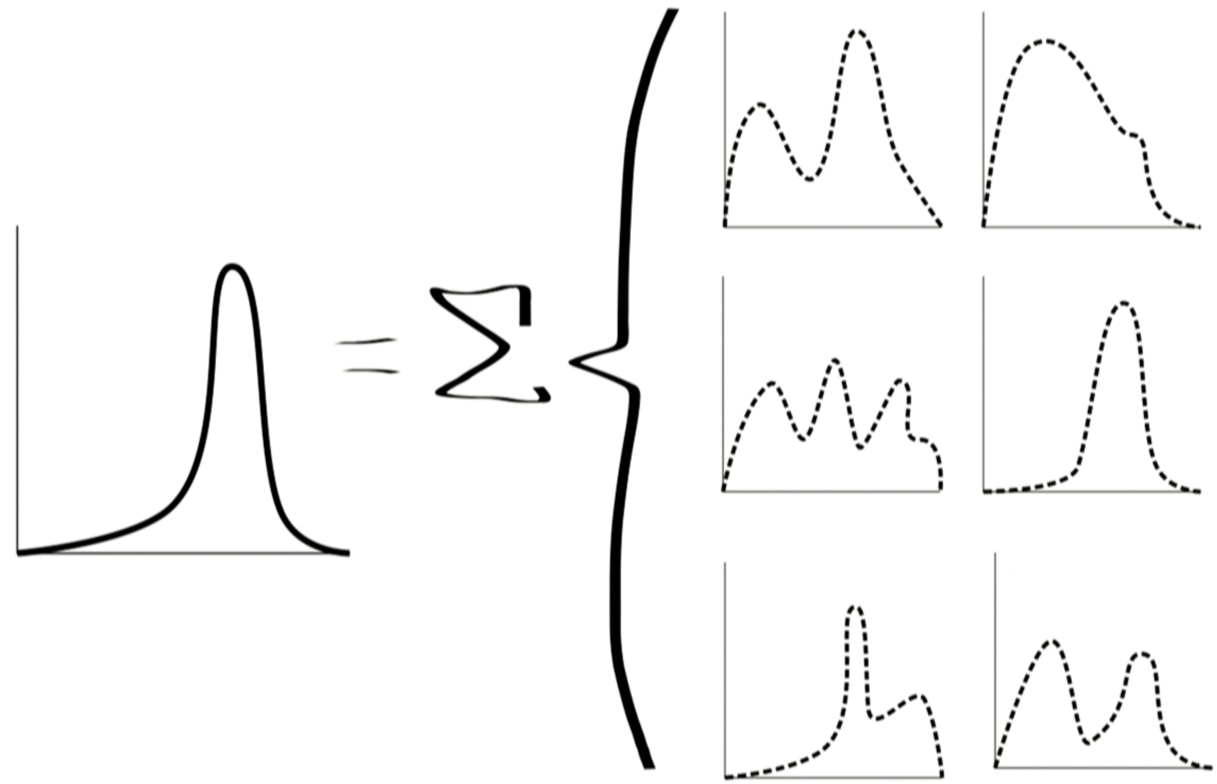
- Considering only the kT or C/A tree introduces an element of randomness into this process, resulting in unnecessary fluctuations in the final state observable.

- Intuitively it makes sense that defining an observable in a way which reflects the ambiguity of this clustering should yield better results.

Solution: Sum over Trees

- We propose that rather than assigning a single number to each event, instead each event should contribute a distribution obtained by summing the observable over many trees.
- When we sum these together, the result is much more stable than the histogram we would have had if we just considered one number per event.





Weighting algorithms

- The only question is: when we add together the result obtained from different trees, how should we weight each tree's contribution?
- Surely they should not all count equally. If they did, then why would we use kT or C/A to find our trees in the first place?
- In theory, one could weight each tree by the product of splitting functions and Sudakovs one would obtain from a parton shower.
  - Work in progress.

Implementation

- Instead, we find a simpler Monte-Carlo procedure works quite well.
  - As in a sequential recombination algorithm, assign every pair of proto-jets a distance measure d_{ij}.
  - However, unlike a normal sequential algorithm (where the pair with the smallest measure is selected clustered), here we suggest that a given pair be randomly selected for merging with probability

$$\Omega_{ij} \equiv \frac{1}{\Omega} \exp\left(-\alpha \frac{d_{ij}}{d_{ij}^{\min}}\right), \quad \alpha = \text{rigidity parameter}$$

  - Thus, paths which deviate from the CA or kT behavior are less likely to occur
  - Repeat many (~100) times, till the distribution stabilizes

- The result is that you get many trees
- The probability of finding a given tree decreases as it becomes less k_T or C/A like
- Available as a Fastjet plugin:

  http://jets.physics.harvard.edu/Qjets

IR/Collinear Safety

- As long as the rigidity variable (alpha) is non-zero, then infinitely soft or collinear particles will not change the observable at hand.
- How will this affect real analytical calculations?
  - Still unknown
  - Perhaps there is a better, more theory-friendly weight?

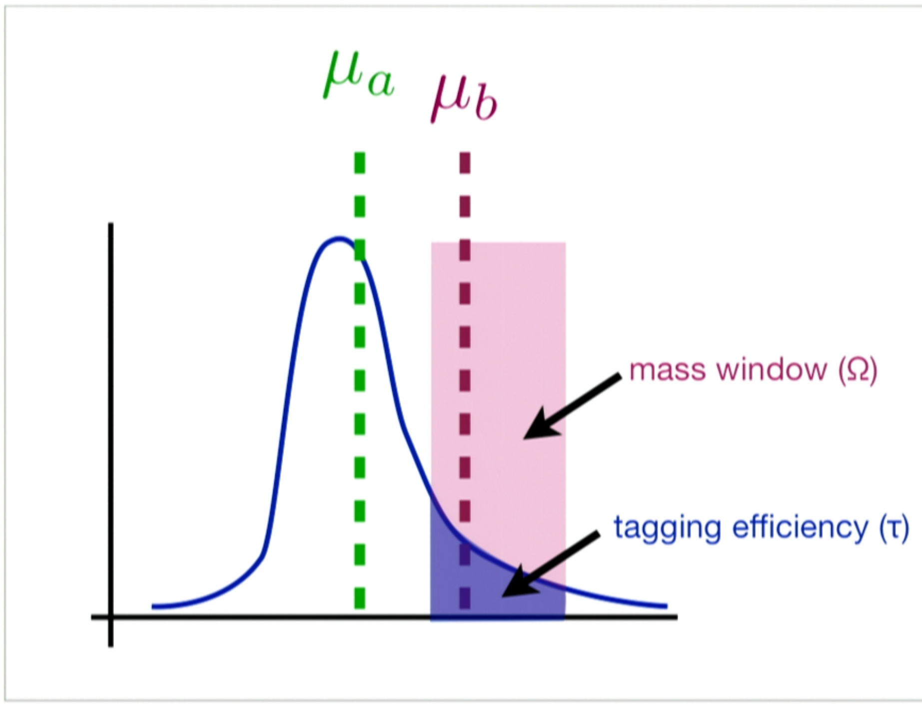
Implementation

- Instead, we find a simpler Monte-Carlo procedure works quite well.
  - As in a sequential recombination algorithm, assign every pair of proto-jets a distance measure d_{ij}.
  - However, unlike a normal sequential algorithm (where the pair with the smallest measure is selected clustered), here we suggest that a given pair be randomly selected for merging with probability

$$\Omega_{ij} \equiv \frac{1}{\Omega} \exp\left(-\alpha \frac{d_{ij}}{d_{ij}^{\min}}\right), \ \alpha = \text{rigidity parameter}$$

  - Thus, paths which deviate from the CA or kT behavior are less likely to occur
  - Repeat many (~100) times, till the distribution stabili

Weighted Interpretations



Better Statistics

- *Weighted events are better for statistics than unweighted events!*
- Consider reconstructing events subject to some reconstruction efficiency
  - Unweighted events $$\frac{\delta N_R}{N_R} = \frac{1}{\sqrt{\epsilon N}}$$
  - Weighted events: $$\frac{1}{\sqrt{N}} \leq \frac{\delta N_R}{N_R} \leq \frac{1}{\sqrt{\epsilon N}}$$

Unweighted events: $\mathrm{P}(x) = \mathrm{Pois}(x|\epsilon n) = \sum_{a=x}^{\infty} \mathrm{Pois}(a|N) \times \mathrm{B}(x|a,\epsilon)$

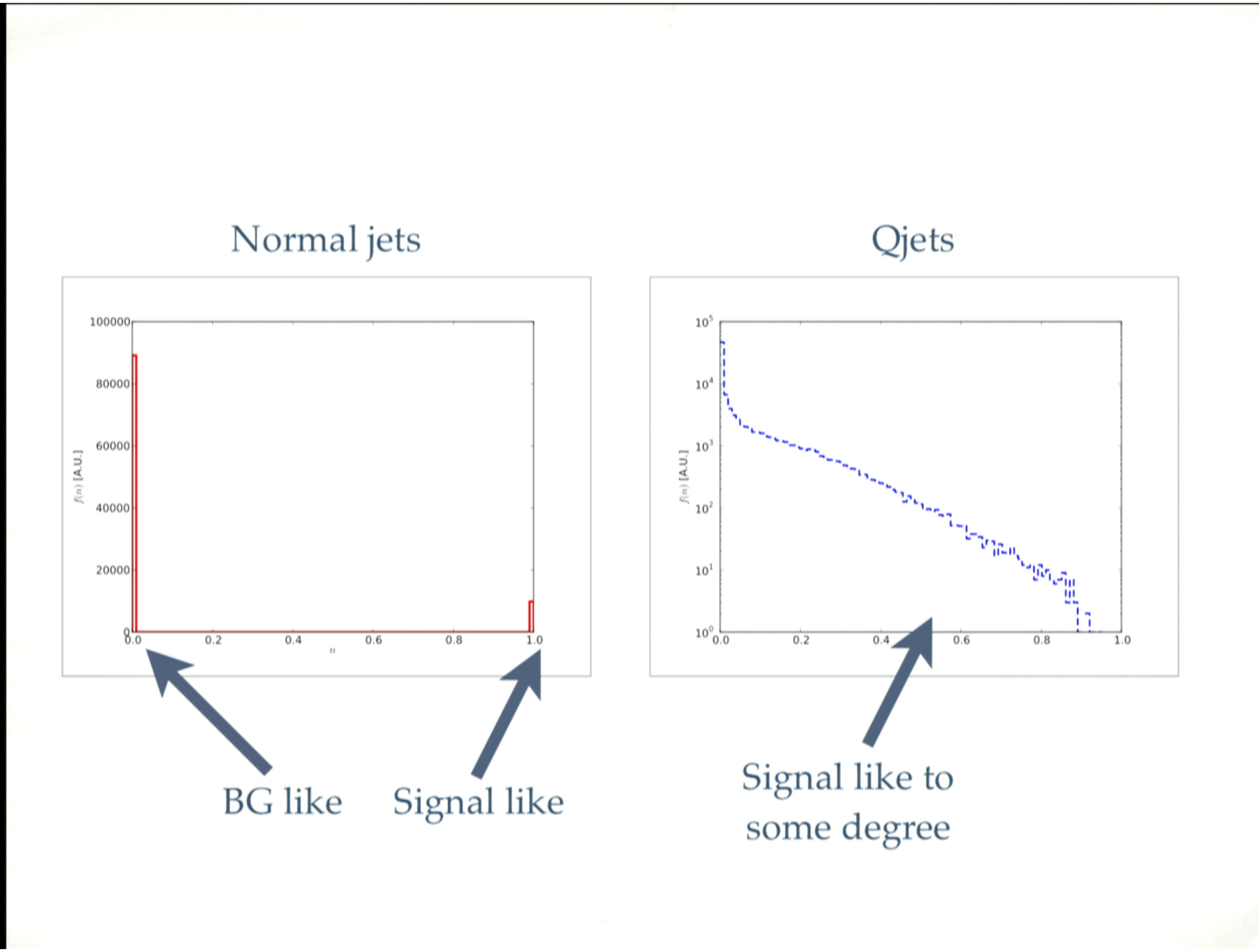
$$\boxed{\frac{\delta N_R}{N_R} = \frac{1}{\sqrt{\epsilon N}}.}$$

$$\mathrm{Pois}(n|N) \equiv \frac{e^{-N} N^n}{n!} \qquad \mathrm{B}(a|n,\epsilon) \equiv \frac{n!}{a!(n-a)!} \epsilon^a (1-\epsilon)^{n-a}$$

Weighted events: $\mathrm{P}(x) = \sum_{a=x}^{\infty} \mathrm{Pois}(a|N) \times \int dN_R f(N_R|N)$

$$\boxed{\frac{\delta N_R}{N_R} = \frac{1}{\sqrt{N}} \times \sqrt{1 + \frac{\sigma_1^2}{\epsilon^2}}}$$

$$\epsilon = \int dN_R f(N_R|1) \qquad \sigma_1^2 = \int dN_R (N_R - \epsilon)^2 f(N_R|1)$$


Normal jets
Qjets
100000
80000
60000
40000
20000
0
0.0
0.2
0.4
0.6
0.8
1.0
f(n) [A.U.]
BG like
Signal like
Signal like to
some degree

Unweighted events: $P(x) = \text{Pois}(x|\epsilon n) = \sum_{a=x}^{\infty} \text{Pois}(a|N) \times B(x|a,\epsilon)$

$$\boxed{\frac{\delta N_R}{N_R} = \frac{1}{\sqrt{\epsilon N}}.}$$

$$\text{Pois}(n|N) \equiv \frac{e^{-N} N^n}{n!} \qquad B(a|n,\epsilon) \equiv \frac{n!}{a!(n-a)!}\epsilon^a (1-\epsilon)^{n-a}$$

Weighted events: $P(x) = \sum_{a=x}^{\infty} \text{Pois}(a|N) \times \int dN_R f(N_R|N)$

$$\boxed{\frac{\delta N_R}{N_R} = \frac{1}{\sqrt{N}} \times \sqrt{1 + \frac{\sigma_1^2}{\epsilon^2}}}$$

$$\epsilon = \int dN_R f(N_R|1) \qquad \sigma_1^2 = \int dN_R (N_R - \epsilon)^2 f(N_R|1)$$

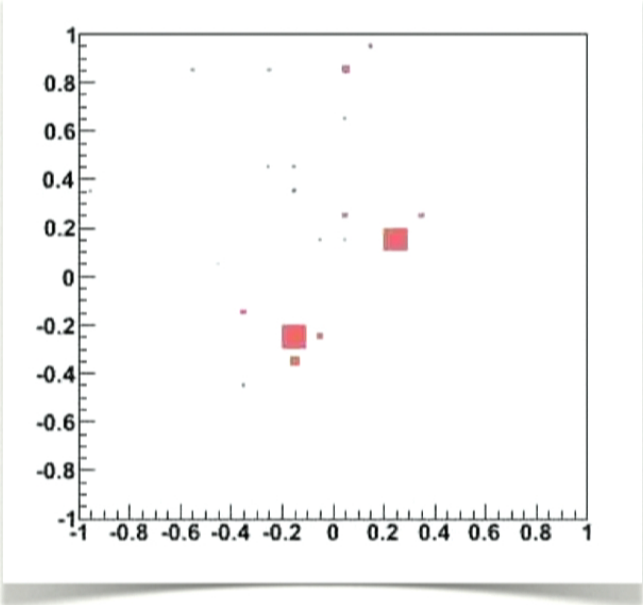
- In practice determine moments of *f* from MC (tau here is the % of the interpretations tagged as signal-like)

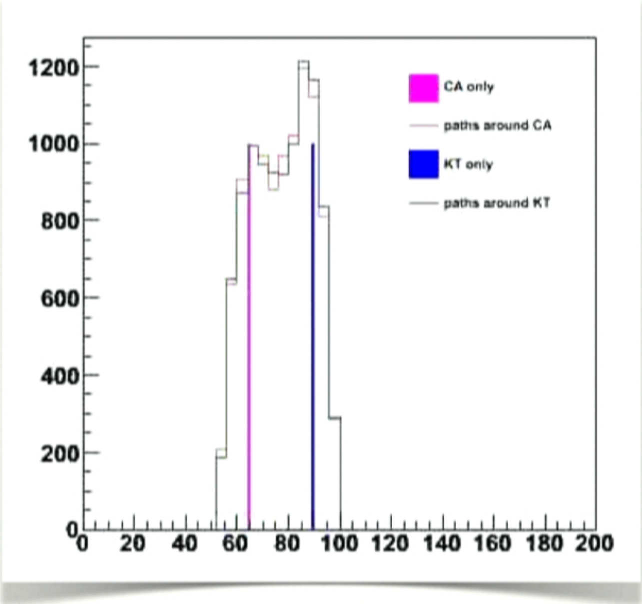
$$\epsilon = \frac{1}{N_{\text{jets}}} \sum_{\text{jets}} \tau, \quad \sigma_1^2 = \frac{1}{N_{\text{jets}}} \sum_{\text{jets}} (\tau - \epsilon)^2$$

- Similar results hold for any measurement whose error drops as 1/sqrt(N) - e.g. avg. jet mass

- Formal discussion in "*The Statistical Properties of Qjets*" with Ellis, Hornig, DK, Roy, Schwartz

Example: Boosted W-Jets with Pruning



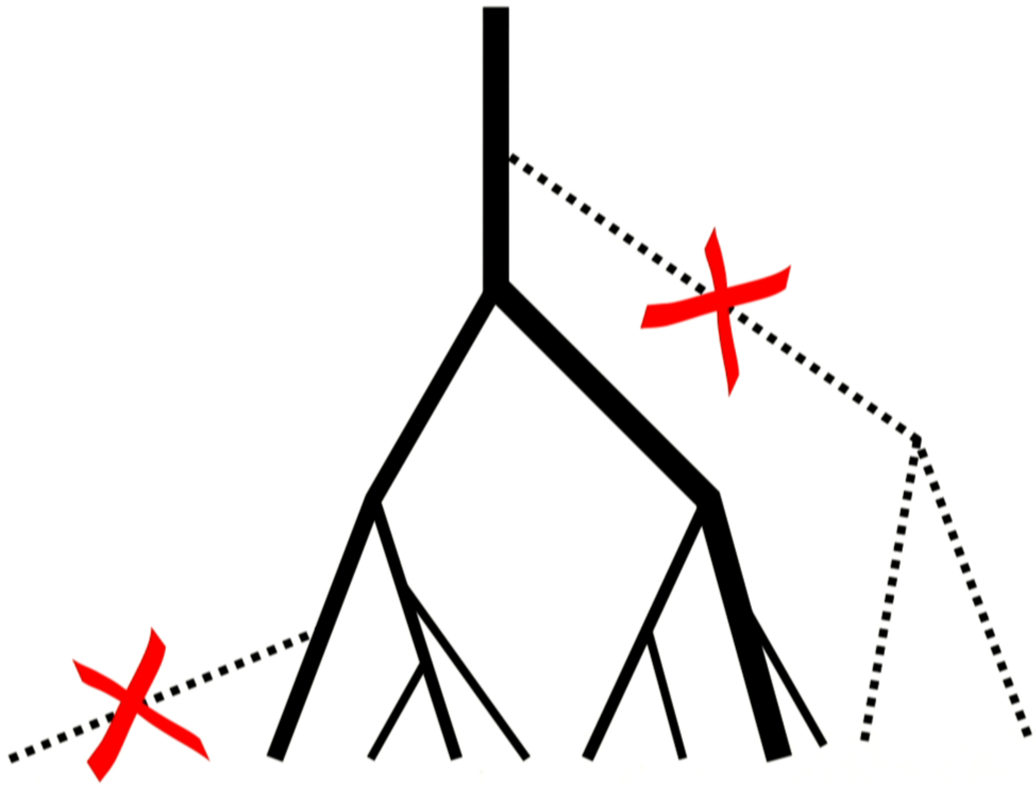


Pruning

- Pruning was introduced to look for boosted heavy objects (e.g., tops, higgses, W's, etc) by cleaning up their mass.
- Intuition: QCD has soft/collinear singularities. Wide-angle emissions should come from hard decays.
  - Remove all parts of the jet which are *both* soft and wide angle.
- Two main advantages:
  - Boosted objects see their mass reconstruction improved
  - Massive QCD jets (a large background) see their mass substantially decreased -> lower backgrounds

Pruning (Ellis, Vermilion, Walsh - 0903.5081, 0912.0033)

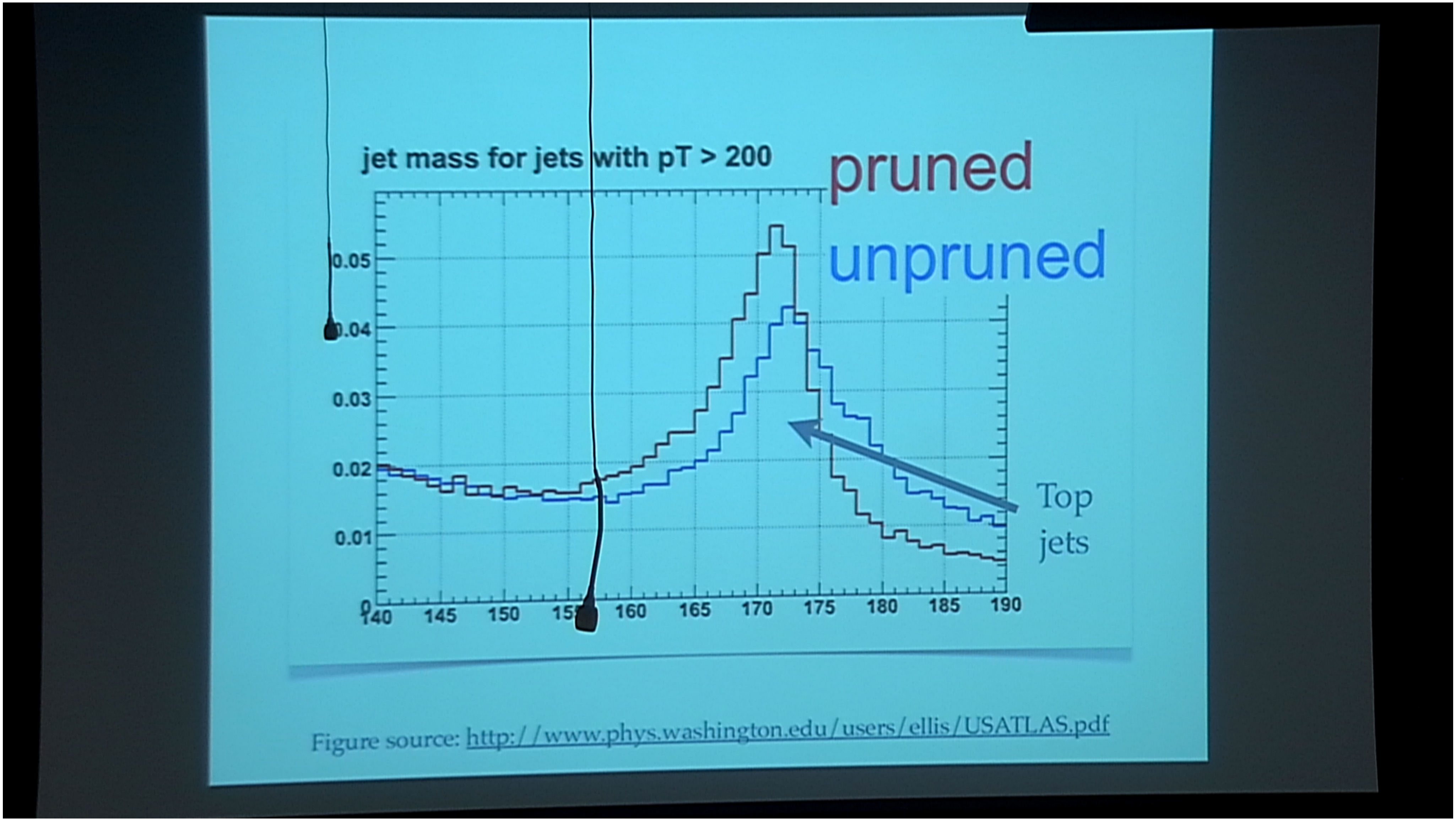
A Pruned Tree



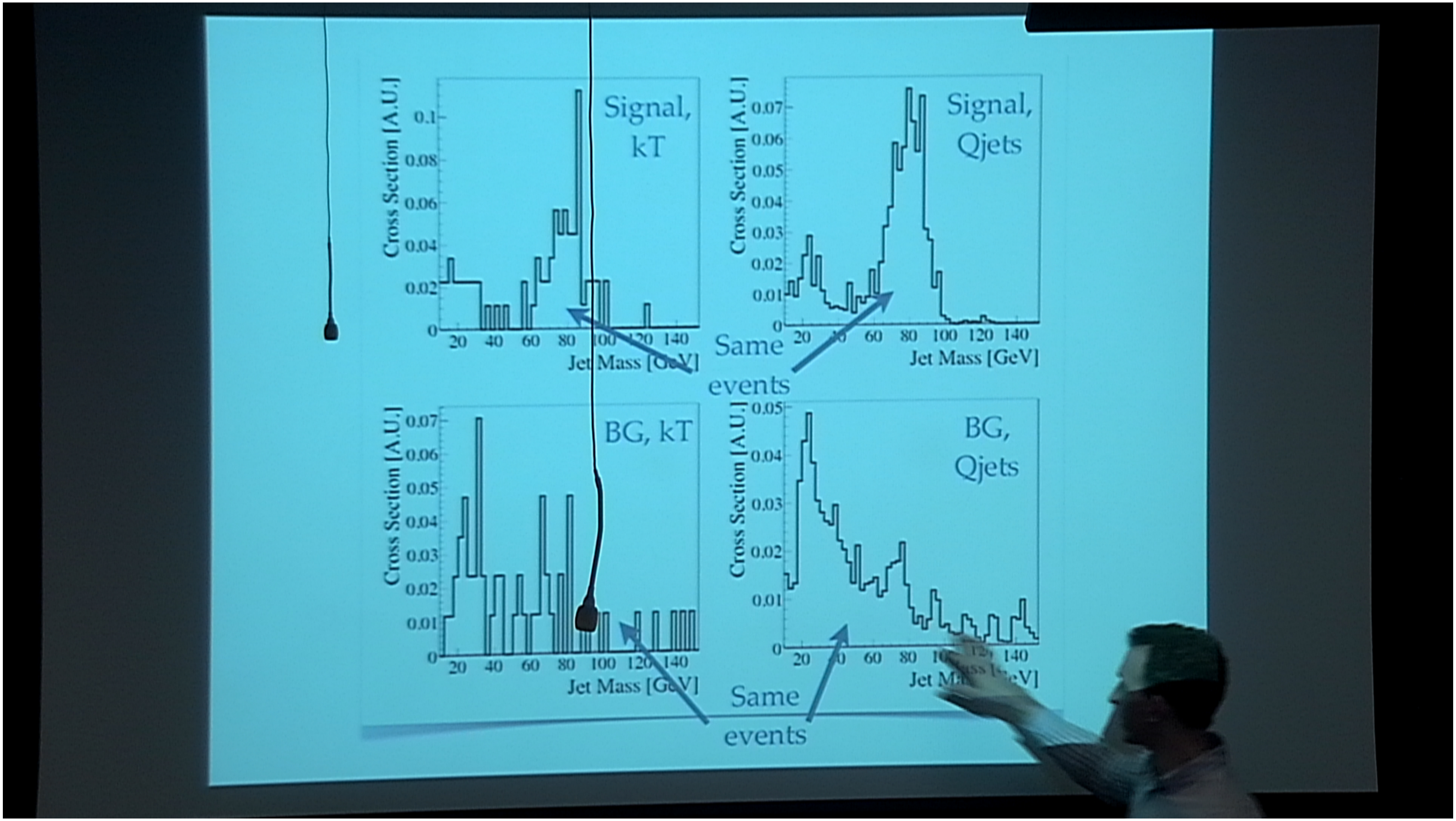
Pruning

- Pruning was introduced to look for boosted heavy objects (e.g., tops, higgses, W's, etc) by cleaning up their mass.
- Intuition: QCD has soft/collinear singularities. Wide-angle emissions should come from hard decays.
  - Remove all parts of the jet which are *both* soft and wide angle.
- Two main advantages:
  - Boosted objects see their mass reconstruction improved
  - Massive QCD jets (a large background) see their mass substantially decreased -> lower backgrounds

Pruning (Ellis, Vermilion, Walsh - 0903.5081, 0912.0033)



jet mass for jets with pT > 200
pruned
unpruned
Top jets
0.05
0.04
0.03
0.02
0.01
0
140
145
150
160
165
170
175
180
185
190
Figure source: http://www.phys.washington.edu/users/ellis/USATLAS.pdf


Signal, kT
Signal, Qjets
BG, kT
BG, Qjets
Same events
Same events
Cross Section [A.U.]
Jet Mass [GeV]

Example 2/3: Signal Discovery &Exclusion

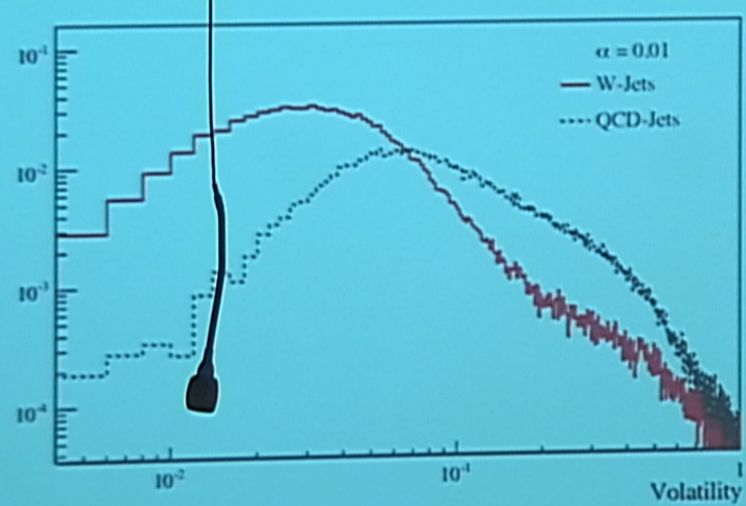
- Signal = boosted W-jets, pT > 500
- BG = light QCD jets, pT > 500
- Measure the signal size in a bin (here 70-90 GeV) and compare it to the size of the BG fluctuations (Poisson stats included)
- Need only ~70% the luminosity to have the same significance

$$S/\delta B \propto \sqrt{N}$$

α	$\frac{\langle S\rangle/\delta B\vert_Q}{\langle S\rangle/\delta B\vert_{cl}}$
0.0	1.07
0.01	1.13
0.1	1.18
1.0	1.14
100	1.06

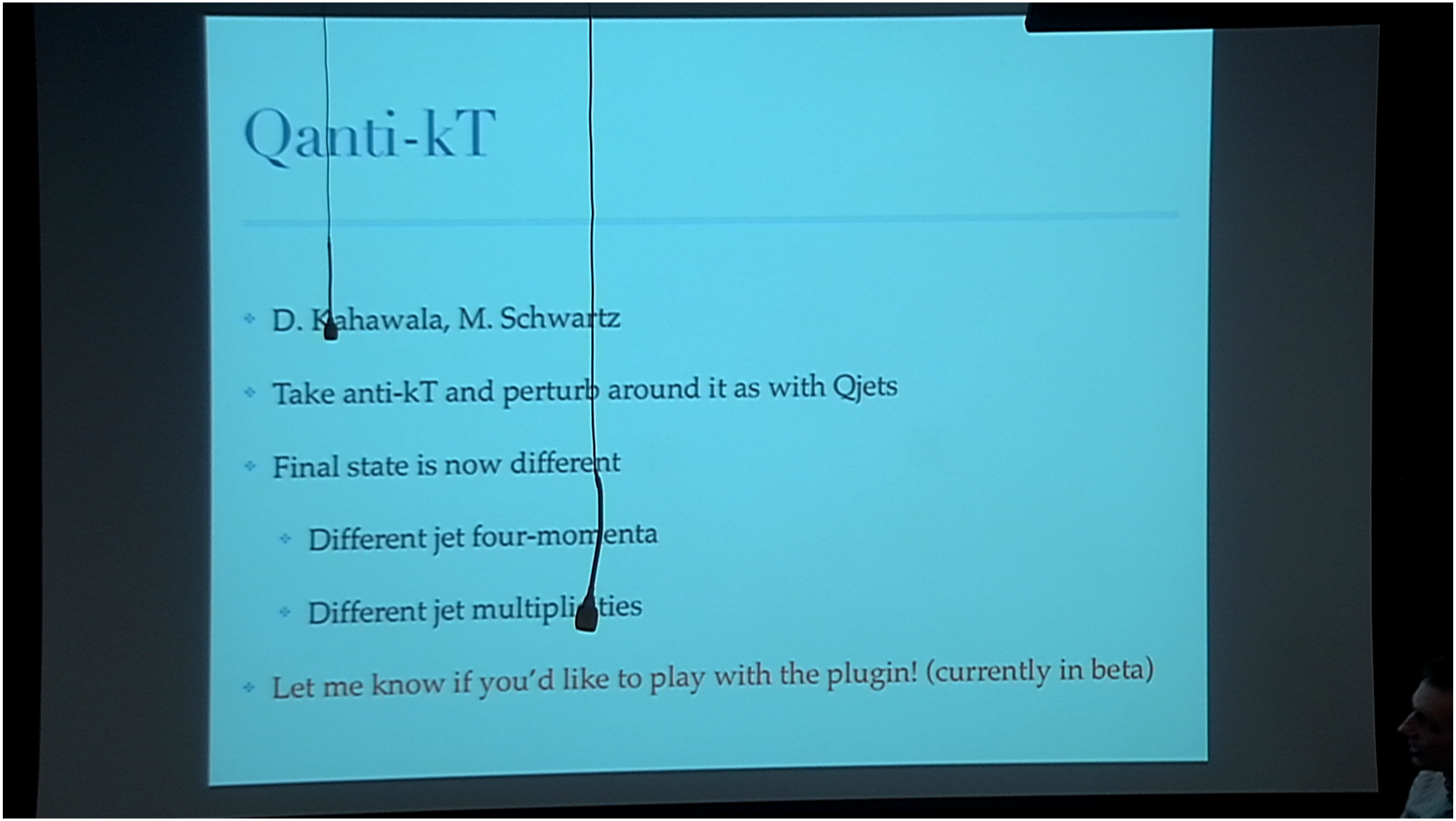
Width to Mass Distribution

- volatility = width of pruned mass distribution



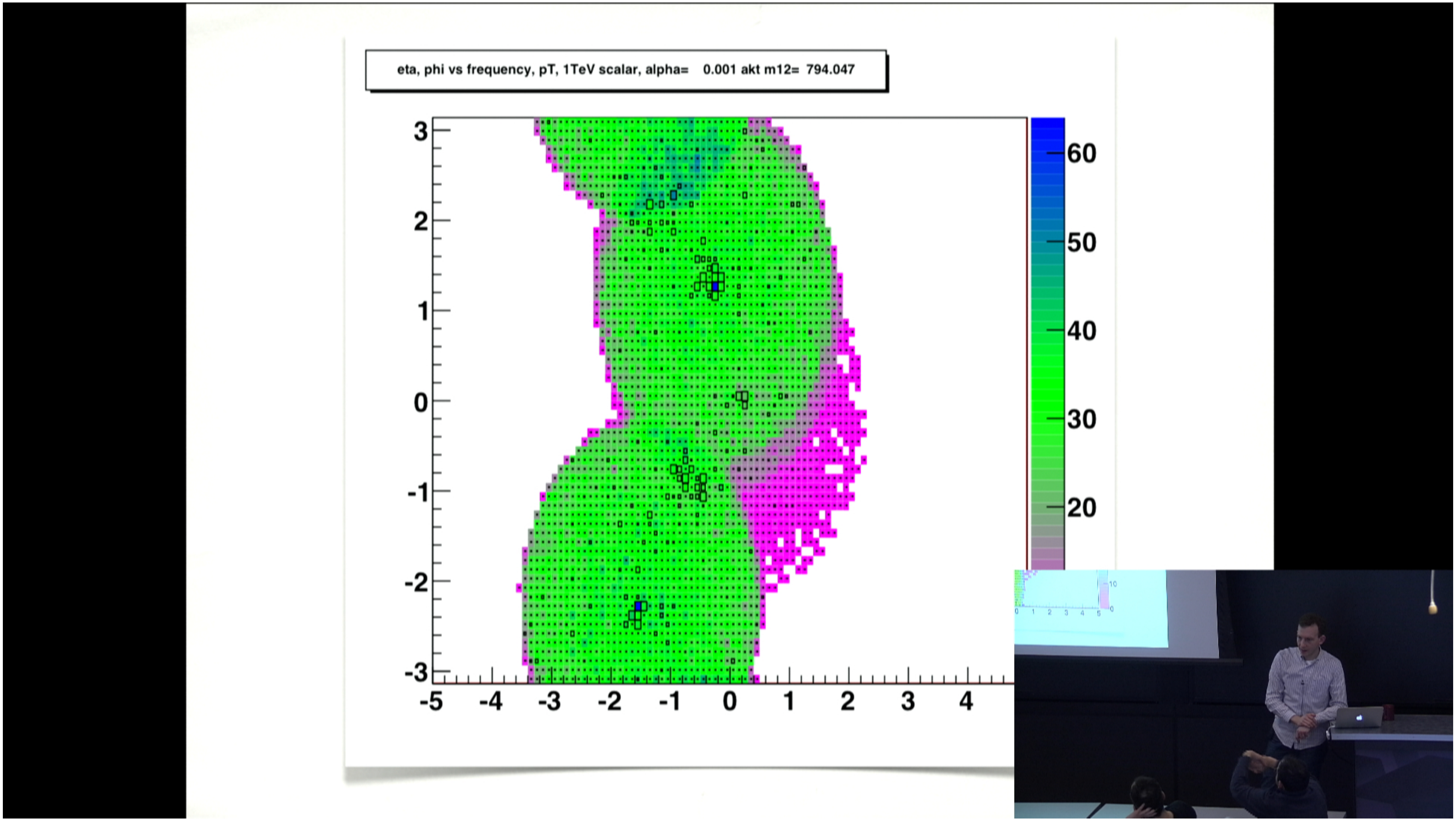
Q-everything

- Perhaps we should consider "summing" over multiple parameters, not just trees.
  - Jet radii, trimming parameters, etc.
- We've only looked at considering multiple tree structures for the radiation inside a jet.
  - Could this help with precision quantities like y_{23}?

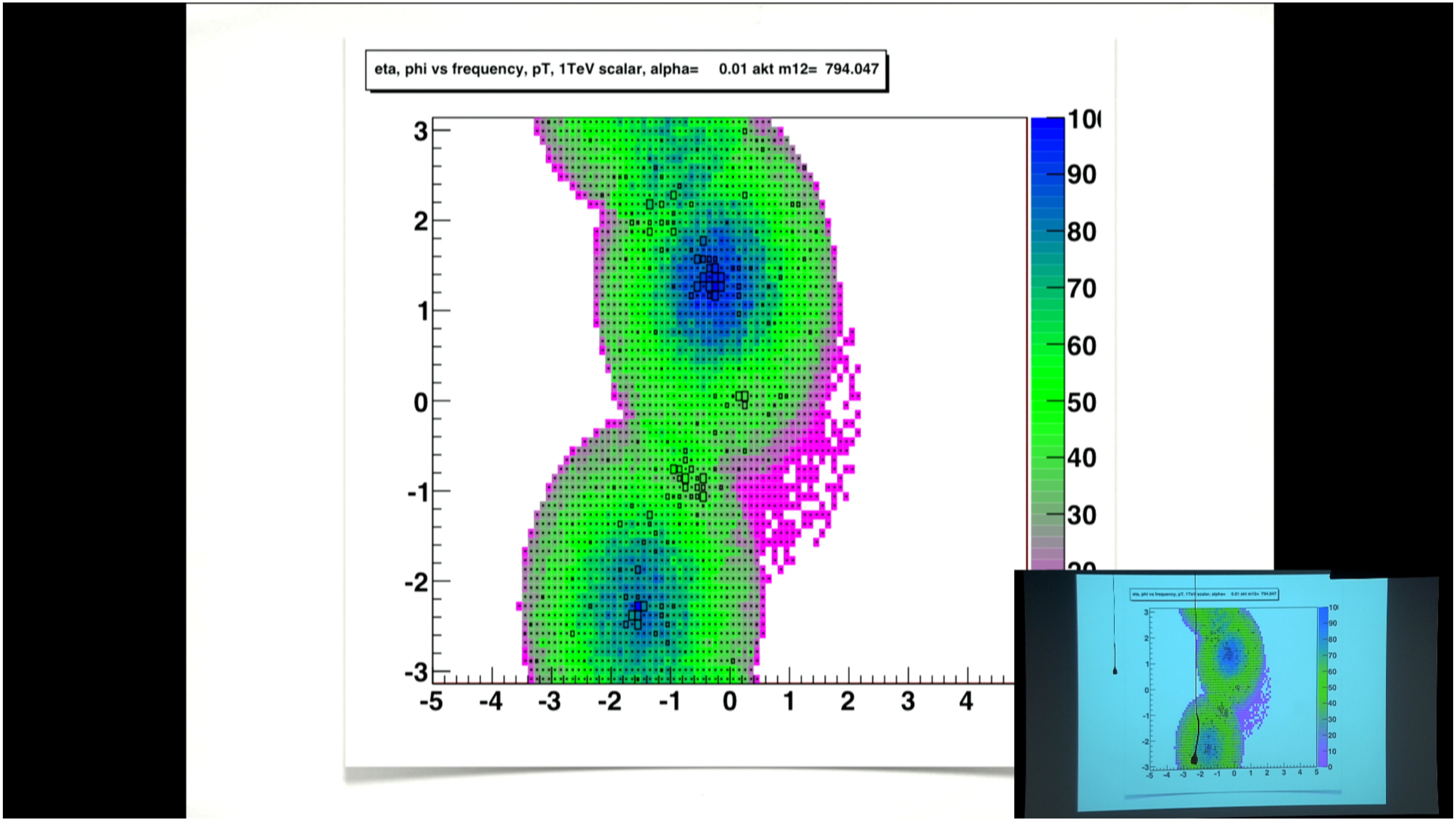

Qanti-kT
D. Kahawala, M. Schwartz
Take anti-kT and perturb around it as with Qjets
Final state is now different
Different jet four-momenta
Different jet multiplicities
Let me know if you'd like to play with the plugin! (currently in beta)

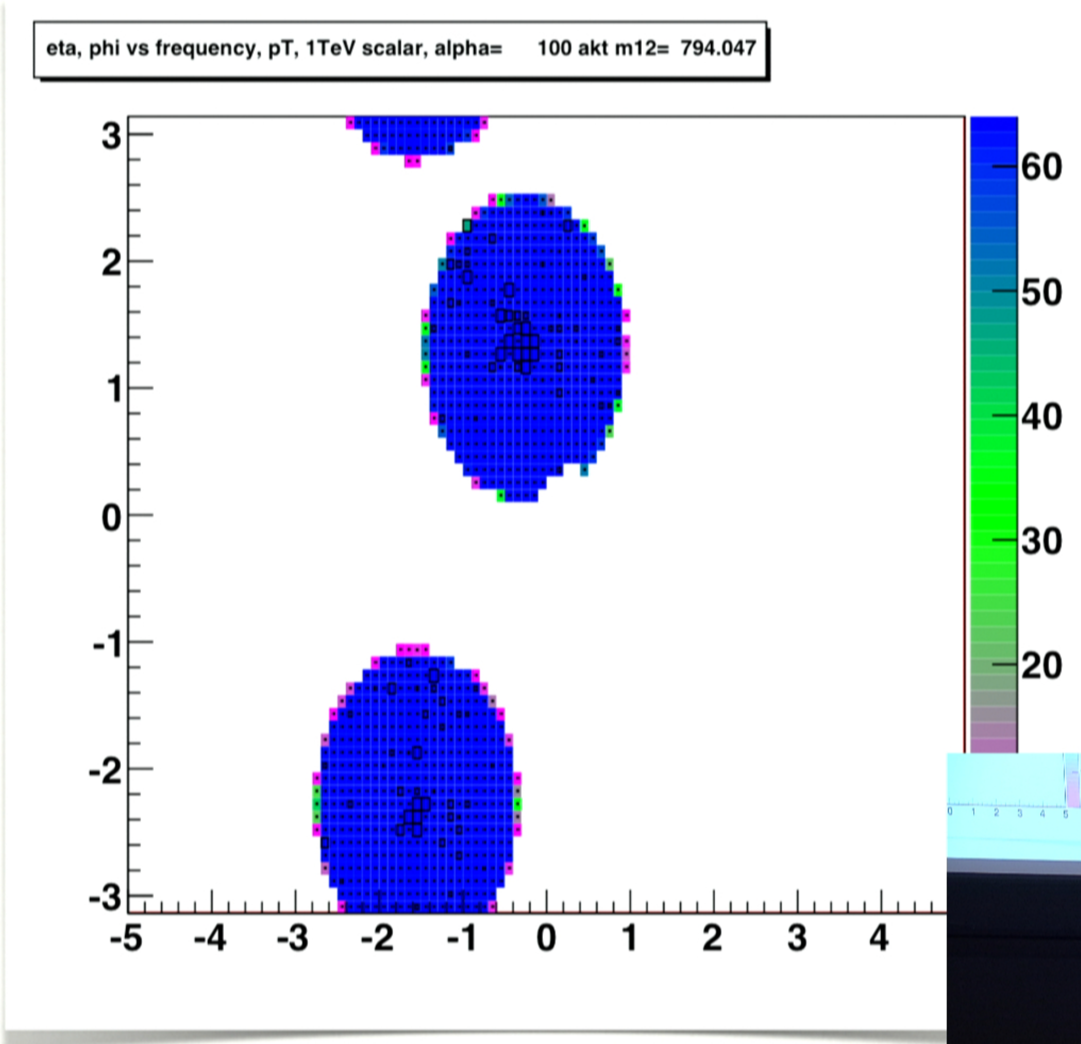
Qanti-kT

- D. Kahawala, M. Schwartz
- Take anti-kT and perturb around it as with Qjets
- Final state is now different
  - Different jet four-momenta
  - Different jet multiplicities
- Let me know if you'd like to play with the plugin! (currently in beta)

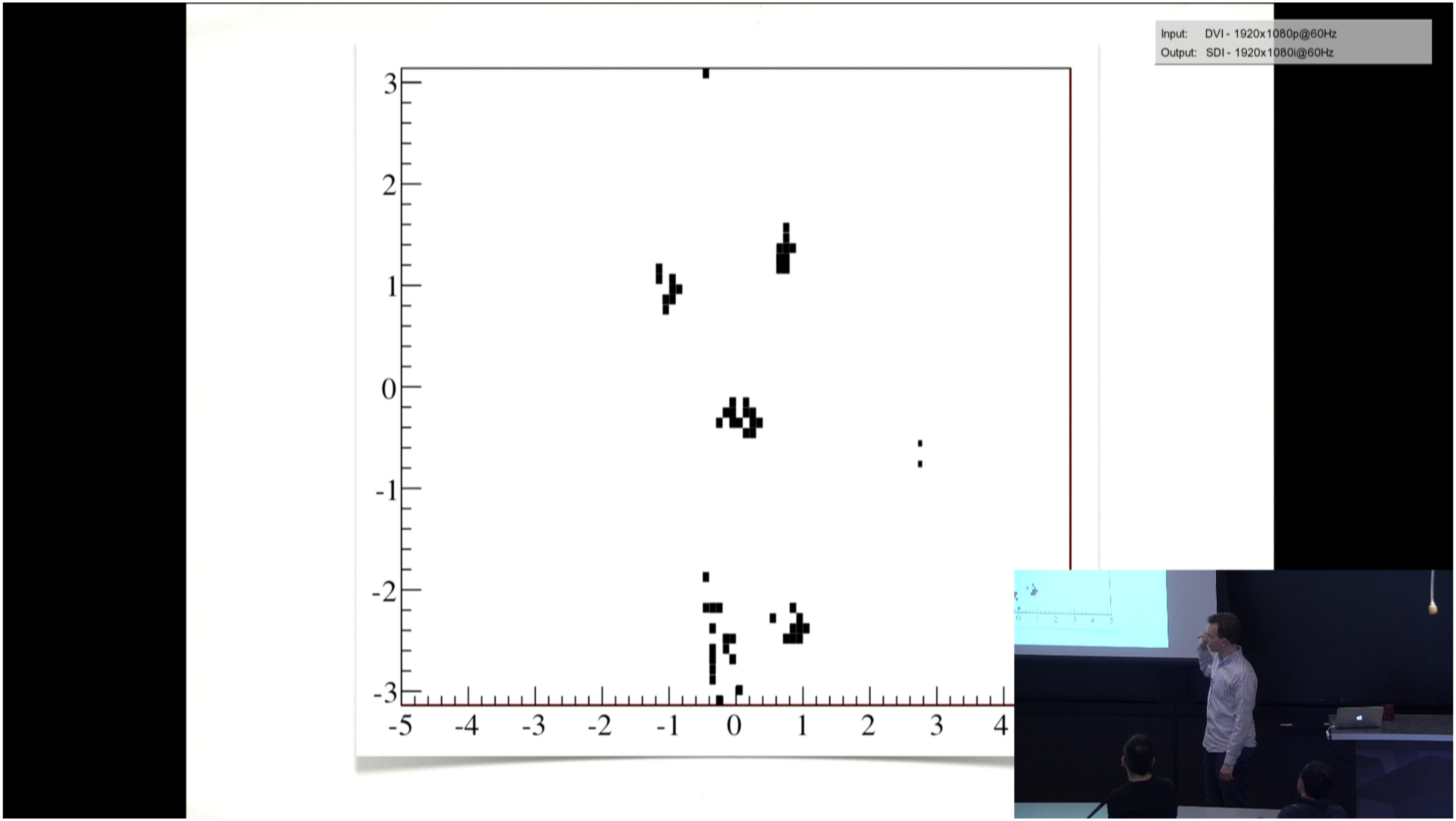


eta, phi vs frequency, pT, 1TeV scalar, alpha= 0.001 akt m12= 794.047
3
2
1
0
-1
-2
-3
-5
-4
-3
-2
-1
0
1
2
3
4
60
50
40
30
20

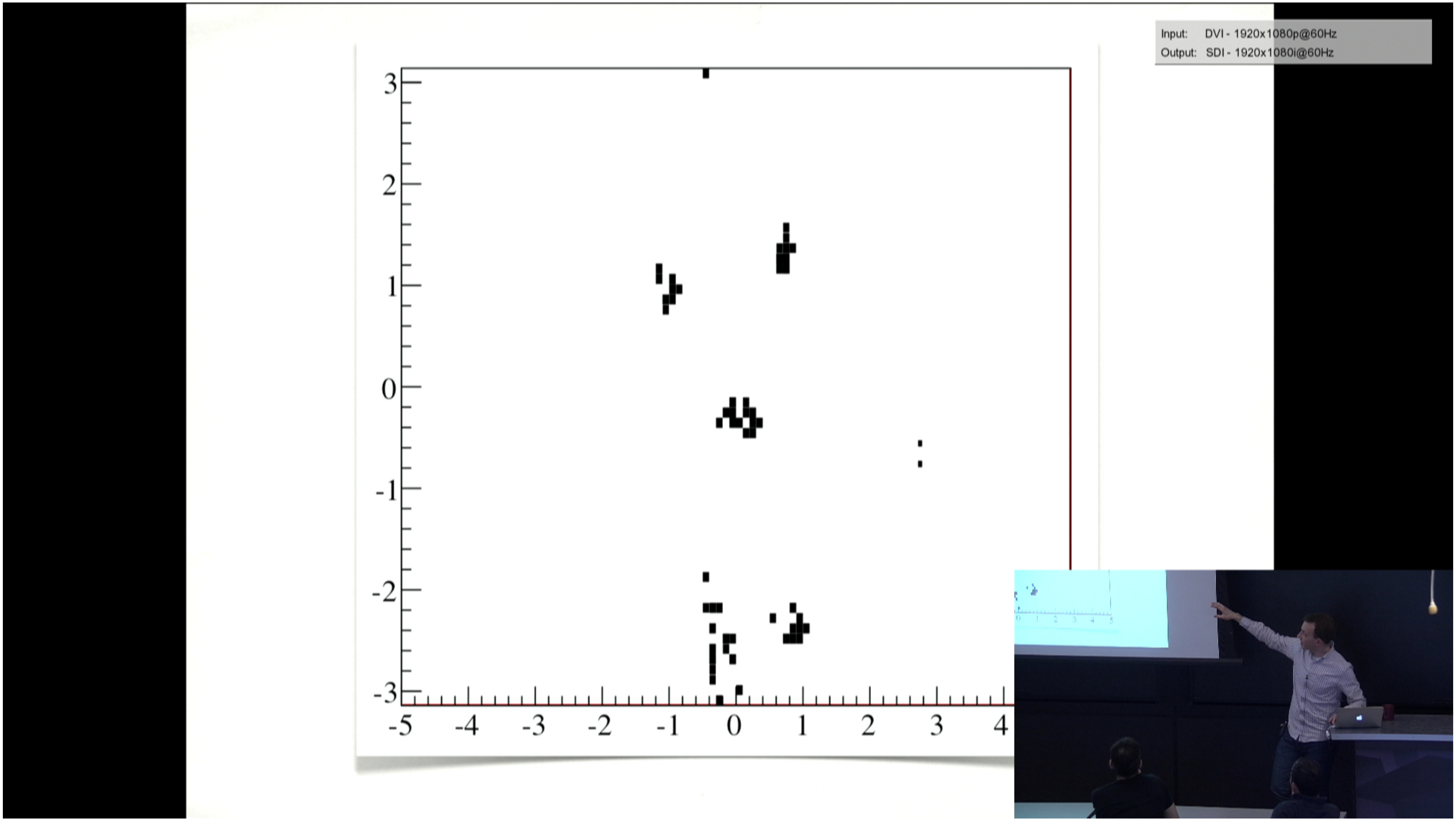

eta, phi vs frequency, pT, 1TeV scalar, alpha=   0.01 akt m12= 794.047
3
2
1
0
-1
-2
-3
-5
-4
-3
-2
-1
0
1
2
3
4
10(
90
80
70
60
50
40
30

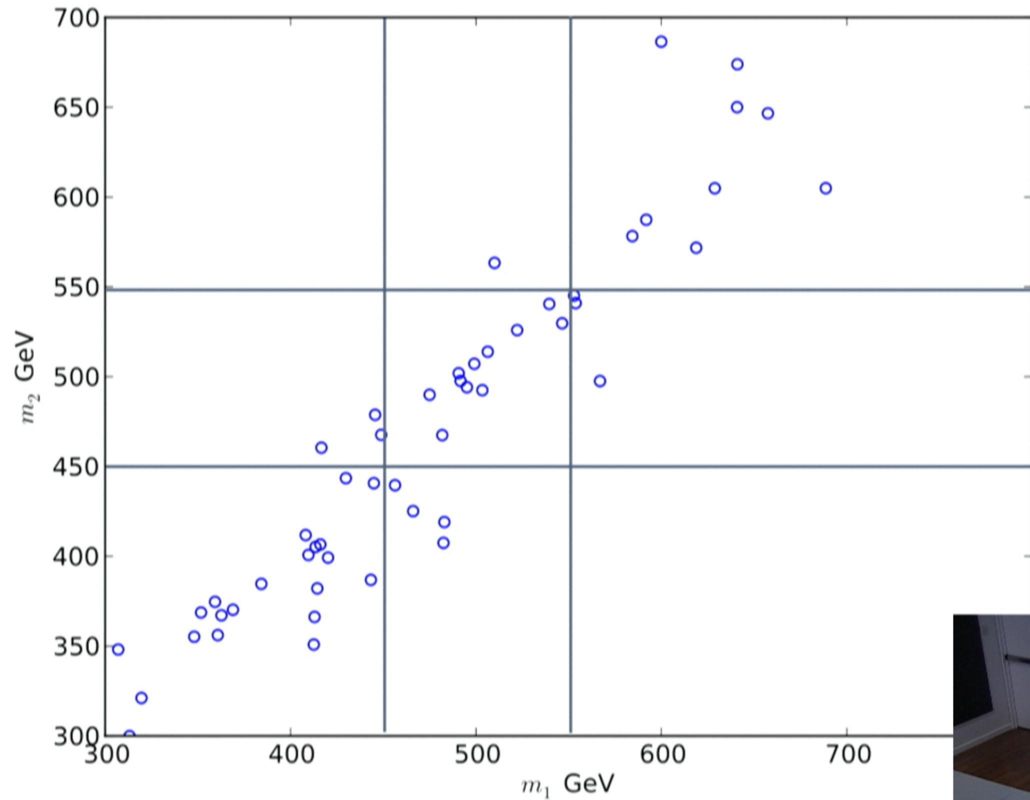

eta, phi vs frequency, pT, 1TeV scalar, alpha= 100 akt m12= 794.047
3
2
1
0
-1
-2
-3
-5 -4 -3 -2 -1 0 1 2 3 4
60
50
40
30
20

Resonance pair production



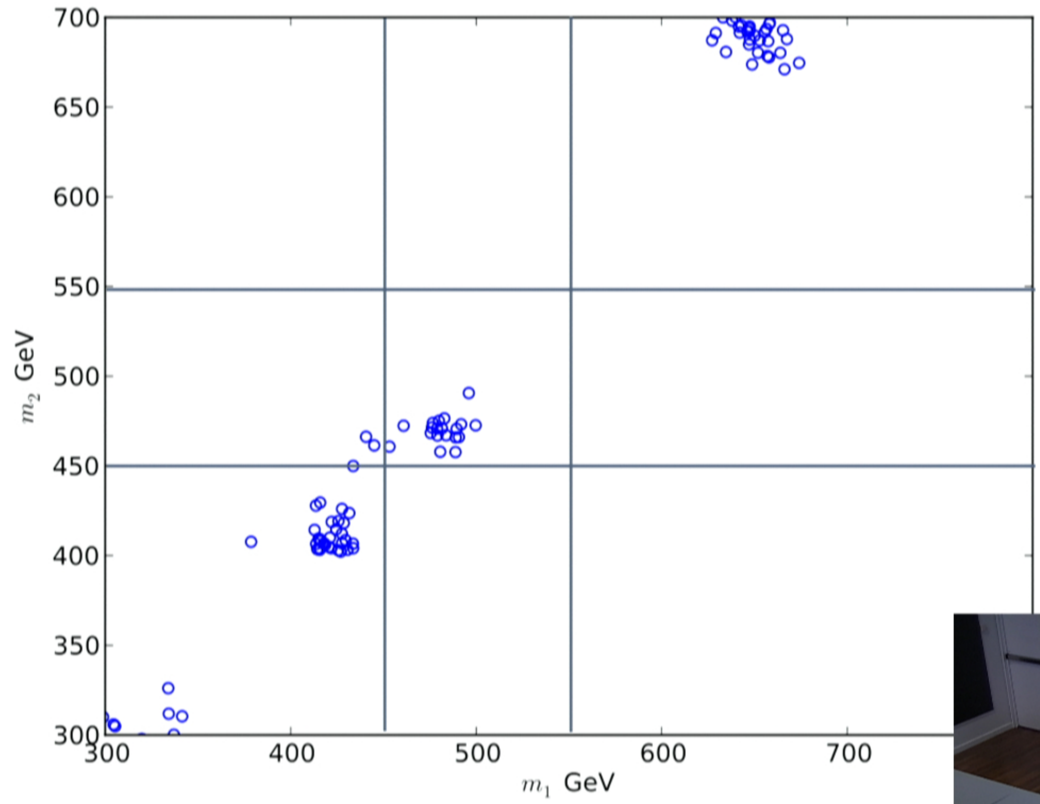
Input: DVI - 1920x1080p@60Hz
Output: SDI - 1920x1080i@60Hz
3
2
1
0
-1
-2
-3
-5
-4
-3
-2
-1
0
1
2
3
4


Input: DVI - 1920x1080p@60Hz
Output: SDI - 1920x1080i@60Hz
3
2
1
0
-1
-2
-3
-5
-4
-3
-2
-1
0
1
2
3
4



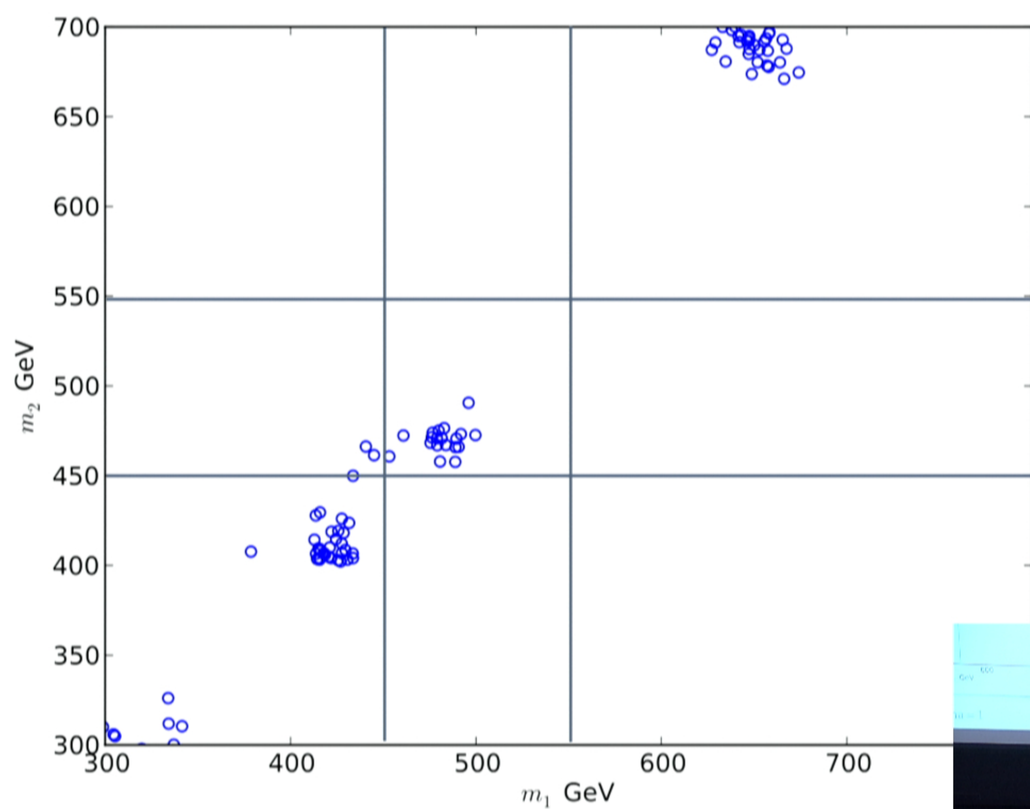
700
650
600
550
500
450
400
350
300
m_2 GeV
300
400
500
600
700
m_1 GeV

alpha = 10^{-6}



700
650
600
550
500
450
400
350
300
300
400
500
600
700
m_1 GeV
m_2 GeV

alpha = 1


700
650
600
550
500
450
400
350
300
m_2 GeV
300
400
500
600
700
m_1 GeV

alpha = 1

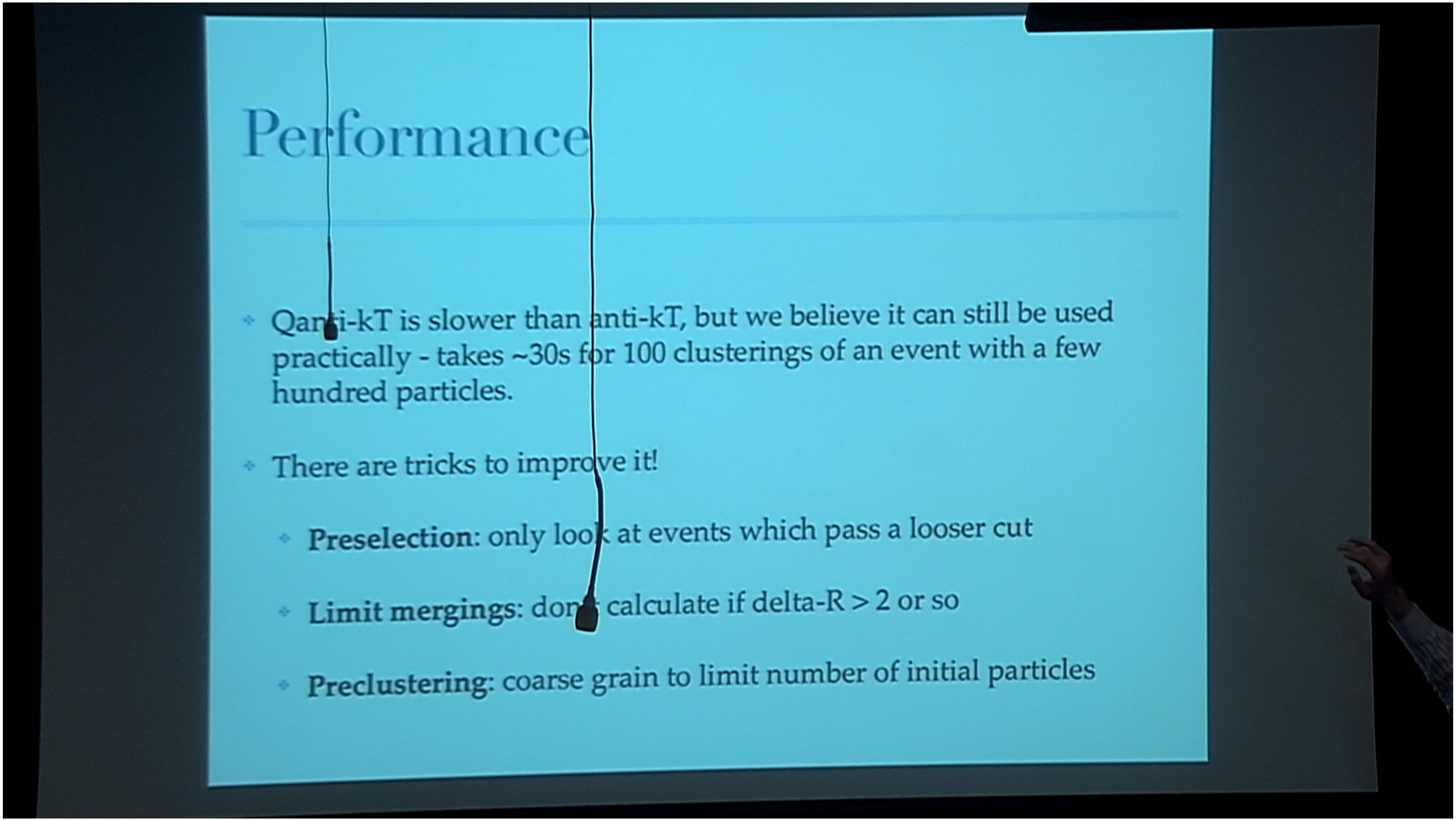
Significant Improvement in Stability

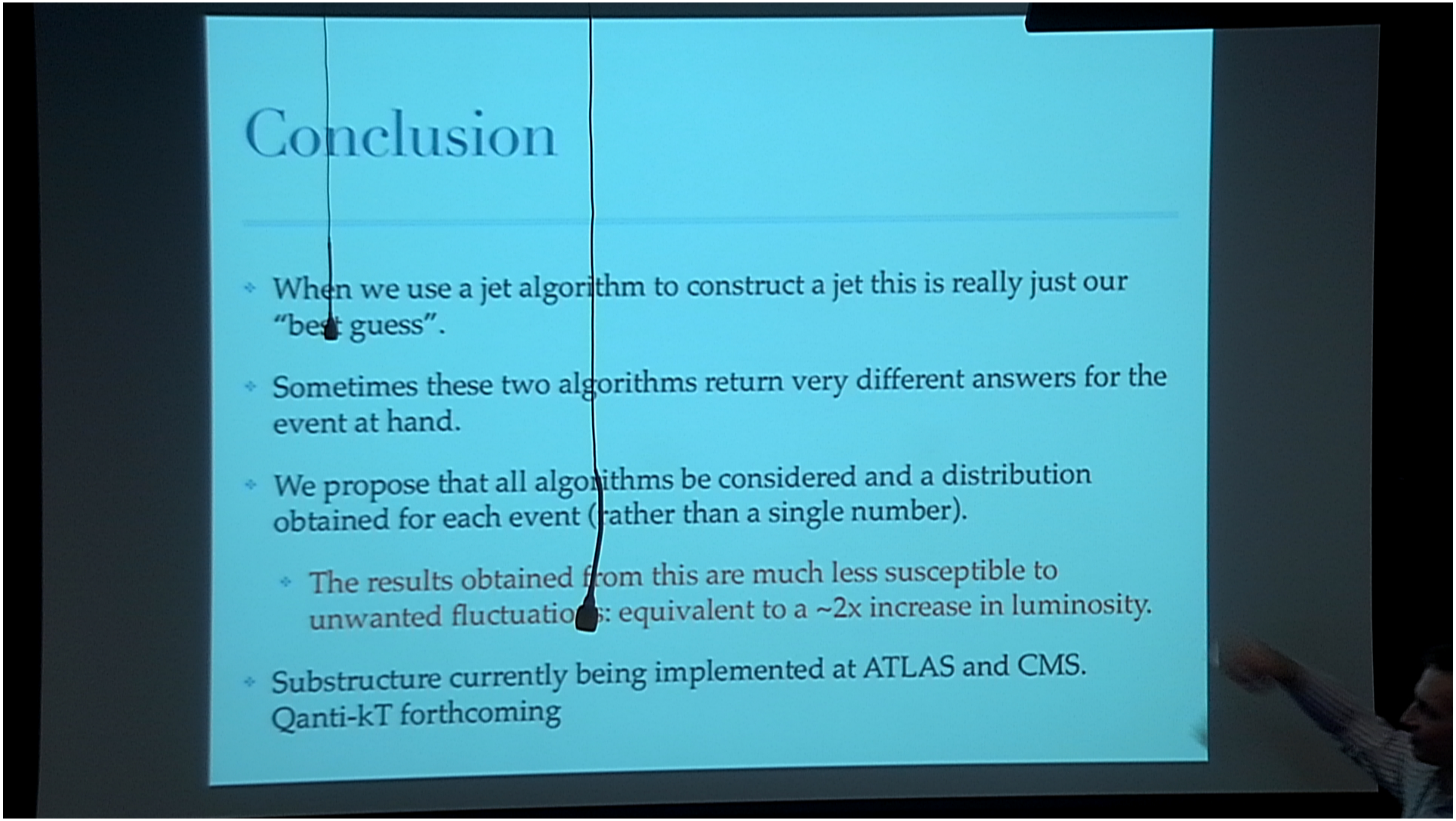
- *Retrofit onto any jet analysis.*
- Can make discoveries/exclusions much sooner!
- S/delta(B) is much larger than with traditional anti-kT.
  - Pair production of resonances: **49**% improvement in S/delta(B)
  - Resonance + Z: **19**% improvement on S/delta(B)
    - Relevant for the Higgs
  - Gets better as multiplicity increases

More ambiguity -> more improvement

- Simple dijets are clearcut - most interpretations are the same
- Scalar + Z: when p_T cut many events close to threshold -> more ambiguity
- Two resonances: lots of ambiguity, overlap

Sample	R	Improvement in $S/\delta B$ (%)				
		$\alpha = 0$	$\alpha = 0.01$	$\alpha = 0.1$	$\alpha = 1$	$\alpha = 100$
$pp \to \phi$	1.10	-86	-2	-9	**3**	-5
$pp \to \phi + Z$ (A)	0.95	-36	-1	**7**	9	1
$pp \to \phi + Z$ (B)	0.65	-42	-2	**19**	10	1
$pp \to \phi + \phi$	0.75	-25	43	**49**	40	1
$pp \to \phi + h$	0.7	-21	-1	**18**	**18**	1


Performance
Qanti-kT is slower than anti-kT, but we believe it can still be used practically - takes ~30s for 100 clusterings of an event with a few hundred particles.
There are tricks to improve it!
Preselection: only look at events which pass a looser cut
Limit mergings: don't calculate if delta-R > 2 or so
Preclustering: coarse grain to limit number of initial particles


Conclusion
When we use a jet algorithm to construct a jet this is really just our "best guess".
Sometimes these two algorithms return very different answers for the event at hand.
We propose that all algorithms be considered and a distribution obtained for each event (rather than a single number).
The results obtained from this are much less susceptible to unwanted fluctuations: equivalent to a ~2x increase in luminosity.
Substructure currently being implemented at ATLAS and CMS. Qanti-kT forthcoming