Dipolarity: Top-Tagging with Color Flow

Martin Jankowiak

with Anson Hook and Jay Wacker

arXiv:hep-ph/1102.1012

SLAC March 9th, 2011

Outline

- Jet substructure
- Sequential jet clustering algorithms
- Some jet substructure techniques
- The HEPTopTagger
- Color flow and pull
- Dipolarity
- Results
- Summary & Outlook

Jet substructure at the LHC

- the excellent resolution of the ATLAS & CMS detectors means that we can “peer inside” jets and measure how energy is distributed within jets

Jet substructure at the LHC

- the excellent resolution of the ATLAS & CMS detectors means that we can "peer inside" jets and measure how energy is distributed within jets

What is this good for?

Jet substructure at the LHC

- the excellent resolution of the ATLAS & CMS detectors means that we can “peer inside” jets and measure how energy is distributed within jets

What is this good for?

- as a probe of QCD

Jet substructure at the LHC

- the excellent resolution of the ATLAS & CMS detectors means that we can “peer inside” jets and measure how energy is distributed within jets

What is this good for?

- as a probe of QCD
- event discrimination

Jet substructure as a probe of QCD

- make jet substructure measurements in real data and compare to perturbative QCD calculations

Jet substructure as a probe of QCD

- make jet substructure measurements in real data and compare to perturbative QCD calculations
- use to tune Monte Carlo event generators

Jet substructure for event discrimination

- the LHC inverse problem:

  how do we connect what we measure (jets) to the hard scattering ?

- use the characteristic energy distribution of signal jets (e.g. top jets) to discriminate against background jets (e.g. QCD jets initiated by light partons)

- especially relevant for boosted objects

Sequential jet clustering algorithms

- using an iterative procedure, combine four-vectors of particles to yield a list of jet four-vectors
- procedure is formulated in coordinates with simple properties under longitudinal boosts (the rapidity y and the azimuthal angle ϕ)

$$y \equiv \frac{1}{2} \ln \frac{E+p_z}{E-p_z}$$

- use the euclidean distance ΔR_{ij} in the $y-\phi$ plane

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

Sequential jet clustering algorithms

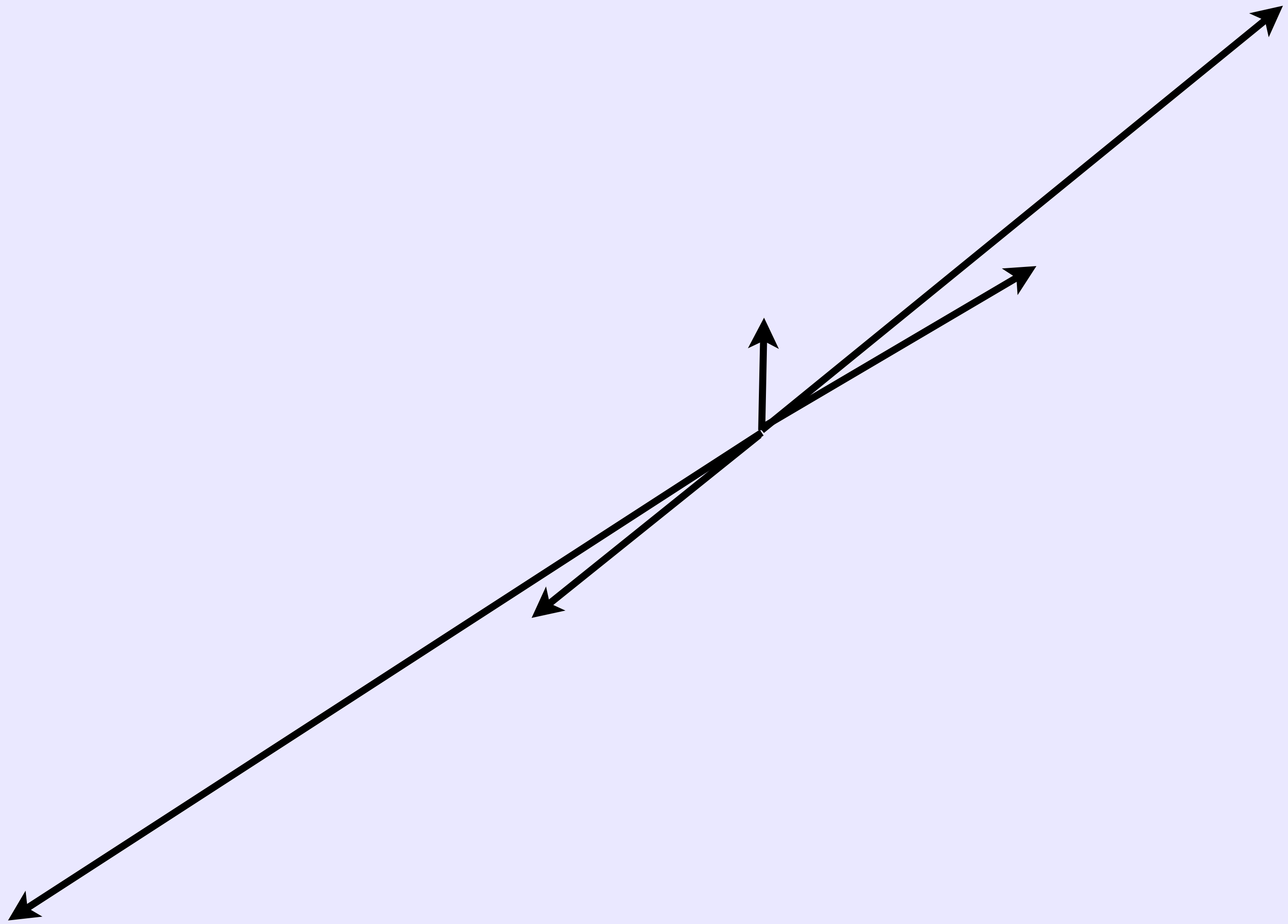
Cambridge-Aachen Algorithm

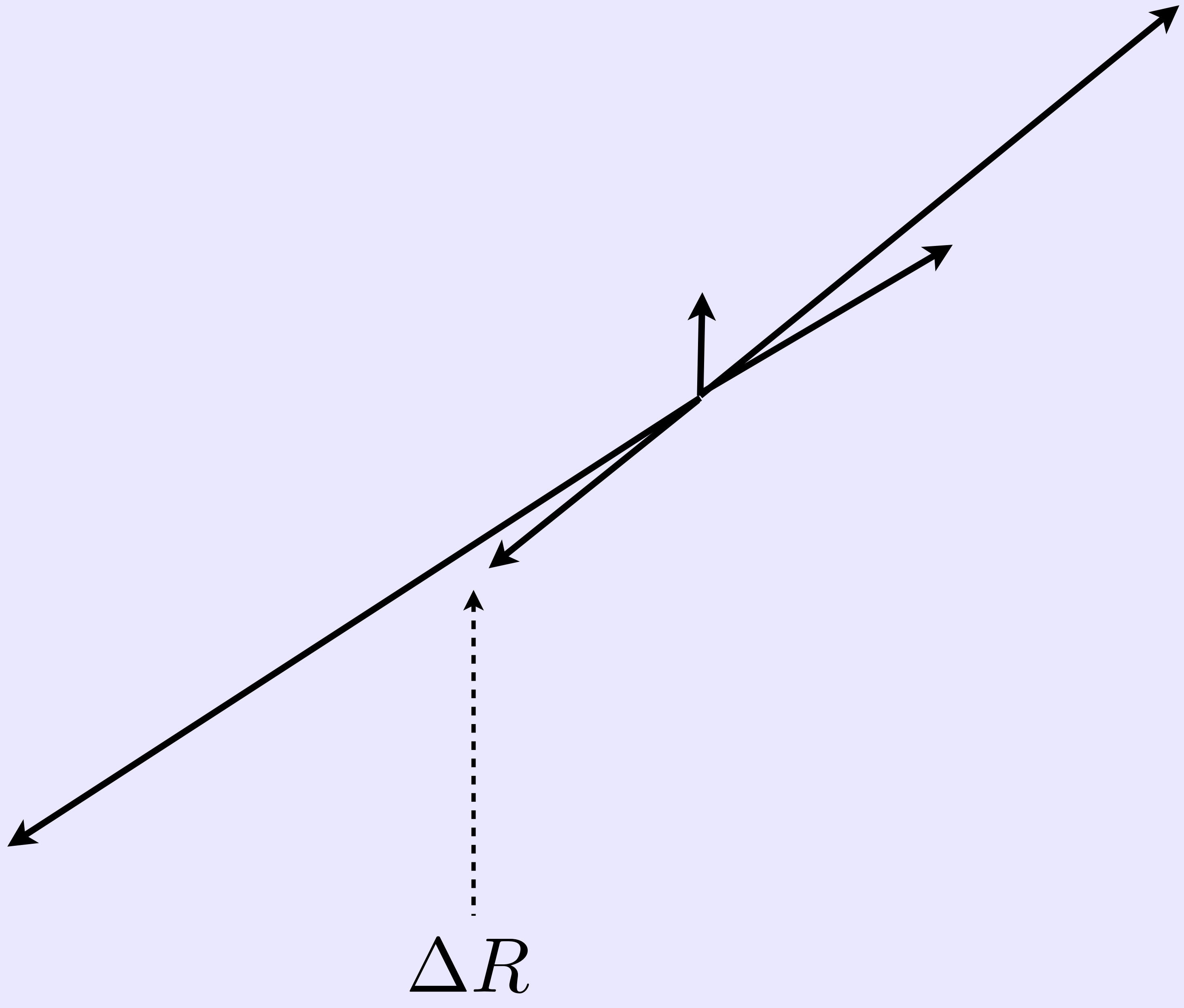
1. find the smallest of the ΔR_{ij}

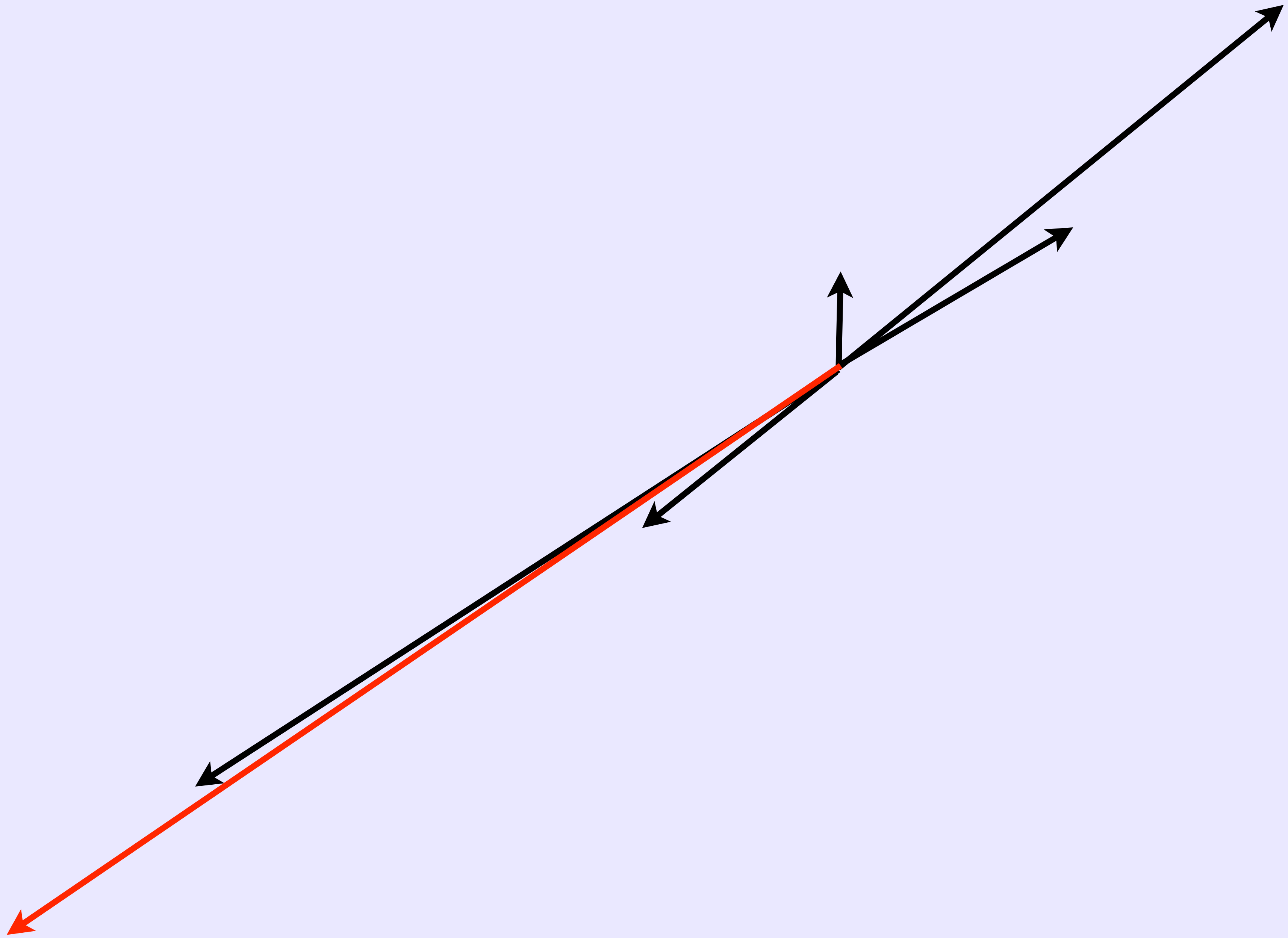
2. combine i and j and return to step 1

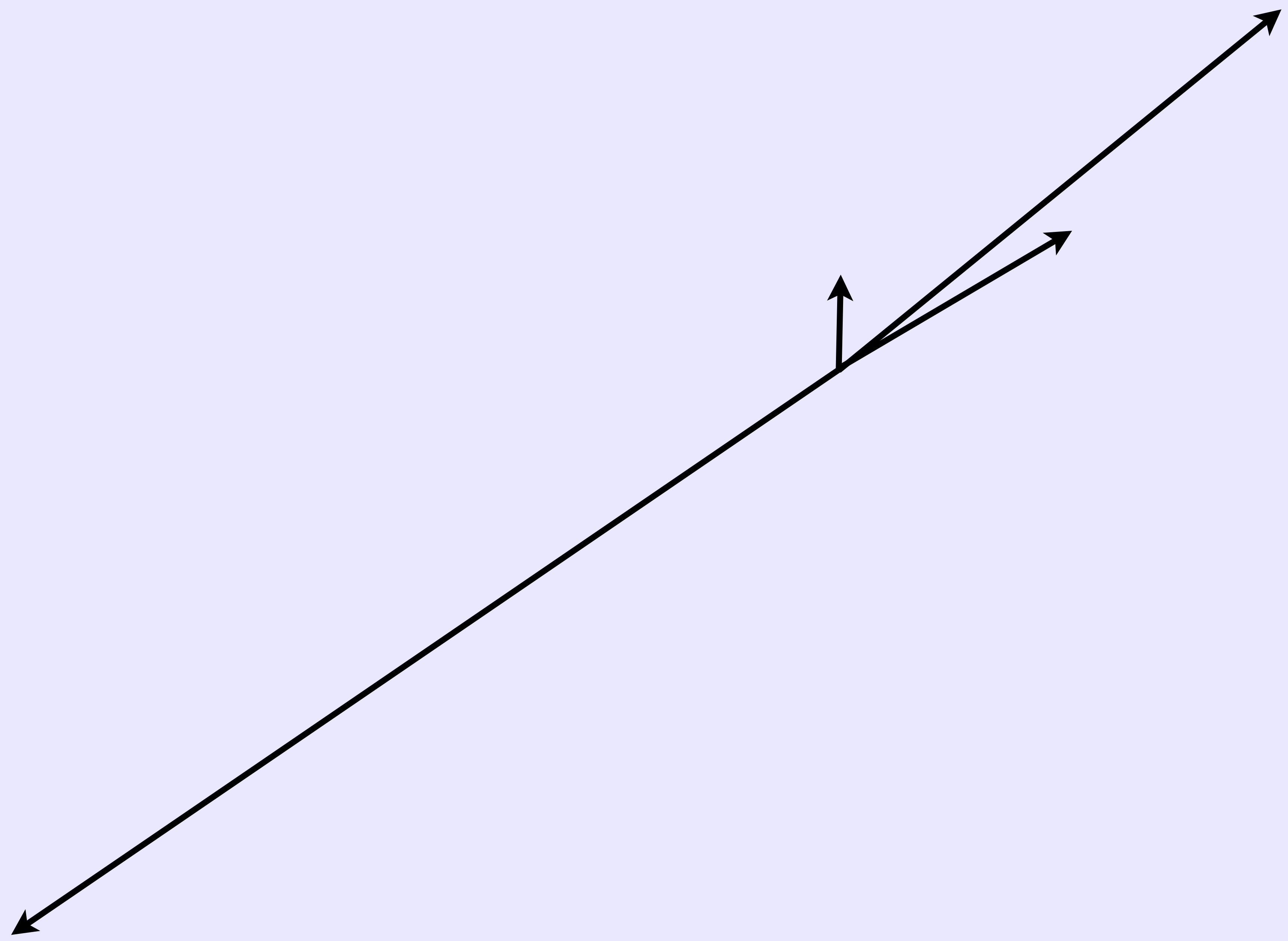
3. continue until all $\Delta R_{ij} > R$

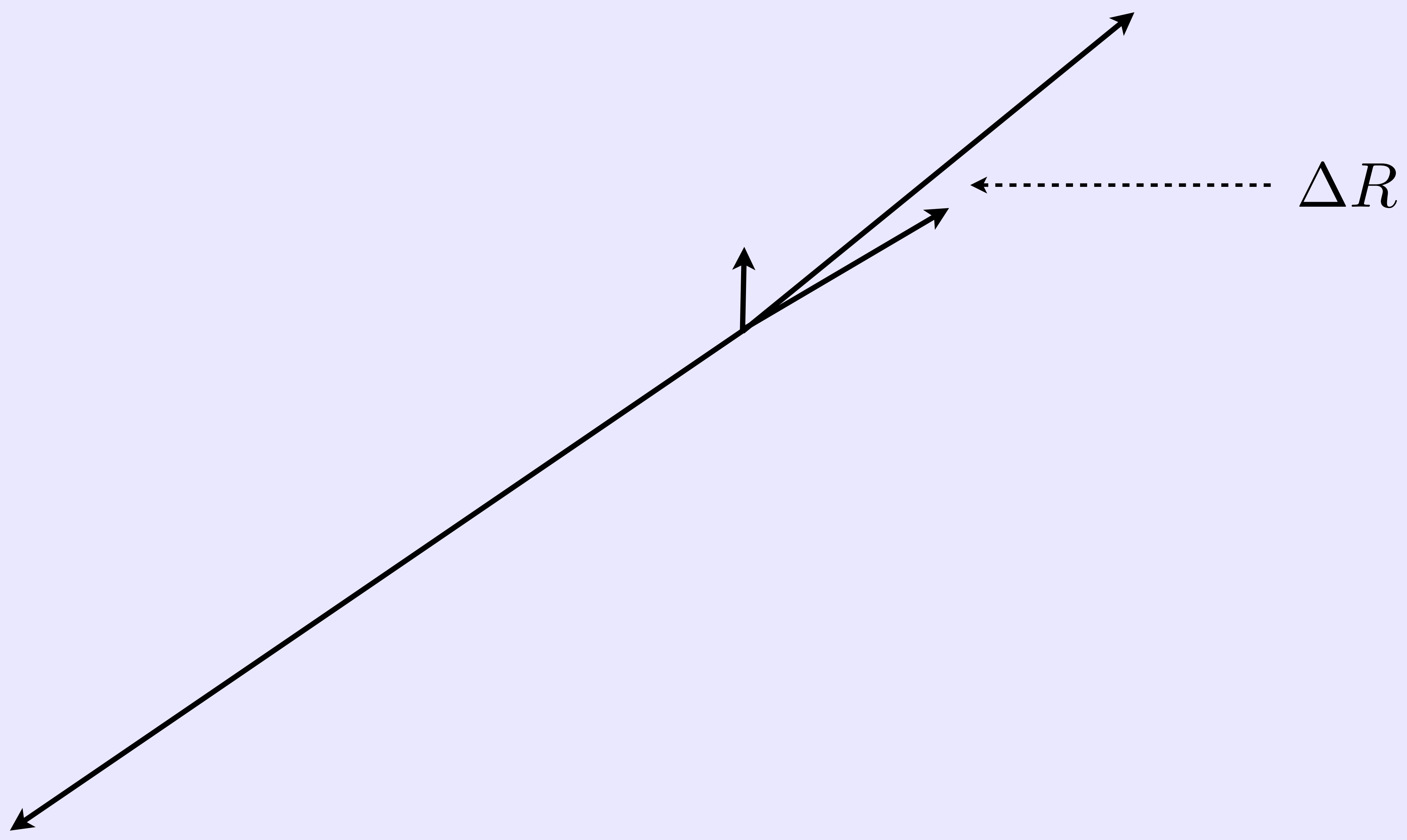
4. the remaining four-vectors define a list of jets

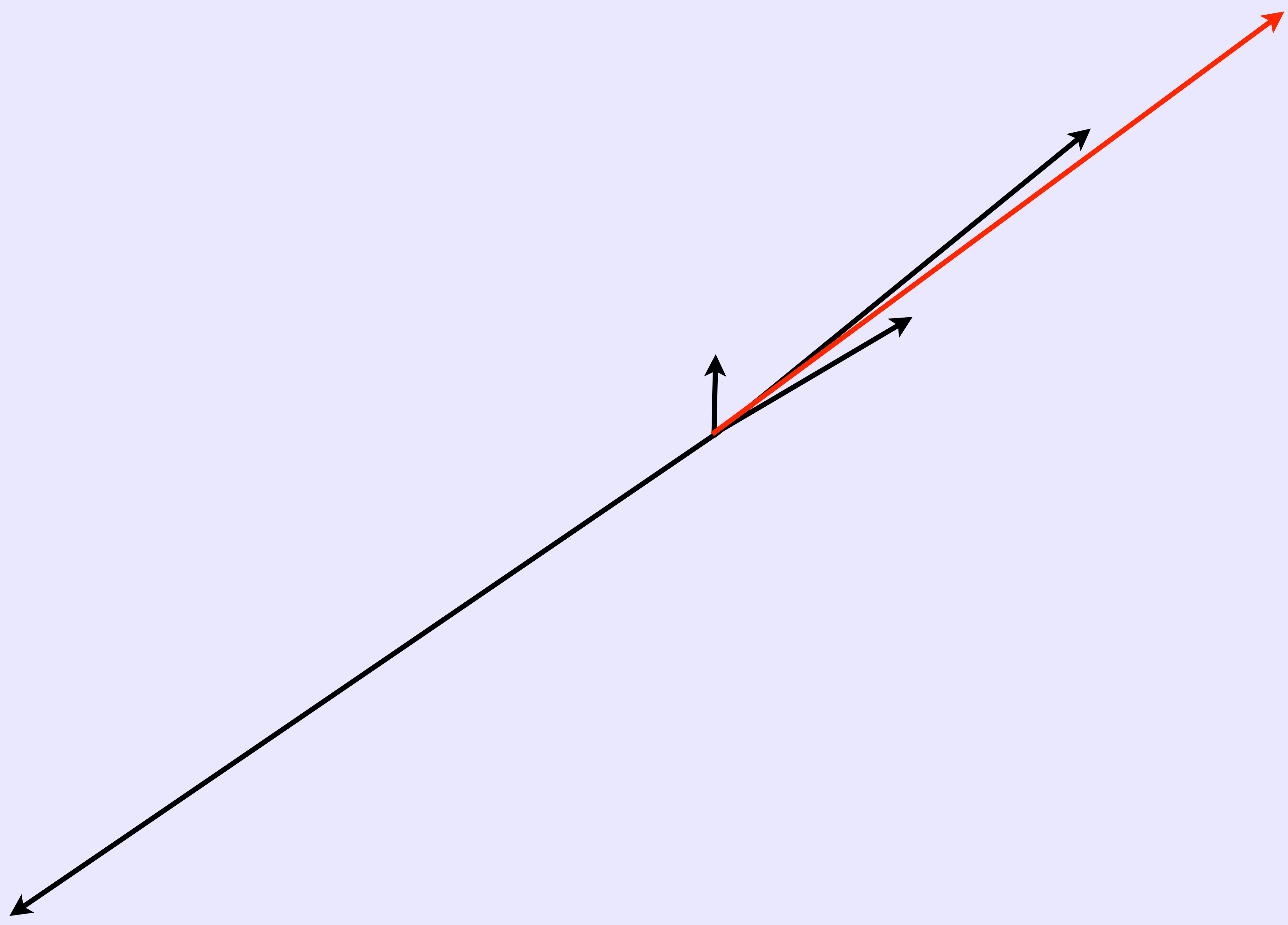



ΔR



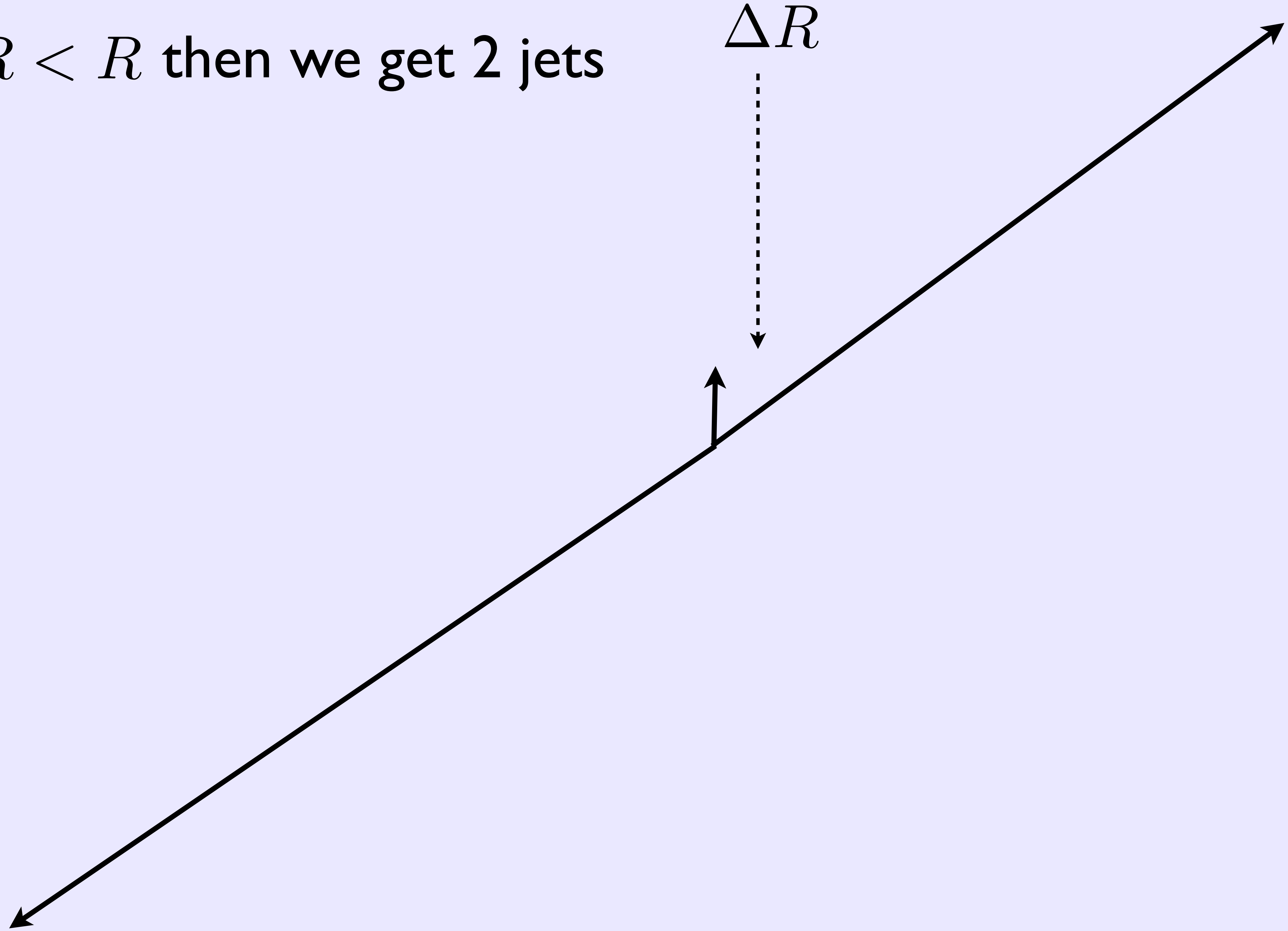



ΔR



If $\Delta R > R$ then we get 3 jets

If $\Delta R < R$ then we get 2 jets



Some jet substructure techniques

Boosted particles

- at the LHC many of the particles considered ‘heavy’ at previous colliders will be produced with transverse momenta far exceeding their rest masses ($W^{\pm}, Z^0, t, h$)

- in many Beyond the Standard Model scenarios boosted particles appear in the decay of heavy resonances (e.g. $\phi \to t\bar{t}$)

Some jet substructure techniques

Boosted higgs search

- for $p_T \gtrsim m_H$ the decay products of the higgs will typically be close together and reconstructed as a single jet
- about 5% of the cross-section for VH has $p_T > 200$ GeV
- backgrounds (V+jets, VV, top pairs) fall faster with p_T than the signal
- can pay to go to the boosted regime if substructure techniques can reduce backgrounds

Jon Butterworth, Adam Davison, Mathieu Rubin, Gavin Salam arXiv/hep-ph:0802.2470

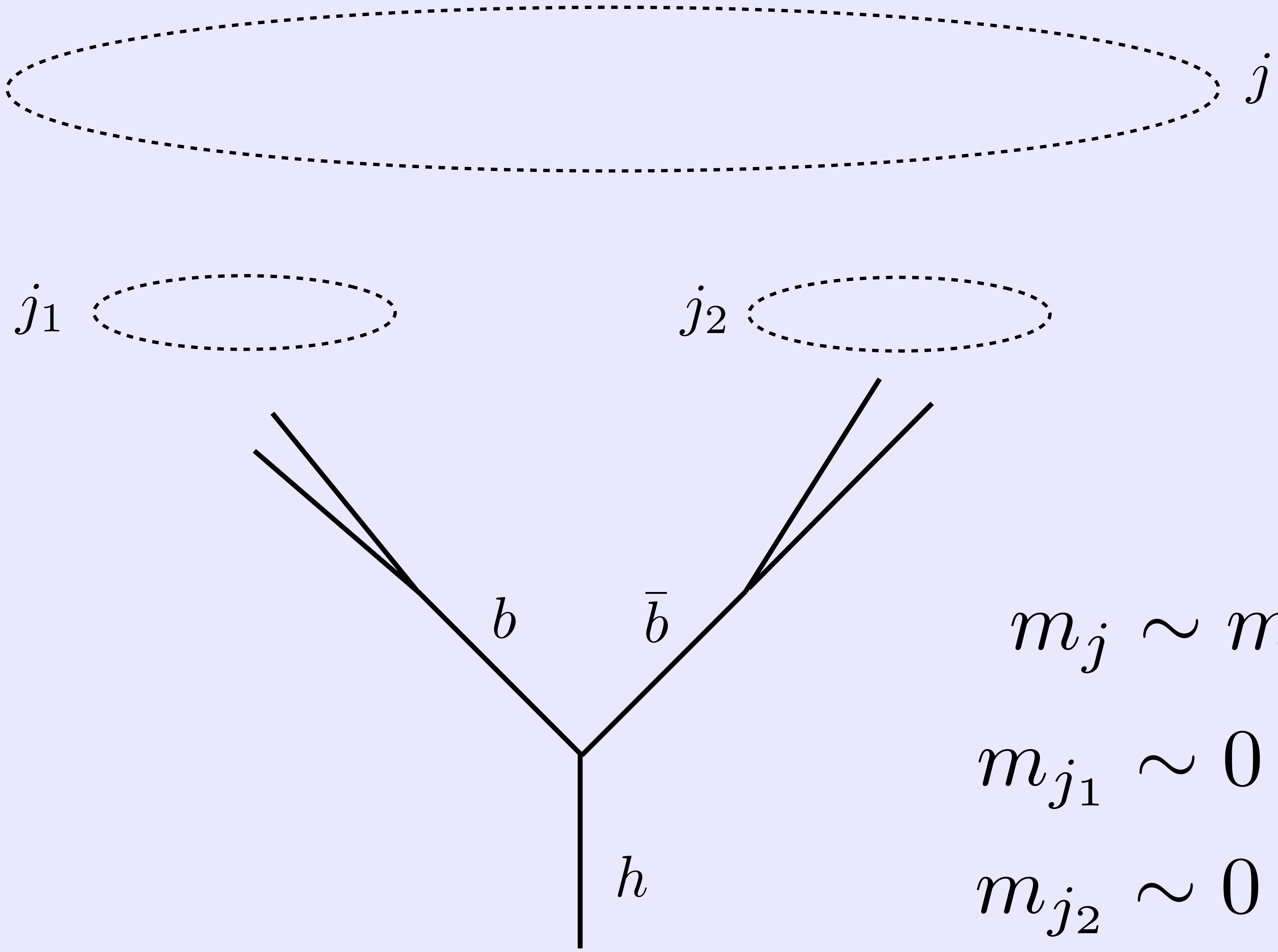
Some jet substructure techniques

- to capture all of the higgs decay products in a single jet, we need to use “fat” jets

- to accurately reconstruct the mass of the higgs, we want to “clean up” our jet to get rid of contamination from the underlying event and pile-up

- use jet substructure techniques to identify the heavy particle neighborhood of the jet

Some jet substructure techniques



$$m_j \sim m_h$$

$$m_{j_1} \sim 0$$

$$m_{j_2} \sim 0$$

Some jet substructure techniques

1. Break a fat (R=1.5) C/A jet j into subjets j_1 and j_2 by undoing the last stage of clustering; label so that $m_{j_1} > m_{j_2}$

Some jet substructure techniques

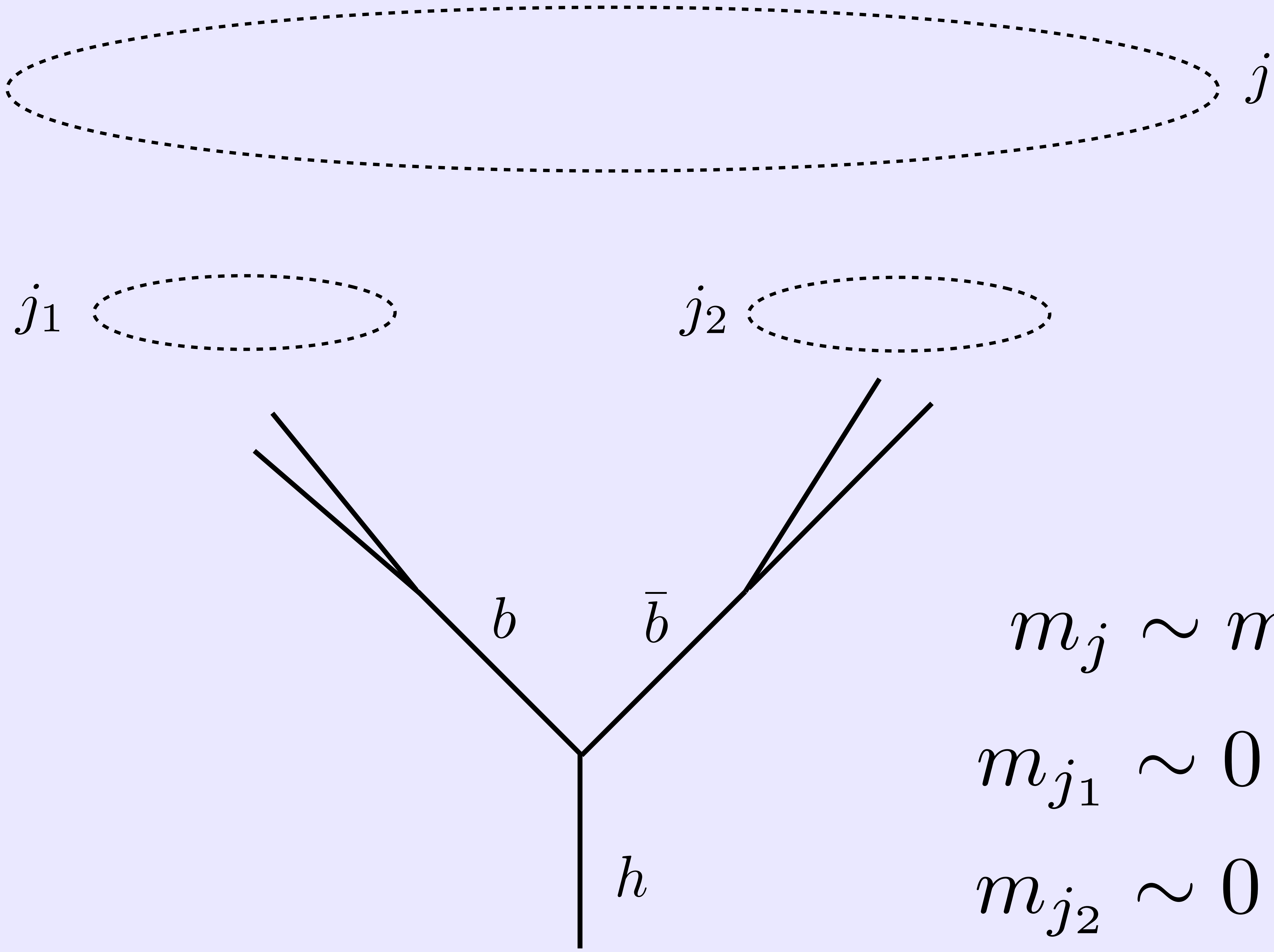
1. Break a fat (R=1.5) C/A jet j into subjets j_1 and j_2 by undoing the last stage of clustering; label so that $m_{j_1} > m_{j_2}$

2. If there is a significant mass drop $m_{j_1} < \mu m_j$ then exit the loop

3. Otherwise redefine j_1 as j and go back to step 1

Here $\mu = 0.67$

Some jet substructure techniques

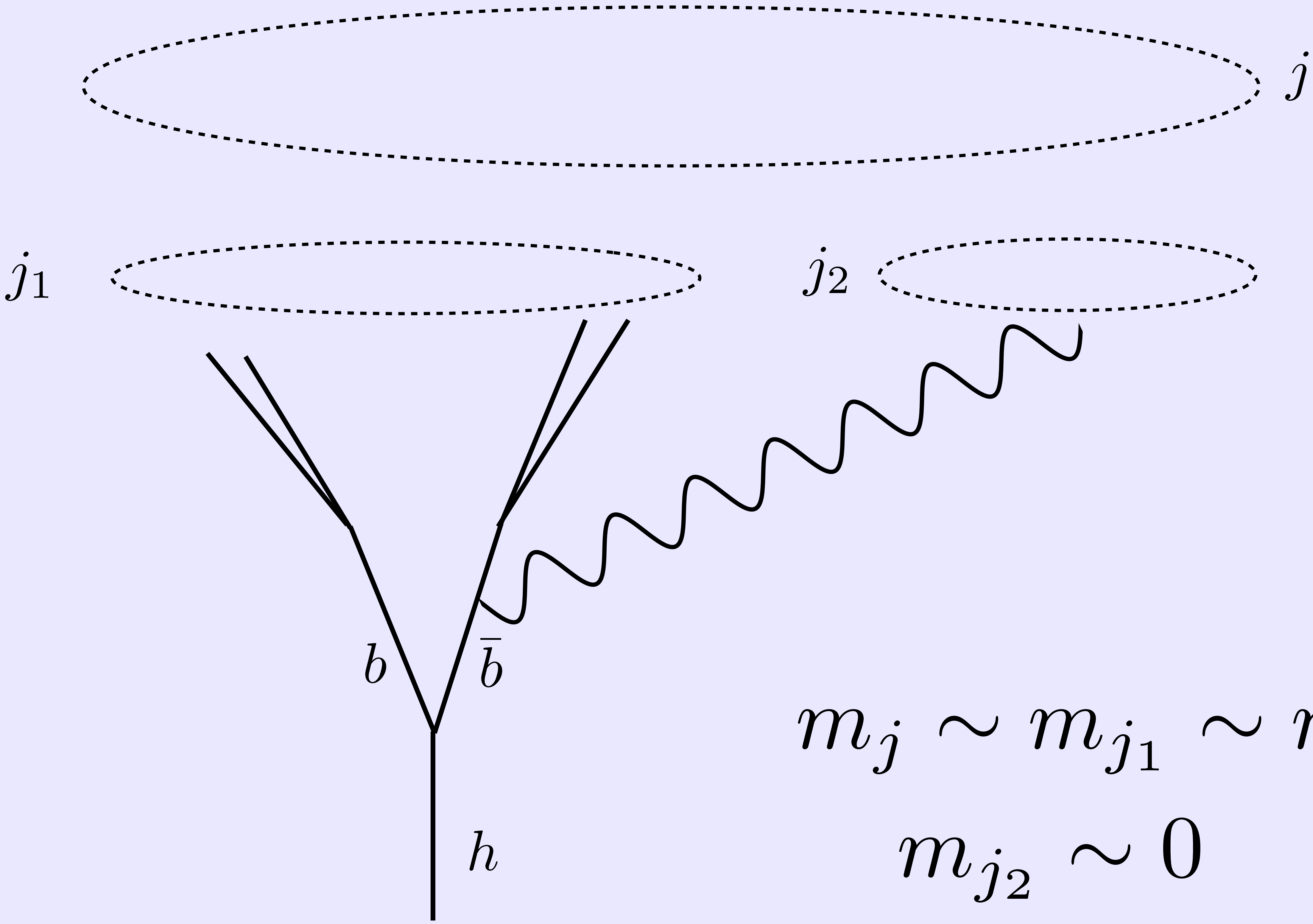


$$m_j \sim m_h$$

$$m_{j_1} \sim 0$$

$$m_{j_2} \sim 0$$

Some jet substructure techniques



$$m_j \sim m_{j_1} \sim m_h$$

$$m_{j_2} \sim 0$$

Some jet substructure techniques

1. Break a fat (R=1.5) C/A jet j into subjets j_1 and j_2 by undoing the last stage of clustering; label so that $m_{j_1} > m_{j_2}$

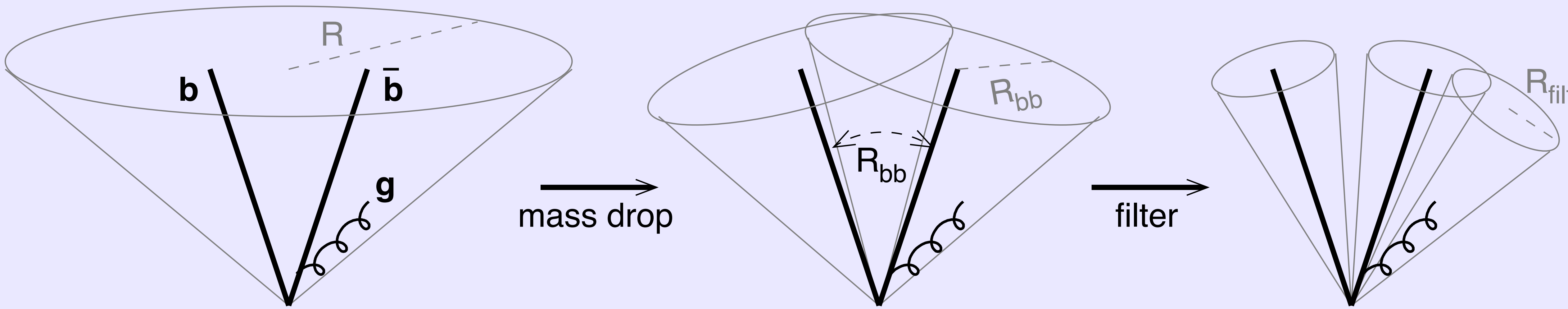
2. If there is a significant mass drop $m_{j_1} < \mu m_j$ then exit the loop

3. Otherwise redefine j_1 as j and go back to step 1

Here $\mu = 0.67$

4. Finally, recluster with $R_{\text{filt}} = \min(0.3, \frac{R_{b\bar{b}}}{2})$ and use the three hardest subjets to calculate the filtered Higgs mass

Some jet substructure techniques



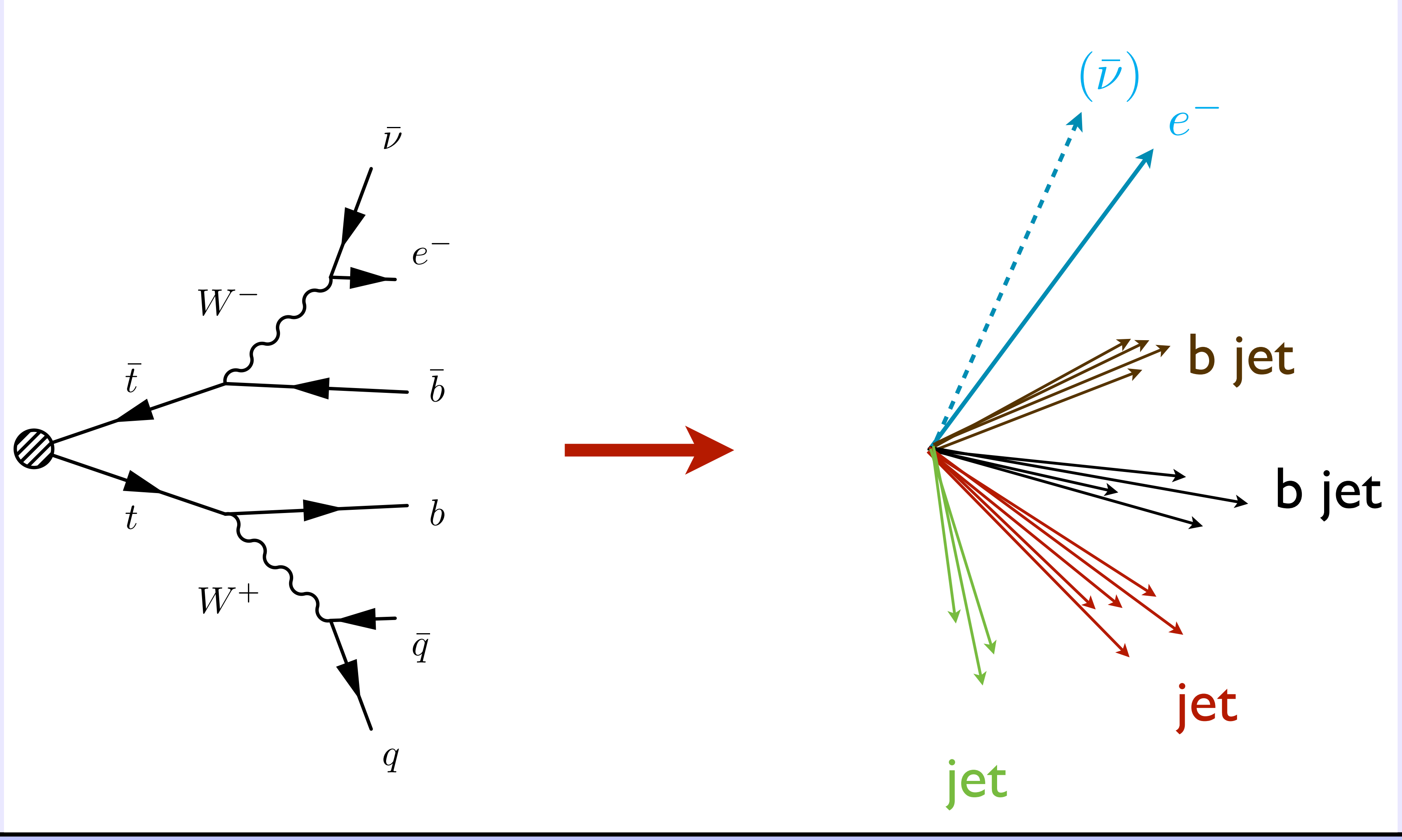
the "BDRS" procedure in diagrams

The HEPTopTagger

HEP = Heidelberg/Eugene/Paris

Tilman Plehn, Gavin Salam, Michael Spannowsky, Michihisa Takeuchi, Dirk Zerwas
arXiv/hep-ph:1006.2833 arXiv/hep-ph:0910.5472

- essentially a generalization of the BDRS procedure to identify the three-pronged hard substructure of a top jet
- designed for intermediate boost $200\,\mathrm{GeV} \lesssim p_T \lesssim 800\,\mathrm{GeV}$



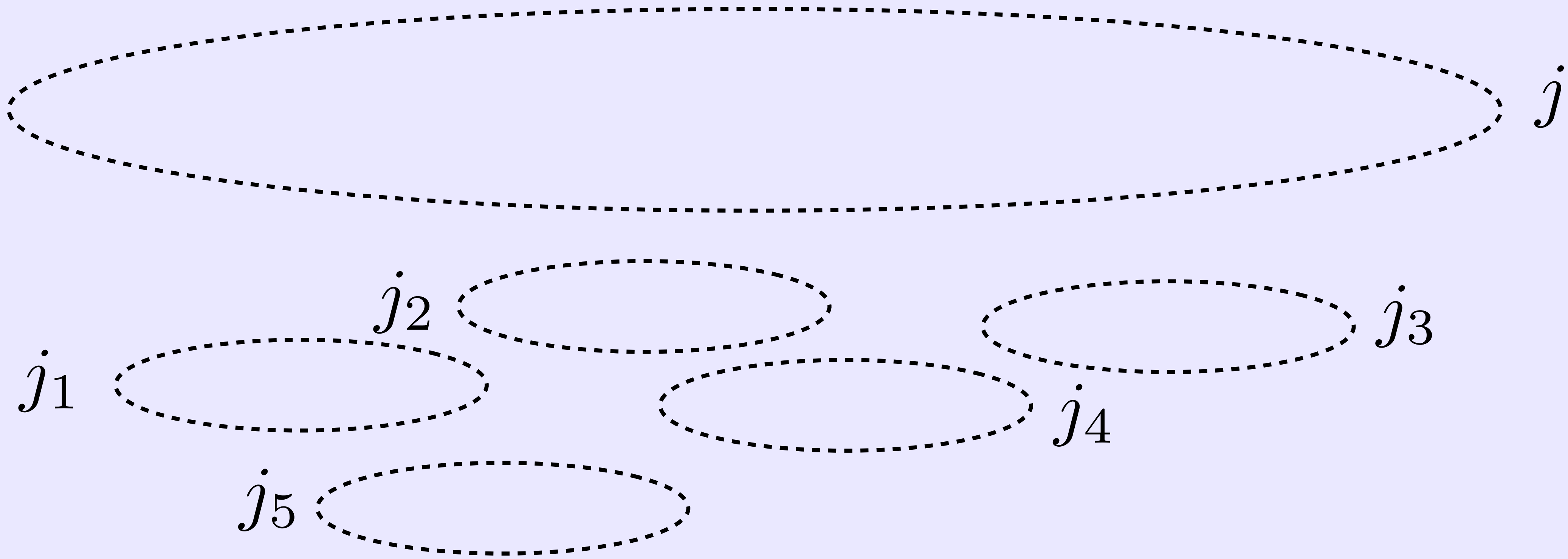
The HEPTopTagger

1. Using the Cambridge/Aachen algorithm cluster the event into fat $R = 1.5$ jets.

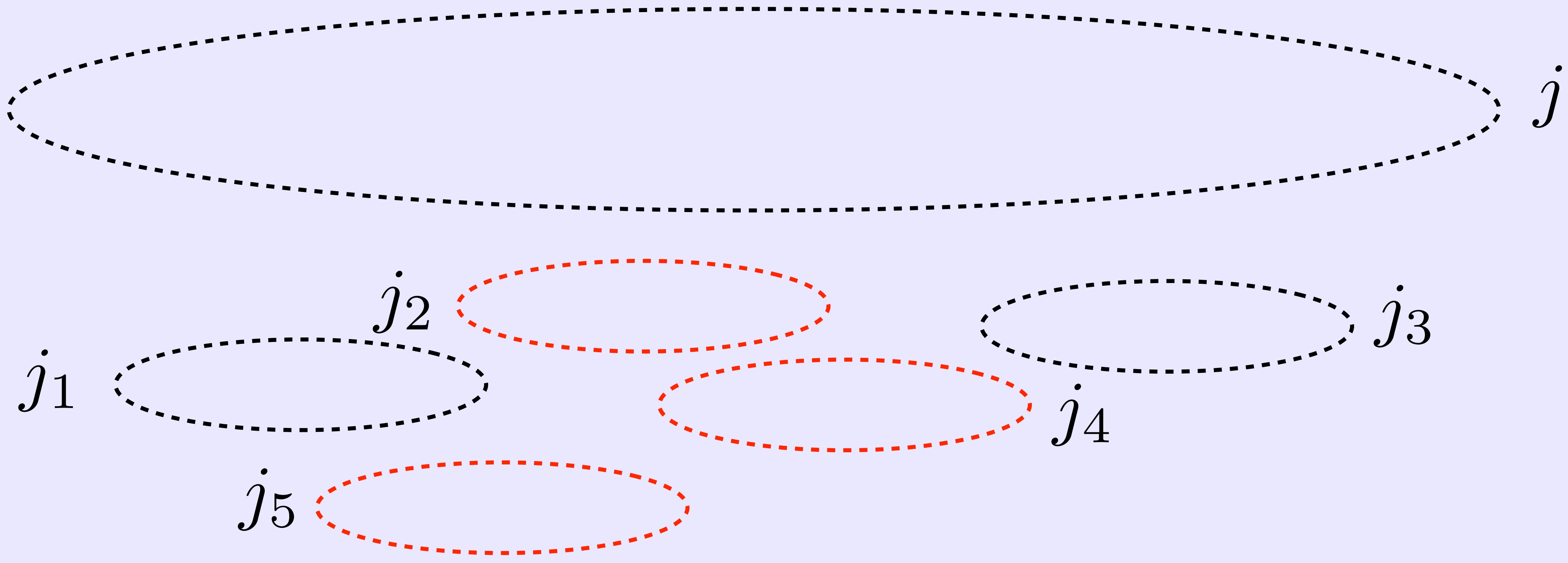
2. Break each fat jet into hard subjets using the following mass-drop criterion. Undo the last stage of clustering to yield two subjets j_1 and j_2 (with $m_{j_1} > m_{j_2}$), keeping both j_1 and j_2 if $m_{j_1} < 0.8 m_j$ and otherwise dropping j_2. Repeat this procedure recursively, stopping when the m_{j_i} drop below 30 GeV.

3. Consider in turn all possible triplets of hard subjets. First, filter each triplet with a resolution $R_{\text{filter}} = \min(0.3, \Delta R_{ij}/2)$. Next, using the five hardest constituent subjets of the filtered triplet calculate the jet mass m_{filt}. Finally, choose the triplet whose m_{filt} lies closest to m_t.

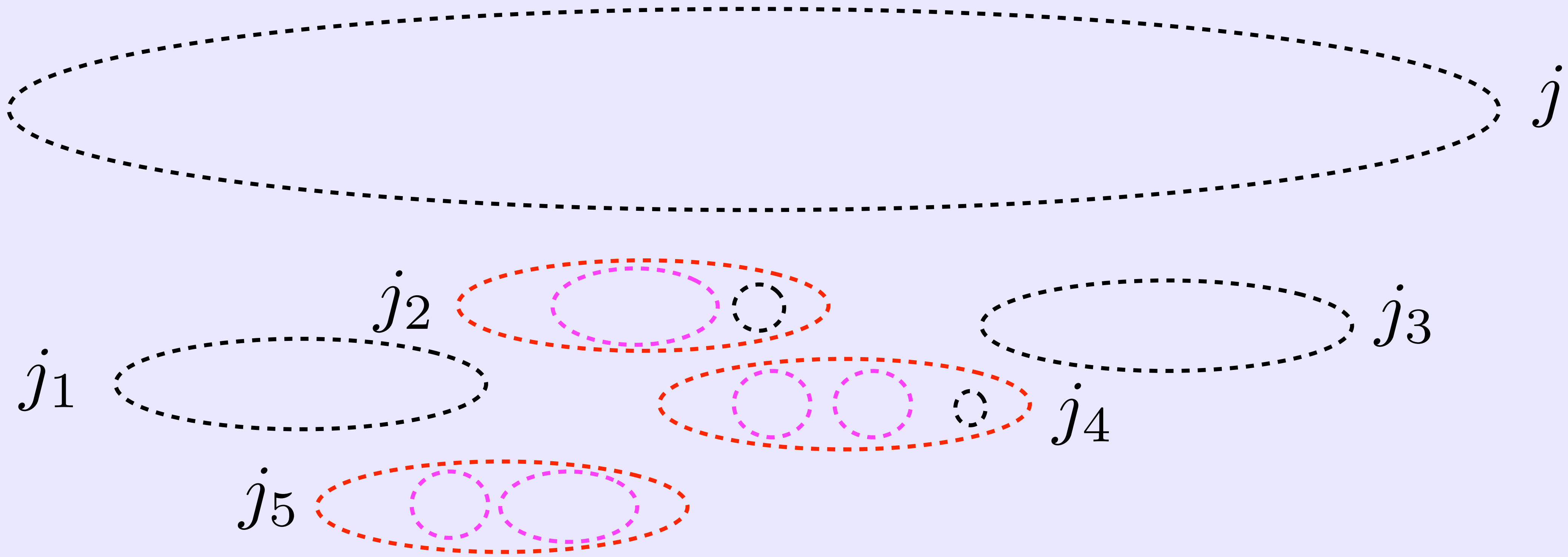
The HEPTopTagger



The HEPTopTagger



The HEPTopTagger



$$\Rightarrow m_t$$

4. Recluster the five filtered constituents chosen in step 3 into exactly three subjets j_1, j_2, and j_3 ordered in descending p_T. Accept the fat jet as a top candidate if it passes any of the following three pairs of mass cuts:

$$i)\quad 0.2 \leq \arctan m_{13} \leq 1.3$$

$$i')\quad R_{\min} \leq \frac{m_{23}}{m_{123}} \leq R_{\max}$$

$$ii)\quad R^2_{\min}\left(1+\frac{m^2_{13}}{m^2_{12}}\right) \leq 1-\frac{m^2_{23}}{m^2_{123}} \leq R^2_{\max}\left(1+\frac{m^2_{13}}{m^2_{12}}\right)$$

$$ii')\quad \frac{m_{23}}{m_{123}} \geq 0.35$$

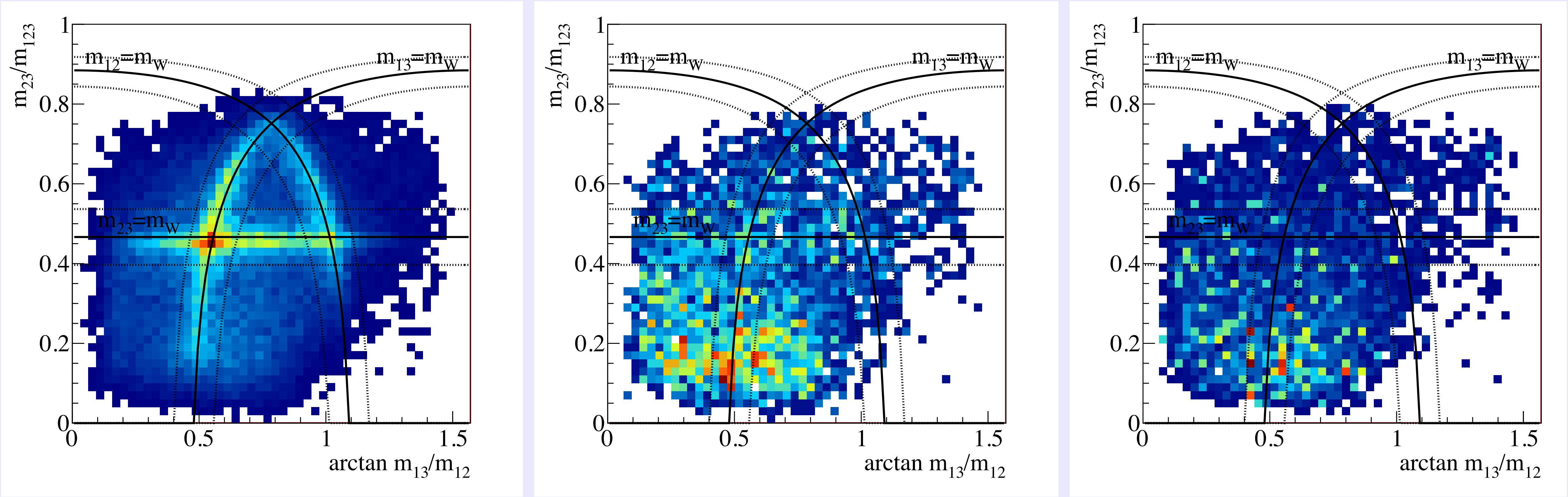
$$iii)\quad R^2_{\min}\left(1+\frac{m^2_{12}}{m^2_{13}}\right) \leq 1-\frac{m^2_{23}}{m^2_{123}} \leq R^2_{\max}\left(1+\frac{m^2_{12}}{m^2_{13}}\right)$$

$$iii')\quad \frac{m_{23}}{m_{123}} \geq 0.35$$

Here $R_{\min} = 85\% \times m_W/m_t$ and $R_{\max} = 115\% \times m_W/m_t$.

5. Finally, require that the total p_T of the three subjets defined in step 4 be greater than 200 GeV.

The HEPTopTagger



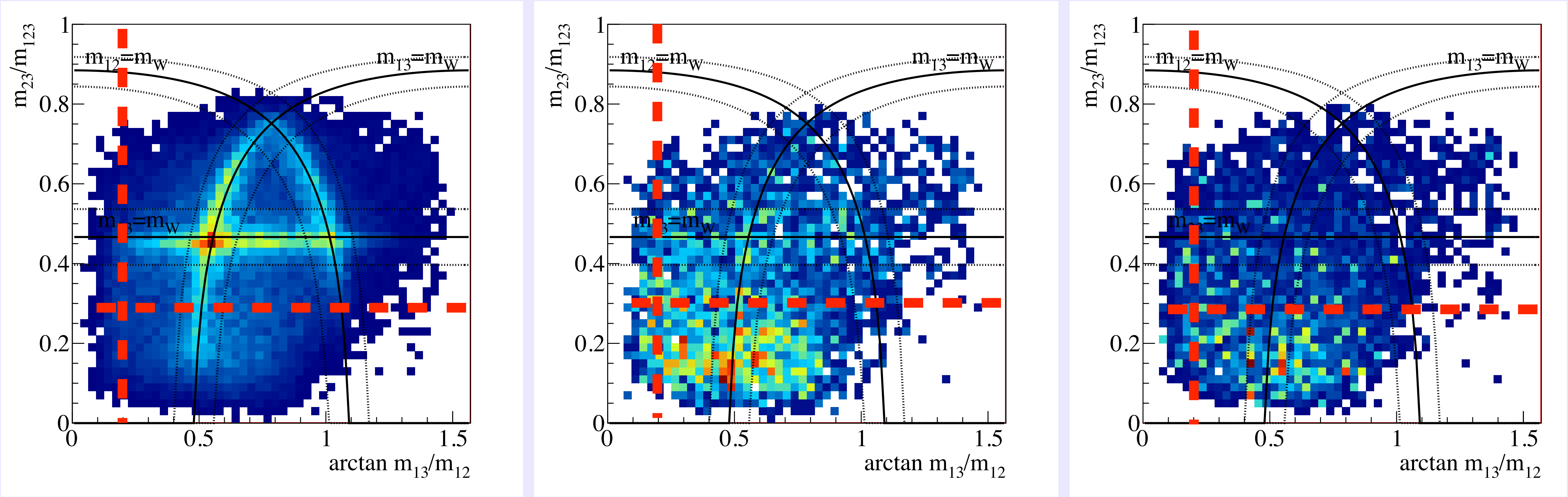
$$i) \quad 0.2 \leq \arctan m_{13} \leq 1.3$$

$$i') \quad R_{\min} \leq \frac{m_{23}}{m_{123}} \leq R_{\max}$$

$$iii) \quad R^2_{\min}\left(1+\frac{m^2_{12}}{m^2_{13}}\right) \leq 1-\frac{m^2_{23}}{m^2_{123}} \leq R^2_{\max}\left(1+\frac{m^2_{12}}{m^2_{13}}\right)$$

$$iii') \quad \frac{m_{23}}{m_{123}} \geq 0.35$$

The HEPTopTagger



$$i) \quad 0.2 \leq \text{arctan}\ m_{13} \leq 1.3$$

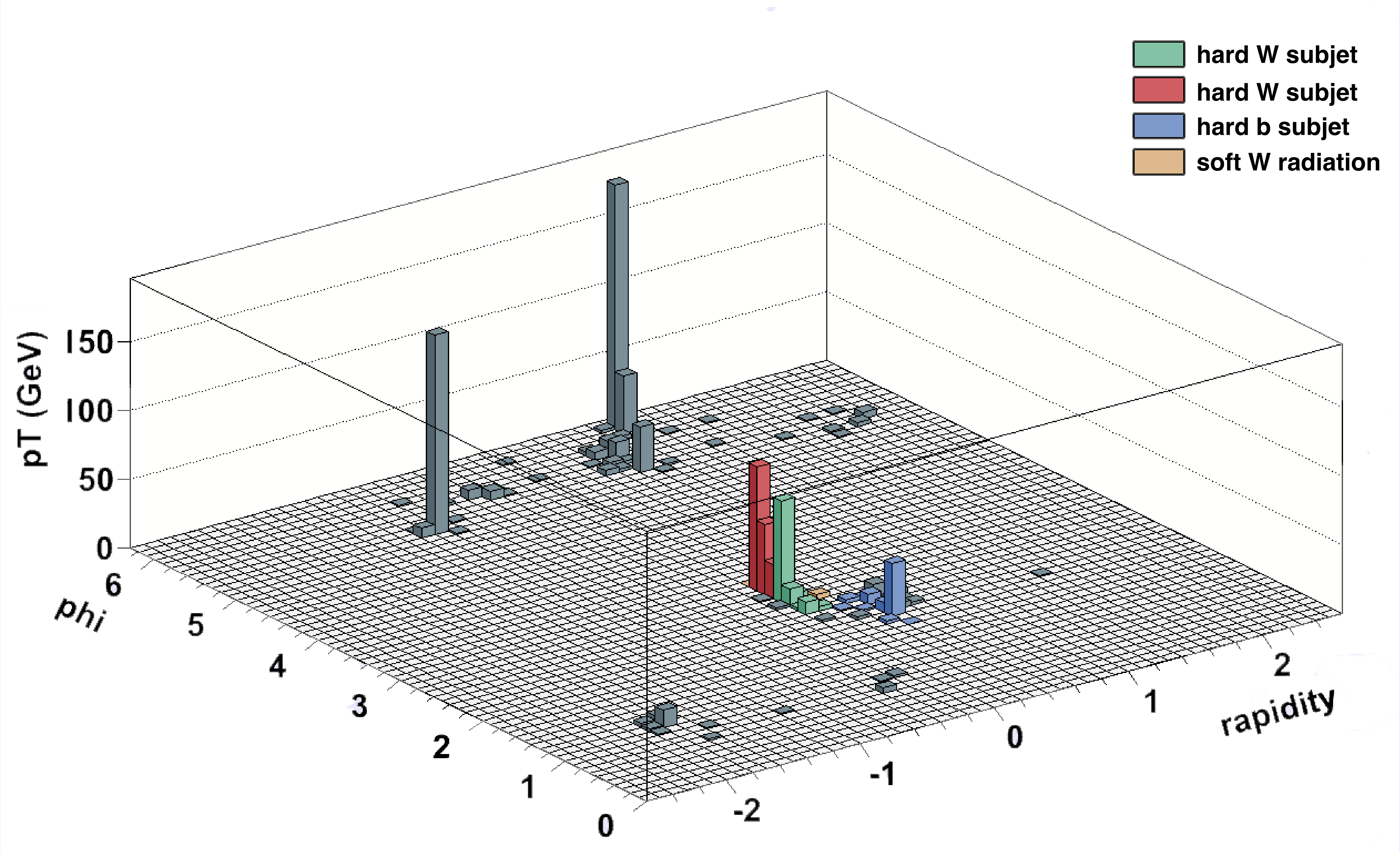
$$i') \quad R_{\min} \leq \frac{m_{23}}{m_{123}} \leq R_{\max}$$

$$iii) \quad R_{\min}^2 \left(1 + \frac{m_{12}^2}{m_{13}^2}\right) \leq 1 - \frac{m_{23}^2}{m_{123}^2} \leq R_{\max}^2 \left(1 + \frac{m_{12}^2}{m_{13}^2}\right)$$

$$iii') \quad \frac{m_{23}}{m_{123}} \geq 0.35$$

The HEPTopTagger

Legoplot for a top jet with hard substructure as identified by the HEPTopTagger



Color flow and pull

- a top jet has more structure than is encoded by kinematic constraints:

$$(p_1 + p_2 + p_3)^2 = m_t^2$$
$$(p_1 + p_2)^2 = m_W^2$$

- the effective operator for the decay $t \to bq\bar{q}$ has a particular color configuration

- in particular the W boson is a color singlet and the color indices of q and $\bar{q}$ are contracted

Color flow and pull

- a top jet has more structure than is encoded by kinematic constraints:

$$(p_1 + p_2 + p_3)^2 = m_t^2$$
$$(p_1 + p_2)^2 = m_W^2$$

- the effective operator for the decay $t \to bq\bar{q}$ has a particular color configuration

- in particular the W boson is a color singlet and the color indices of q and $\bar{q}$ are contracted

Question: can we use color information to improve background rejection in top tagging algorithms?

Color flow and pull

- in a QCD event radiation is controlled by
  i) the kinematics of the hard partons and by
  ii) how color indices are contracted together (color flow)

- how does a color singlet radiate?

- apart from some color algebra, QED ~ QCD;
  so let's first ask this question in the context of QED

Color flow and pull

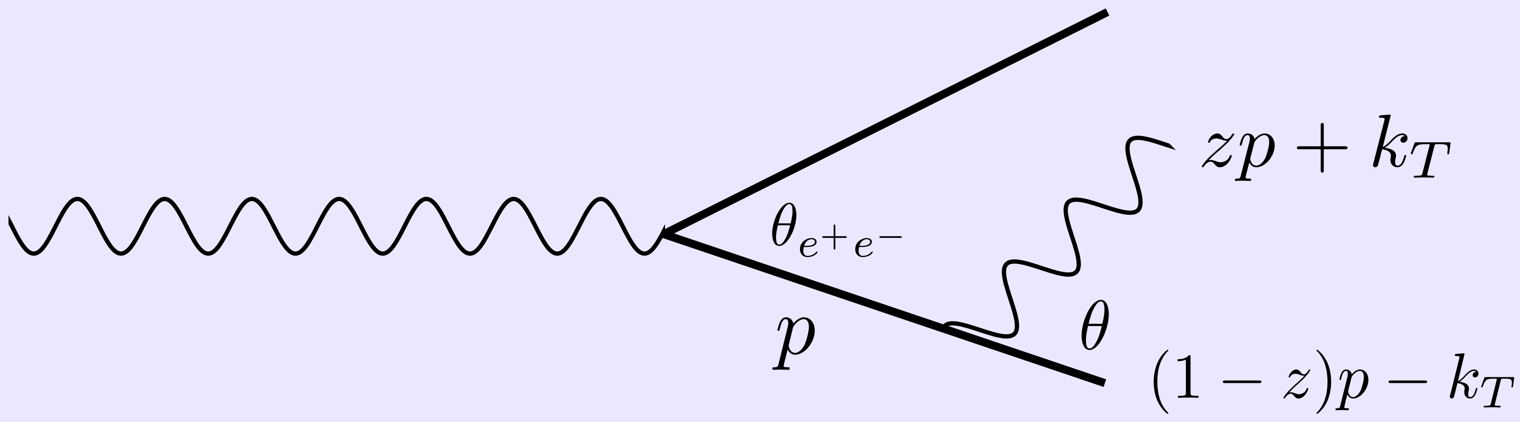
The Chudakov Effect:

Soft bremsstrahlung from e^+e^- pairs is suppressed

Heuristic explanation:

- there is an energy imbalance at the vertex $\Delta E \sim k_T^2/zp \sim zp\theta^2$
- time available for emission is $\Delta t \sim 1/\Delta E$
- in which time the pair separates $\Delta b \sim \theta_{e^+e^-}\Delta t$
- for emission photon must resolve this distance:

$$\Delta b > \lambda/\theta \sim (zp\theta)^{-1} \Rightarrow \theta_{e^+e^-}(zp\theta^2)^{-1} > (zp\theta)^{-1} \Rightarrow \theta_{e^+e^-} > \theta$$



Color flow and pull

The Chudakov Effect:

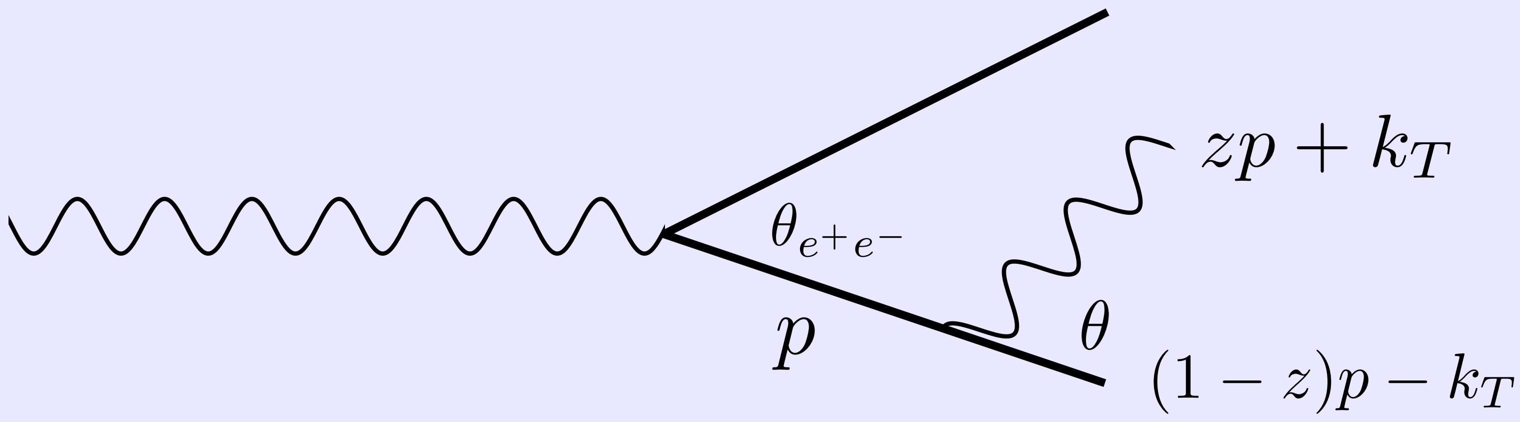
Soft bremsstrahlung from e^+e^- pairs is suppressed

Heuristic explanation:

- there is an energy imbalance at the vertex $\Delta E \sim k_T^2/zp \sim zp\theta^2$
- time available for emission is $\Delta t \sim 1/\Delta E$
- in which time the pair separates $\Delta b \sim \theta_{e^+e^-}\Delta t$
- for emission photon must resolve this distance:

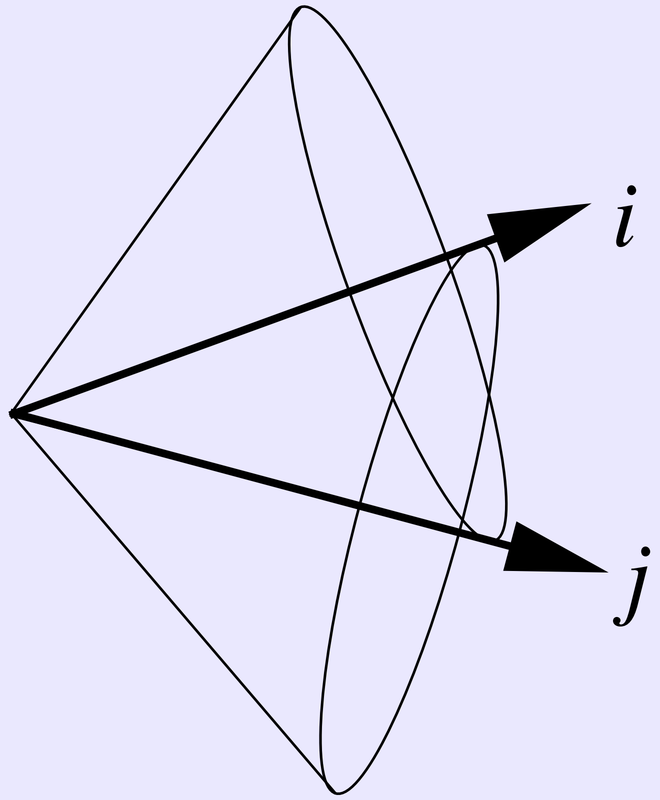
$$\Delta b > \lambda/\theta \sim (zp\theta)^{-1} \Rightarrow \theta_{e^+e^-}(zp\theta^2)^{-1} > (zp\theta)^{-1} \Rightarrow \boxed{\theta_{e^+e^-} > \theta}$$

angular ordering



Color flow and pull

- this so-called angular ordering property of soft emission is common to all gauge theories
- soft emissions that are not angular-ordered are suppressed by deconstructive interference
- the radiation from a pair of partons i and j in a color singlet configuration is mostly limited to two cones centered around i and j

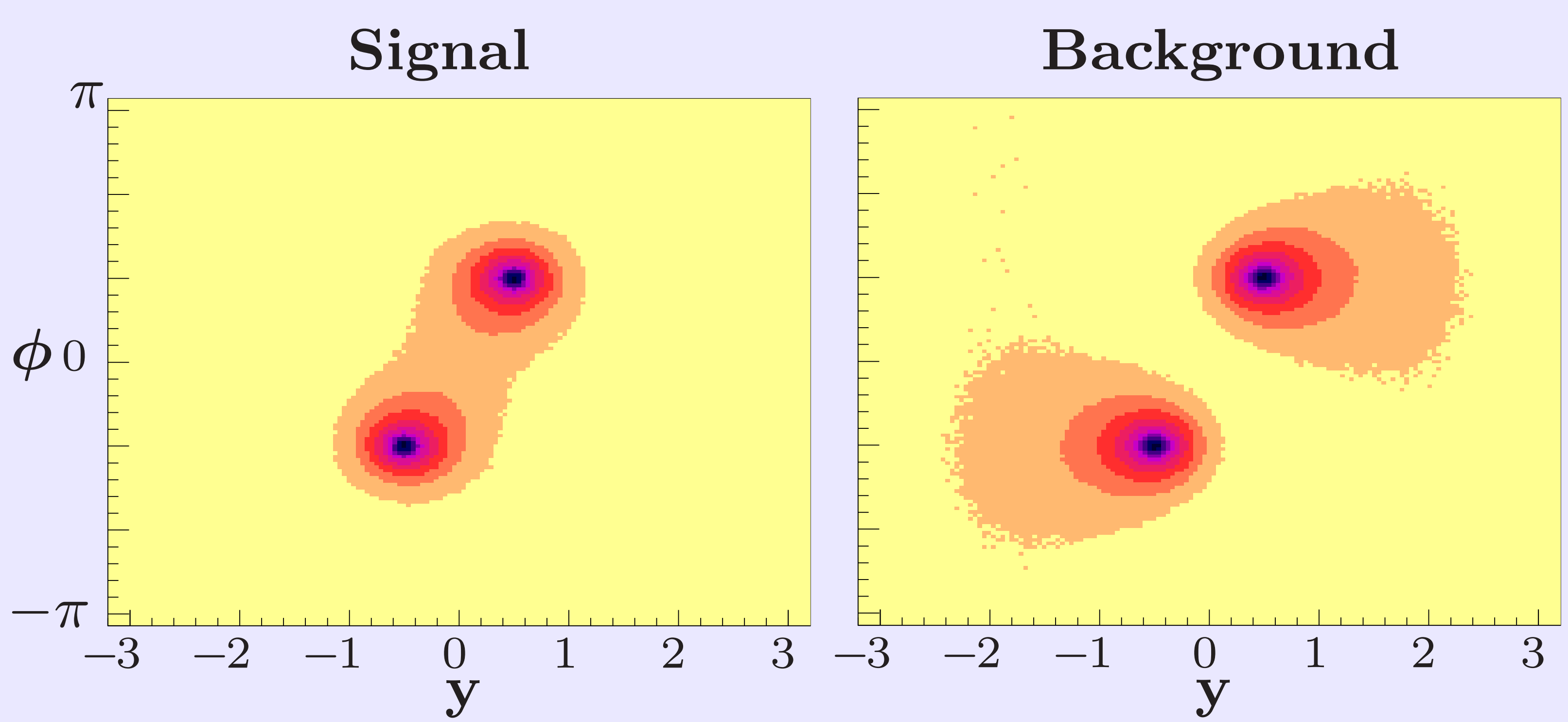


Color flow and pull

- this observation led to the introduction of the jet observable "pull"

$$\vec{t} = \sum_{i \in \text{jet}} \frac{p_T^i \, |r_i|}{p_T^{\text{jet}}} \, \vec{r}_i$$

- unfortunately, pull does not seem well suited to top-tagging



Jason Gallicchio and Matthew Schwartz arXiv:hep-ph/1001.5027

Color flow and pull

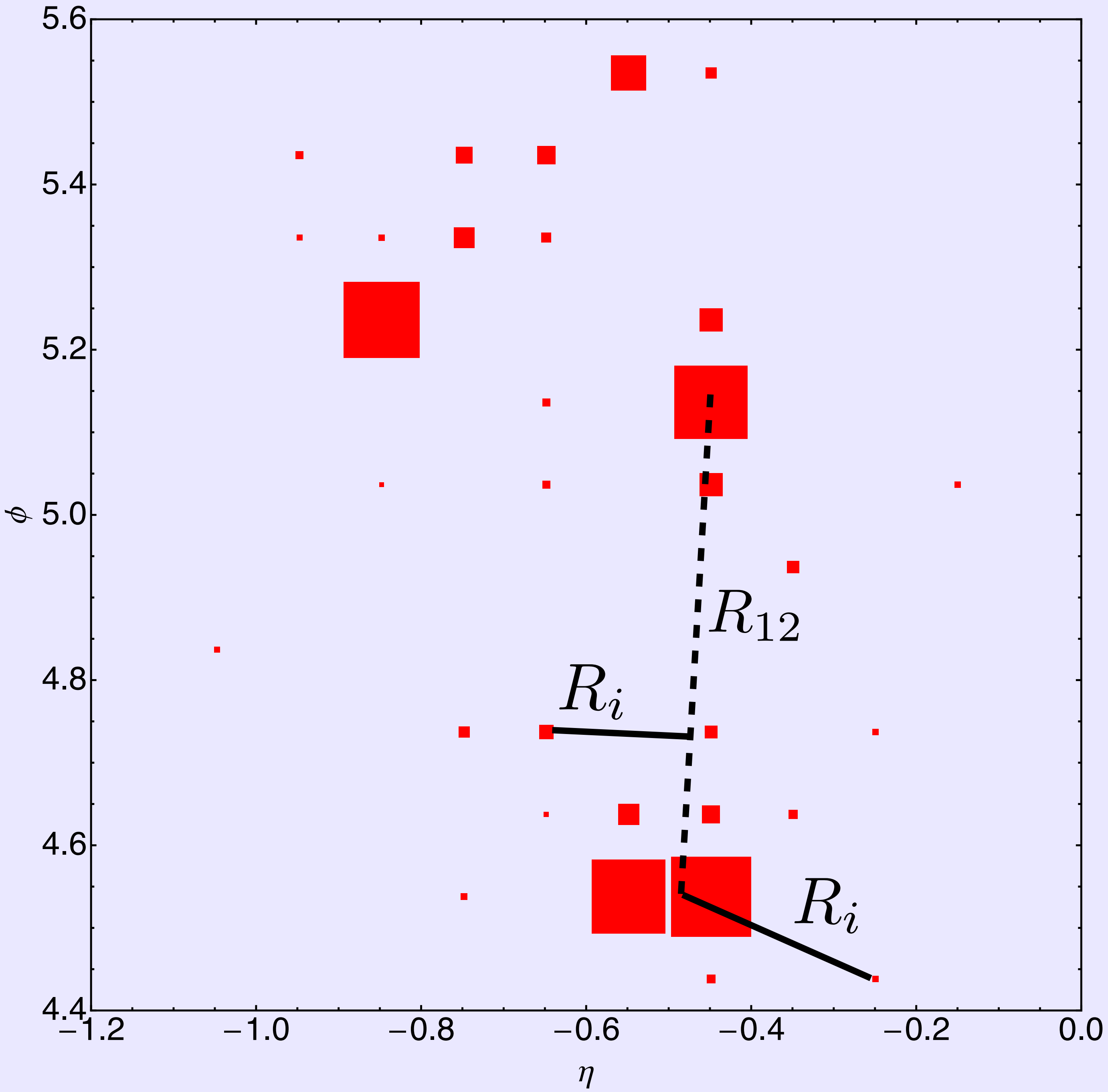
- D0 experiment has looked at the pull of hadronic W bosons in $t\bar{t}$ events
- Results in good agreement with Monte Carlo
- fraction of uncolored W bosons measured to be $f = 0.56 \pm 0.42$

Measurement of color flow in $\bar{t}t$ events from $p\bar{p}$ collisions at $\sqrt{s}$ = 1.96 TeV

arXiv/hep-ex:1101.0648v1

Dipolarity

- instead of something like pull, consider the entire radiation pattern of the W simultaneously



$$\mathcal{D} \equiv \frac{1}{R_{12}^2} \sum_{i \in J} \frac{p_{Ti}}{p_{T_J}} R_i^2$$

R_{12} is the separation between the two W subjets

p_{Ti} is the transverse momentum of cell i

p_{T_J} is the transverse momentum of the W

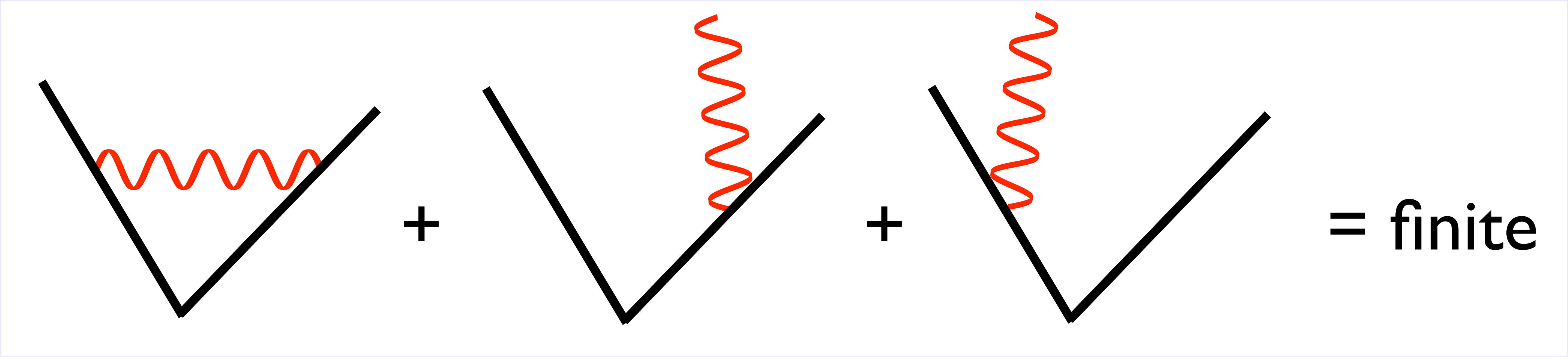
R_i is the distance between cell i and the line segment that spans the W subjets

Dipolarity

- dipolarity is essentially a two-subjet observable

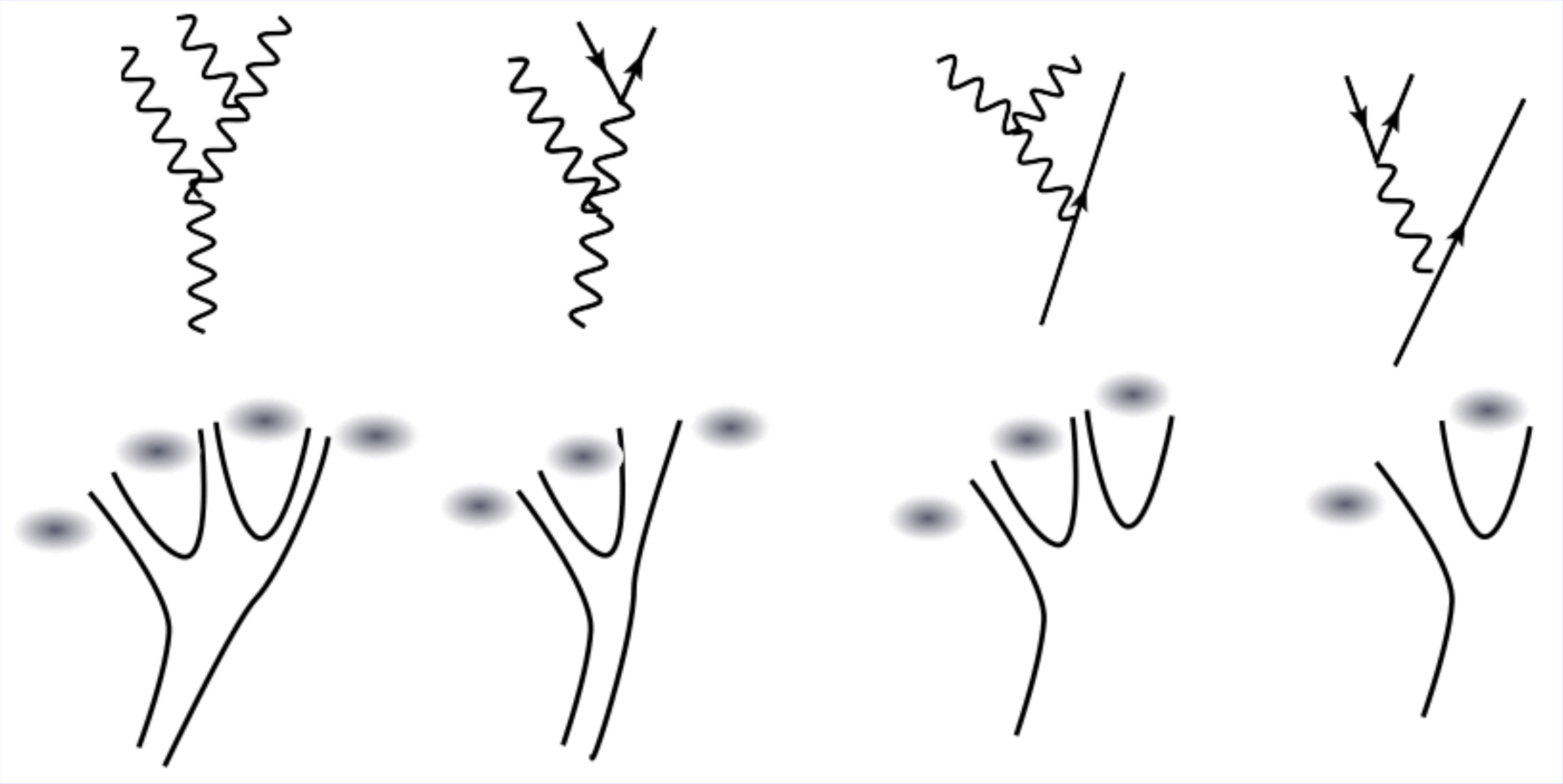
- expectation: top jets will yield small values of $\mathcal{D}$ whereas QCD jets will yield larger values of $\mathcal{D}$

- provided the two subjets are chosen in an IRC safe way, dipolarity is IRC safe as well



Dipolarity

- by incorporating dipolarity into the HEPToptagger, we can try to beat down QCD backgrounds



- even if a QCD fakes the kinematics of the top well, it will typically have a different color configuration

Dipolarity

- incorporate dipolarity into the HEPTopTagger by modifying step 4

Dipolarity

4. Recluster the five filtered constituents chosen in step 3 into exactly three subjets j_1, j_2, and j_3 ordered in descending p_T. Accept the fat jet as a top candidate if it passes any of the following three pairs of mass cuts:

$$i)\quad 0.2 \leq \arctan m_{13} \leq 1.3$$

$$i')\quad R_{\min} \leq \frac{m_{23}}{m_{123}} \leq R_{\max}$$

$$ii)\quad R_{\min}^2 \left(1 + \frac{m_{13}^2}{m_{12}^2}\right) \leq 1 - \frac{m_{23}^2}{m_{123}^2} \leq R_{\max}^2 \left(1 + \frac{m_{13}^2}{m_{12}^2}\right)$$

$$ii')\quad \frac{m_{23}}{m_{123}} \geq 0.35$$

$$iii)\quad R_{\min}^2 \left(1 + \frac{m_{12}^2}{m_{13}^2}\right) \leq 1 - \frac{m_{23}^2}{m_{123}^2} \leq R_{\max}^2 \left(1 + \frac{m_{12}^2}{m_{13}^2}\right)$$

$$iii')\quad \frac{m_{23}}{m_{123}} \geq 0.35$$

Here $R_{\min} = 85\% \times m_W/m_t$ and $R_{\max} = 115\% \times m_W/m_t$.

Dipolarity

4. Recluster the five filtered constituents chosen in step 3 into exactly three subjets j_1, j_2, and j_3 ordered in descending p_T. Accept the fat jet as a top candidate if it passes any of the following three pairs of mass cuts:

$$\boxed{\begin{aligned} &i) \quad 0.2 \leq \arctan m_{13} \leq 1.3 \\ &i') \quad R_{\min} \leq \frac{m_{23}}{m_{123}} \leq R_{\max} \end{aligned}} \Rightarrow j_W = j_2 + j_3$$

$$ii) \quad R_{\min}^2 \left(1 + \frac{m_{13}^2}{m_{12}^2}\right) \leq 1 - \frac{m_{23}^2}{m_{123}^2} \leq R_{\max}^2 \left(1 + \frac{m_{13}^2}{m_{12}^2}\right)$$

$$ii') \quad \frac{m_{23}}{m_{123}} \geq 0.35$$

$$iii) \quad R_{\min}^2 \left(1 + \frac{m_{12}^2}{m_{13}^2}\right) \leq 1 - \frac{m_{23}^2}{m_{123}^2} \leq R_{\max}^2 \left(1 + \frac{m_{12}^2}{m_{13}^2}\right)$$

$$iii') \quad \frac{m_{23}}{m_{123}} \geq 0.35$$

Here $R_{\min} = 85\% \times m_W/m_t$ and $R_{\max} = 115\% \times m_W/m_t$.

Dipolarity

4. Recluster the five filtered constituents chosen in step 3 into exactly three subjets j_1, j_2, and j_3 ordered in descending p_T. Accept the fat jet as a top candidate if it passes any of the following three pairs of mass cuts:

$$i)\quad 0.2 \leq \arctan m_{13} \leq 1.3$$

$$i')\quad R_{\min} \leq \frac{m_{23}}{m_{123}} \leq R_{\max}$$

$$ii)\quad R^2_{\min}\left(1+\frac{m^2_{13}}{m^2_{12}}\right) \leq 1-\frac{m^2_{23}}{m^2_{123}} \leq R^2_{\max}\left(1+\frac{m^2_{13}}{m^2_{12}}\right)$$

$$ii')\quad \frac{m_{23}}{m_{123}} \geq 0.35$$

$$\Rightarrow j_W = j_1 + j_2$$

$$iii)\quad R^2_{\min}\left(1+\frac{m^2_{12}}{m^2_{13}}\right) \leq 1-\frac{m^2_{23}}{m^2_{123}} \leq R^2_{\max}\left(1+\frac{m^2_{12}}{m^2_{13}}\right)$$

$$iii')\quad \frac{m_{23}}{m_{123}} \geq 0.35$$

Here $R_{\min} = 85\% \times m_W/m_t$ and $R_{\max} = 115\% \times m_W/m_t$.

Dipolarity

- incorporate dipolarity into the HEPTopTagger by modifying step 4
- calculate the dipolarity of the pair of subjets identified in step 4
- if more than one pair of subjets passes the mass cuts, choose the smaller dipolarity
- make a dipolarity cut $\mathcal{D} < \mathcal{D}_{\rm max}$
- need to make one more choice: what radiation goes into the sum?

Dipolarity

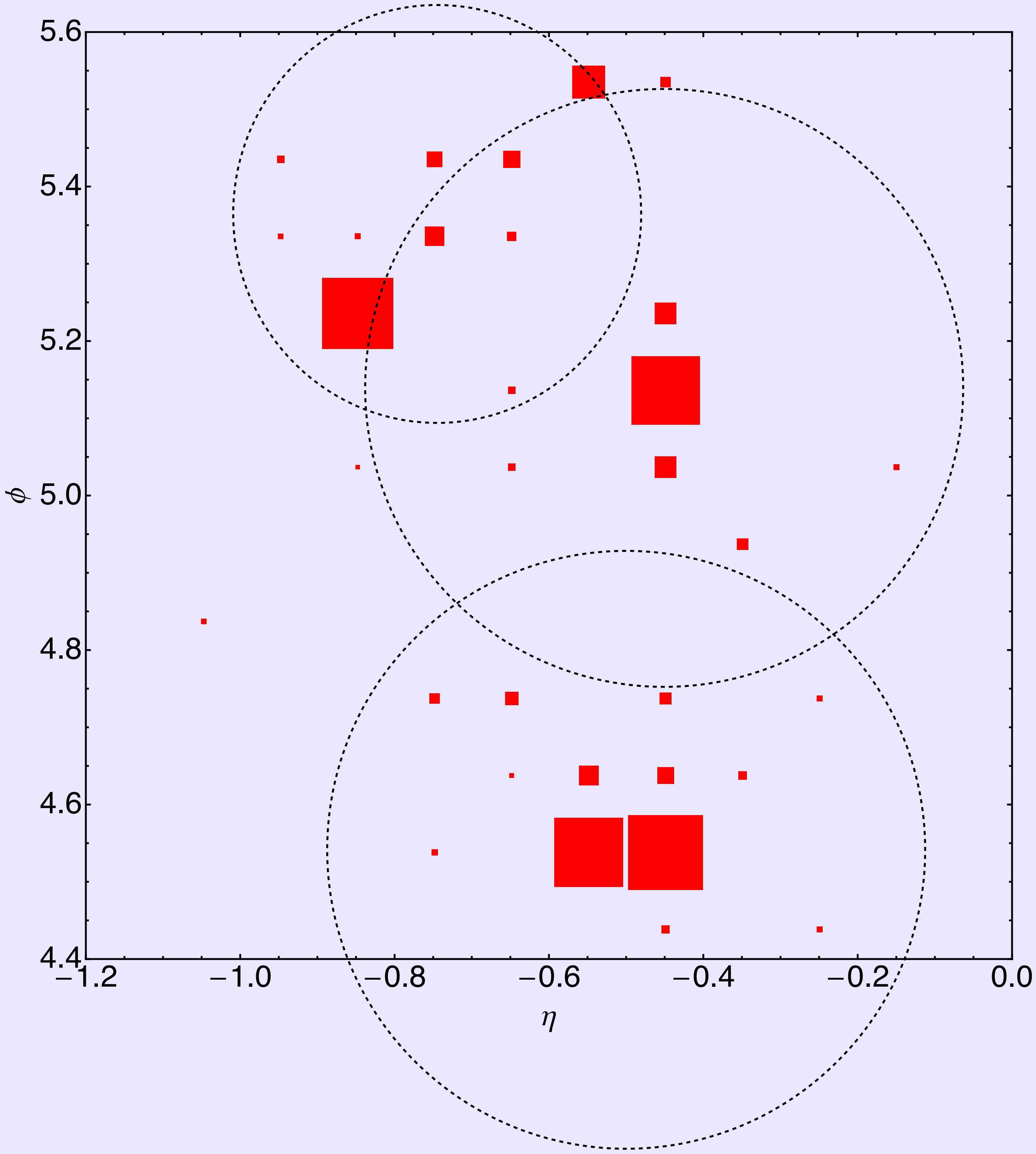
- we find that the criterion used to select the radiation that enters the sum has significant impact on the utility of dipolarity of as a discriminant

- angular ordering implies that most of the radiation from the W is within the pair of cones of radius ΔR

- choose our cones to be somewhat smaller, $\Delta R/\sqrt{2}$, to minimize contamination from the underlying event

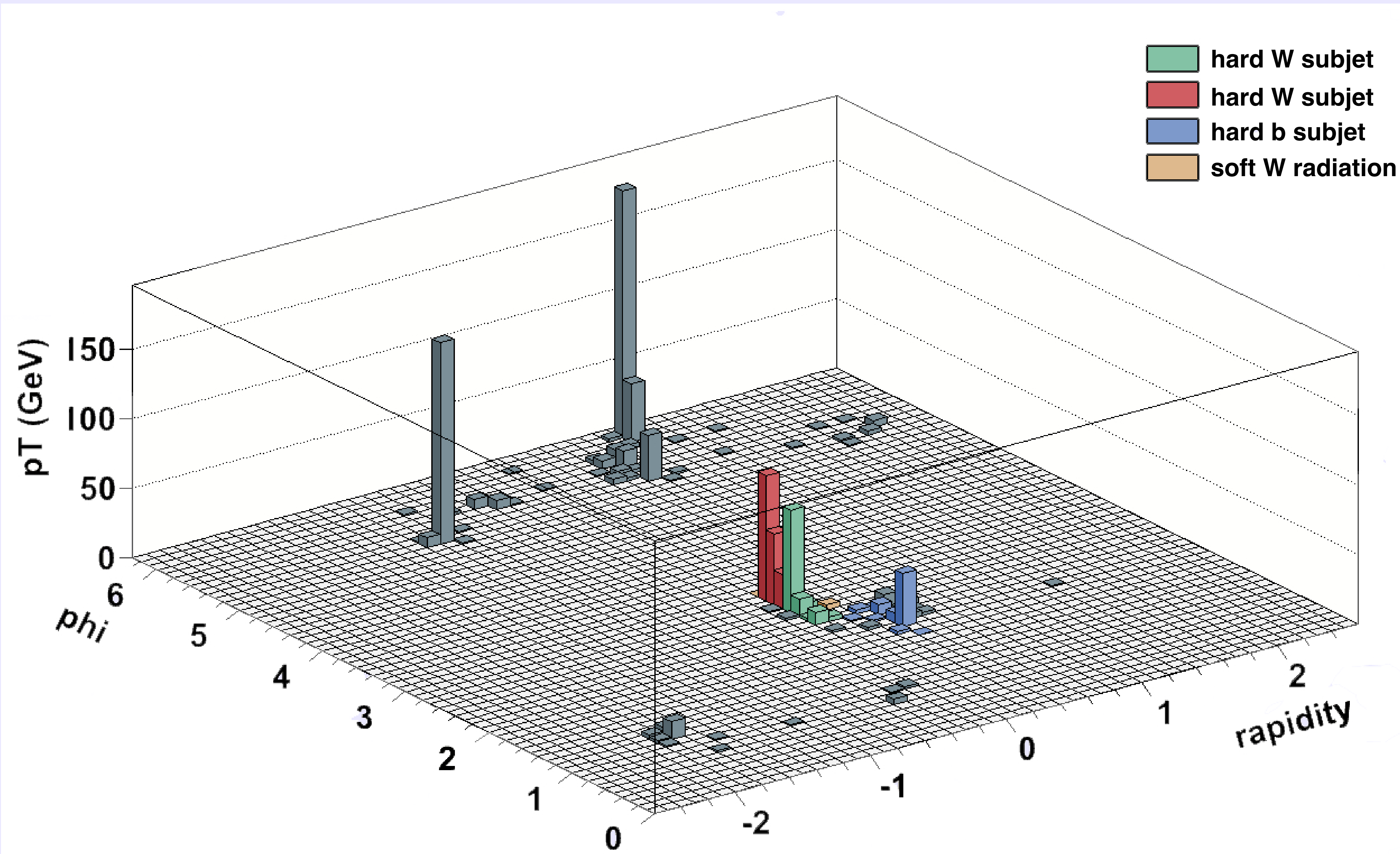
- also remove any radiation in the neighborhood of the b subjet

Dipolarity



Dipolarity

Legoplot for a top jet with hard substructure as identified by the HEPTopTagger



Results

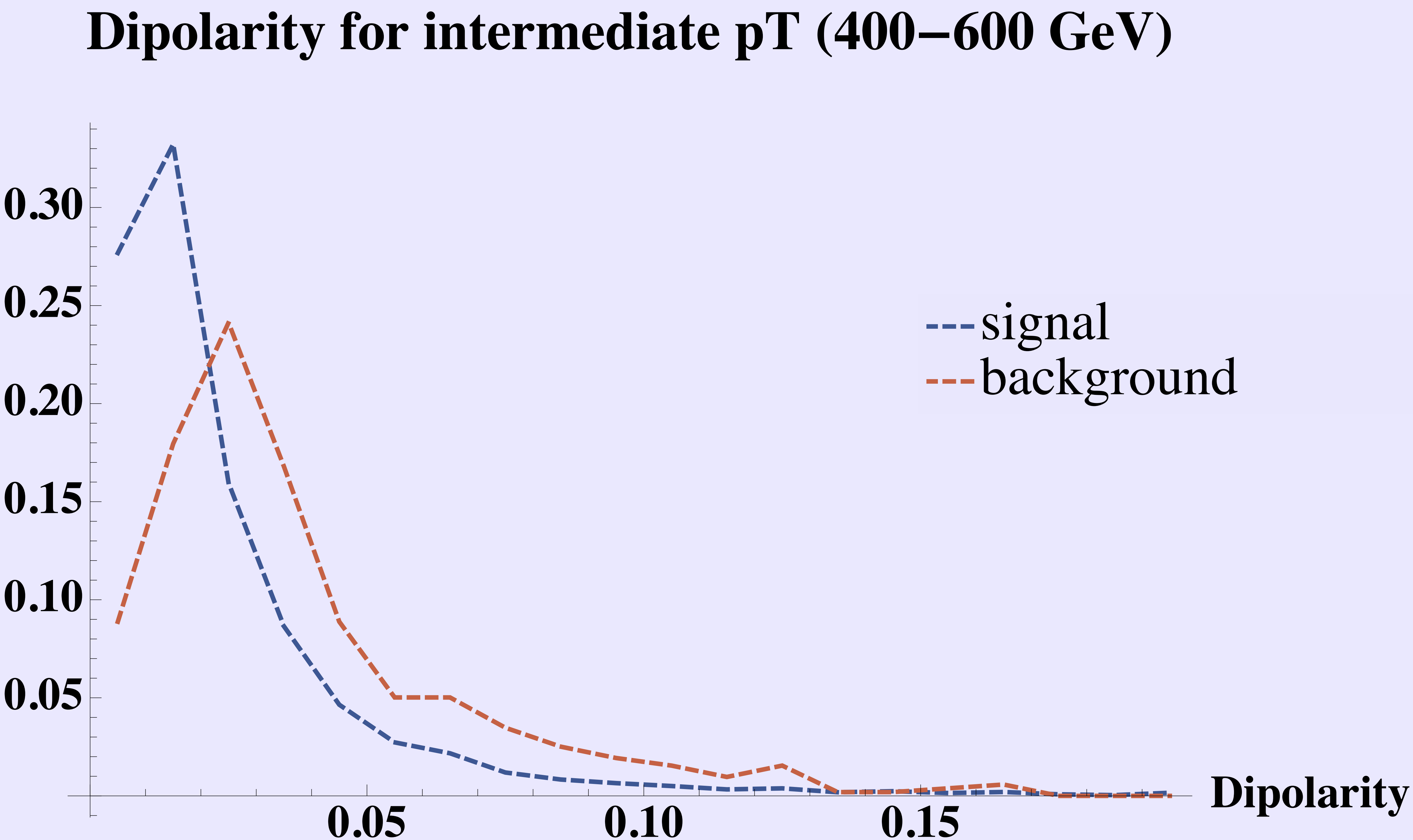
- test the modified top-tagger on three event samples from BOOST 2010:

    1. HERWIG 6.510 - angular ordering
    2. PYTHIA 'DW' - Q^2 ordering
    3. PYTHIA 'Perugia' - p_T ordering

- jet clustering with FastJet 2.4.2

- zeroth order detector mock-up by binning particles into 0.1×0.1 cells in $y-\phi$ space

Results



for HERWIG events passing default mass cuts

Results

- we want to see whether dipolarity cuts are essentially orthogonal to the kinematic cuts imposed by the HEPTopTagger

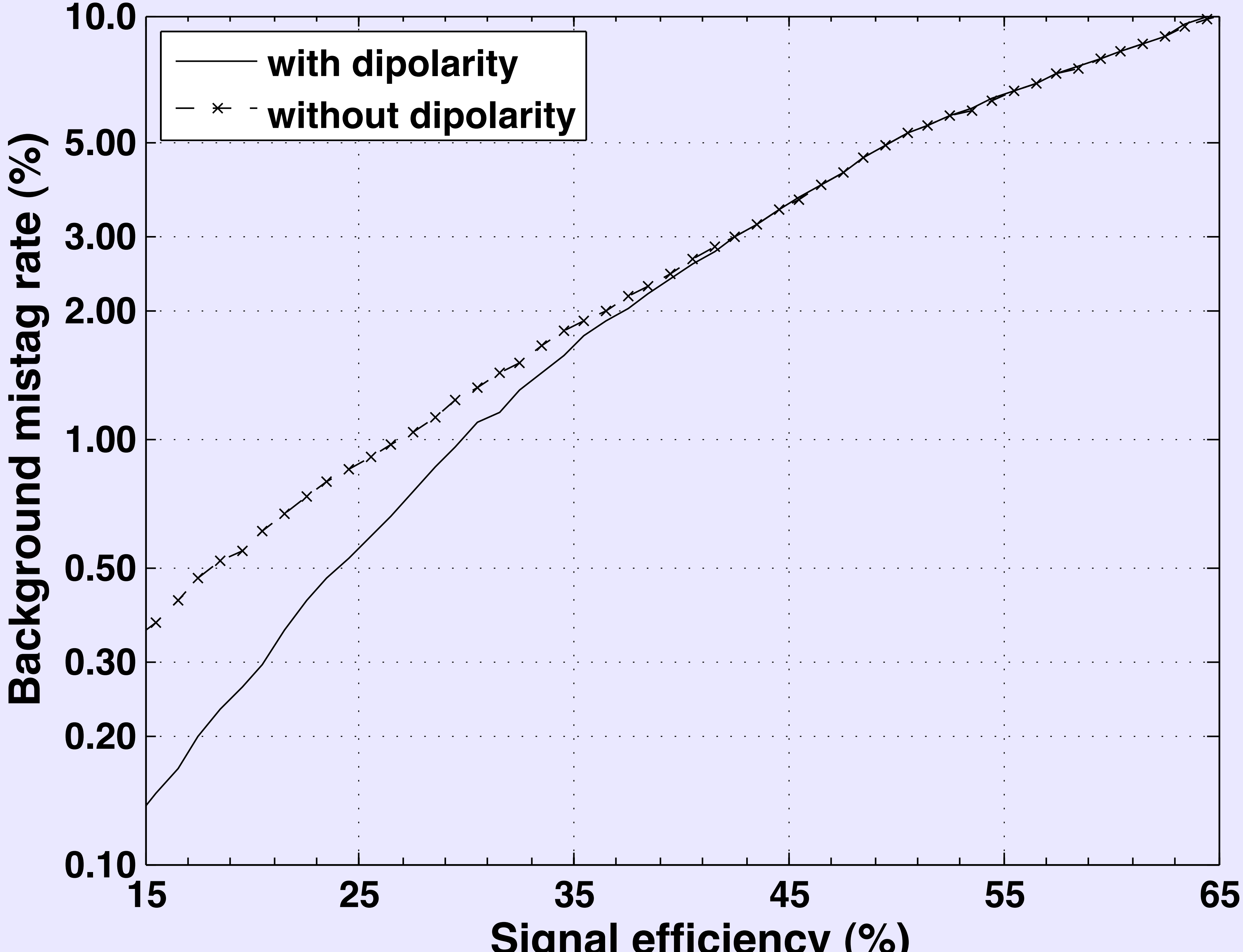
- so include cuts on the reconstructed mass of the top so that the HEPTopTagger is using a full compliment of kinematic cuts

- optimize the cuts use Monte Carlo code to finely sample the space of cuts

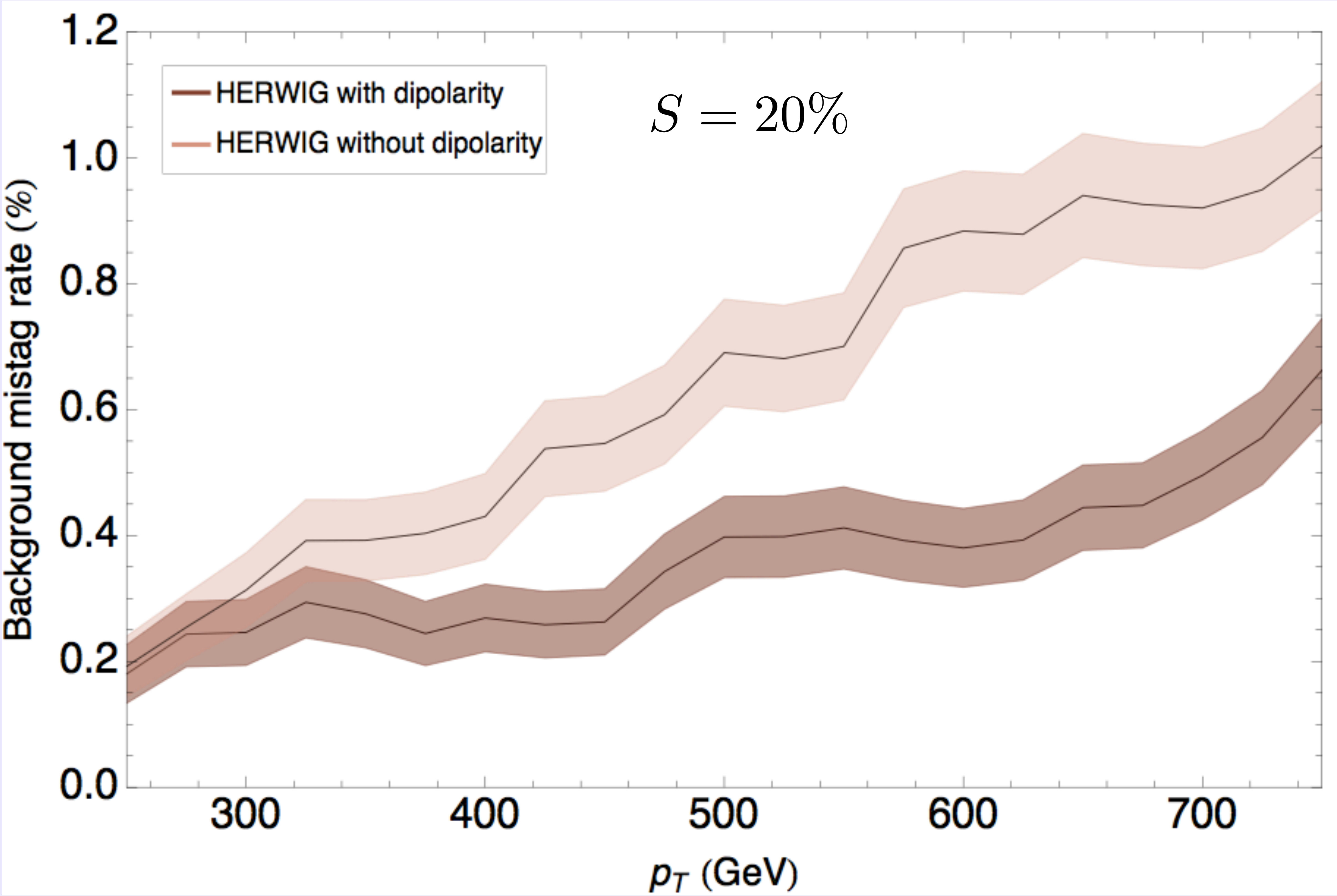
- at each signal efficiency S choose cuts so that the background mistag rate B is minimized

Results

- improves background rejection at lower S

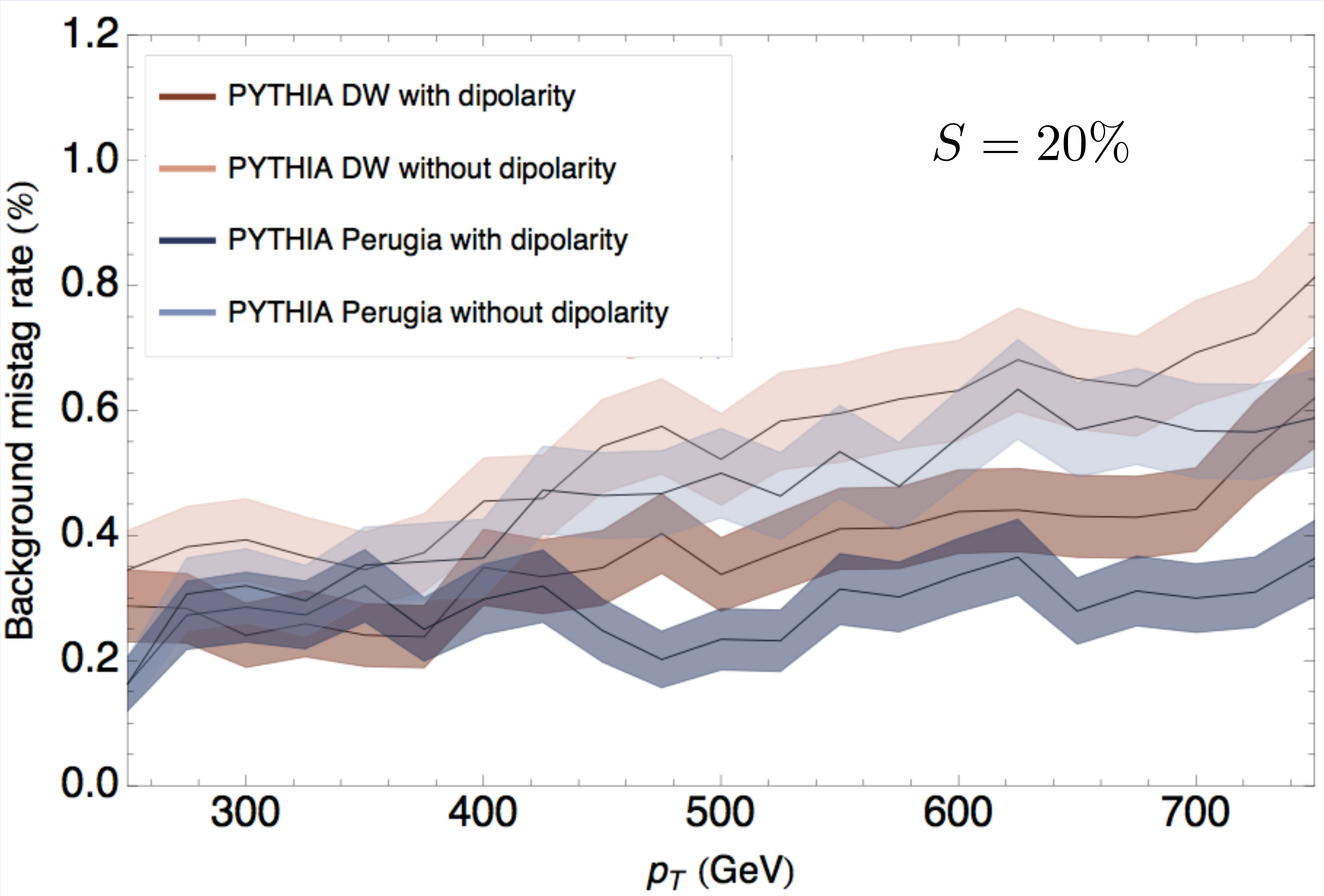


Results



HERWIG with dipolarity
HERWIG without dipolarity
$S = 20\%$
Background mistag rate (%)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
300
400
500
600
700
p_T (GeV)

Results



PYTHIA DW with dipolarity
PYTHIA DW without dipolarity
PYTHIA Perugia with dipolarity
PYTHIA Perugia without dipolarity
$S = 20\%$
Background mistag rate (%)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
300
400
500
600
700
p_T (GeV)

Results

- for intermediate to high p_T $(400\ \mathrm{GeV} < p_T < 800\ \mathrm{GeV})$ and for lower signal efficiencies including dipolarity cuts can improve background rejection

- there is sizable disagreement between the different Monte Carlo event samples

- this disagreement probably has its origin in the details of the parton showers (not e.g. the underlying event models)

- this is not surprising - theoretical understanding of color coherence (and its inclusion in MC) is limited

Summary & Outlook

- introduced a jet observable "dipolarity" that can distinguish between different color configurations in jets with significant mass drops

- incorporating dipolarity in the HEPTopTagger improves background rejection

- due to theoretical uncertainties, the ultimate utility of dipolarity awaits real data

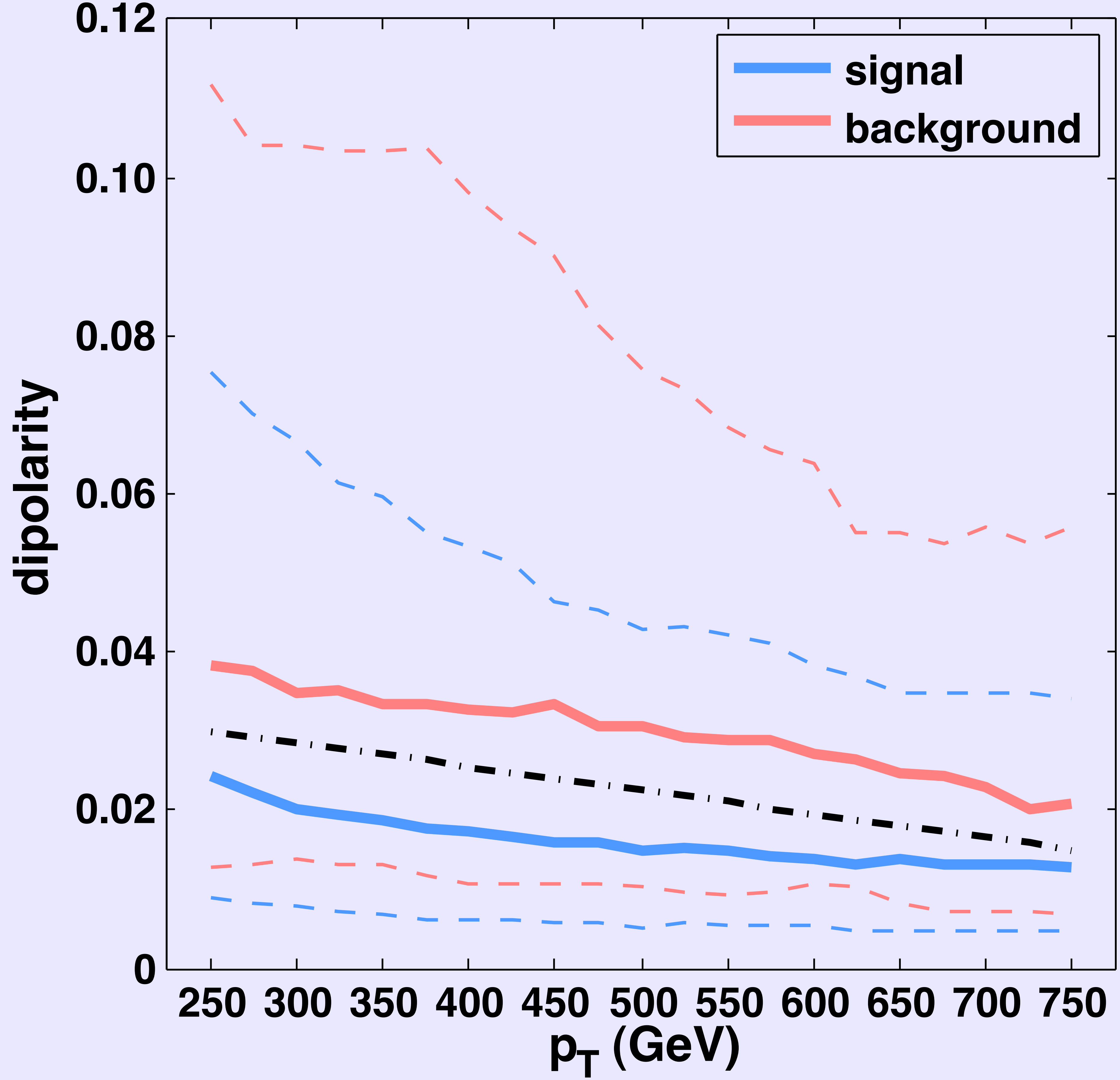
- dipolarity should have other applications outside of top-tagging (e.g. W/Z physics, heavy Higgs)

Summary & Outlook

- theoretical understanding of color flow and other jet substructure observables should benefit from confrontation with LHC data

- e.g. CMS just published CMS PAS JME-10-013 'Study of Jet Substructure in pp Collisions at 7 TeV in CMS'
    - measured mistag rate for a W tagging and top tagging algorithm
    - good agreement with Monte Carlo (especially Herwig++)

backup slides


0.12
0.10
0.08
0.06
0.04
0.02
0
dipolarity
250
300
350
400
450
500
550
600
650
700
750
p_T (GeV)
signal
background

Sequential jet clustering algorithms

1. find the smallest of the d_{ij} and d_{iB}
2. if it is d_{ij} combine i and j and return to step 1
3. if it is d_{iB} declare i to be a jet and remove it from the list of four-vectors, returning to step 1
4. continue until there are no particles left

$p = 0 \Rightarrow$ Cambridge-Aachen

$p = 1 \Rightarrow$ kT

$p = -1 \Rightarrow$ anti-kT

$$d_{ij} = \xcancel{\min(p_{ti}^{2p}, p_{tj}^{2p})} \frac{\Delta R_{ij}^2}{R^2}$$

$$d_{iB} = \xcancel{p_{ti}^{2p}}\ 1$$

Some jet substructure techniques

1. Break a fat (R=1.5) C/A jet j into subjets j_1 and j_2 by undoing the last stage of clustering; label so that $m_{j_1} > m_{j_2}$

2. If (i) there is a significant mass drop $m_{j_1} < \mu m_j$ and (ii) the splitting is not too asymmetric then exit the loop

$$y = \frac{\min(p_{tj_1}^2, p_{tj_2}^2)}{m_j^2} \Delta R_{j_1,j_2}^2 > y_{\rm cut} \qquad y \approx \frac{\min(p_{Tj_1},\ p_{Tj_2})}{\max(p_{Tj_1},\ p_{Tj_2})}$$

3. Otherwise redefine j_1 as j and go back to step 1

Here $\mu = 0.67$ and $y_{\rm cut} = 0.09$

4. Finally, recluster with $R_{\rm filt} = \min(0.3, \frac{R_{b\bar{b}}}{2})$ and use the three hardest subjets to calculate the filtered Higgs mass