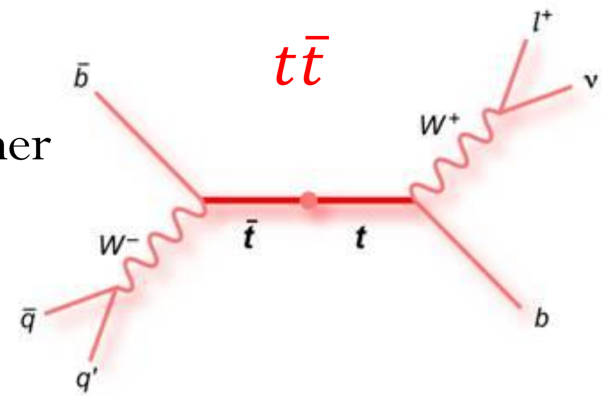
Tagging Boosted Top Decays Using Jet Substructure at 1.4TeV

Alasdair Winter
University of Birmingham
08/03/2017

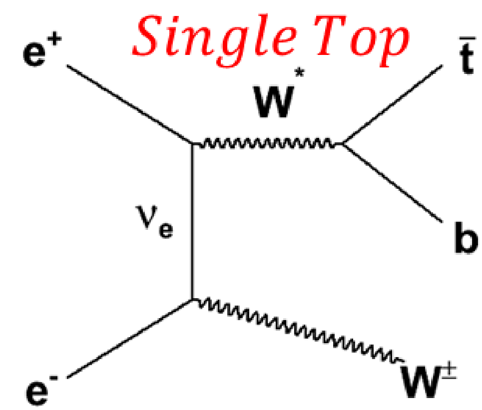
Overview

- Aims:
  - identify Top decays in boosted events
  - distinguish different Top type events from each other and background
  - select events without biasing the Top mass
  - Investigate impact on A_{FB}, CP violation



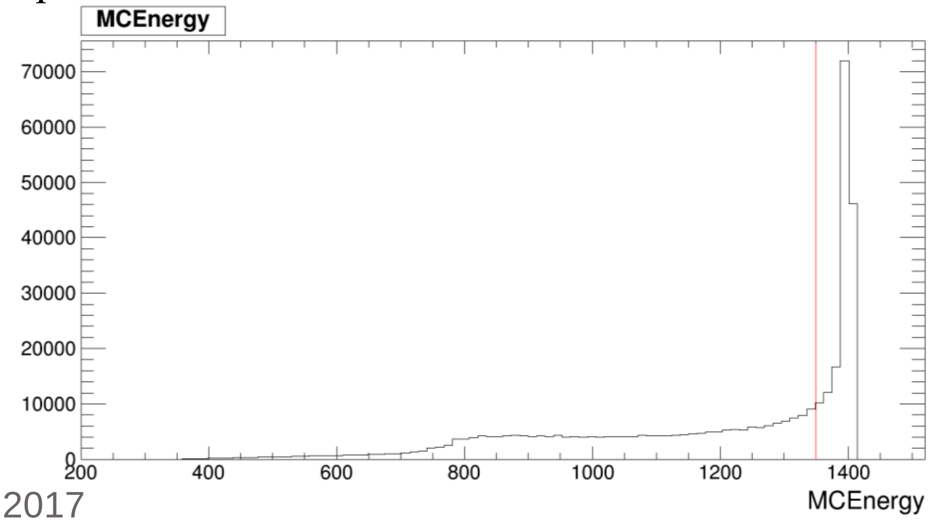
- Boosted topology makes conventional Top tagging techniques a challenge- btagging alone no longer reliable!

- Approach based on the concept of Fat Jets and jet substructure



MC Sample Selection

- Samples are inclusive 6 fermion final states (*qqqqlv*) consistent with semileptonic $t\bar{t}$ decays
- Signal definition: events most closely resembling genuine $t\bar{t}$ or single Top
  - MC truth requirements:
    - 2 b quarks
    - 2 additional quarks
    - Combination of three quarks (minimum 1 b quark) consistent with Top mass of 168-178GeV
    - Exclude events where the isolated lepton is a Tau



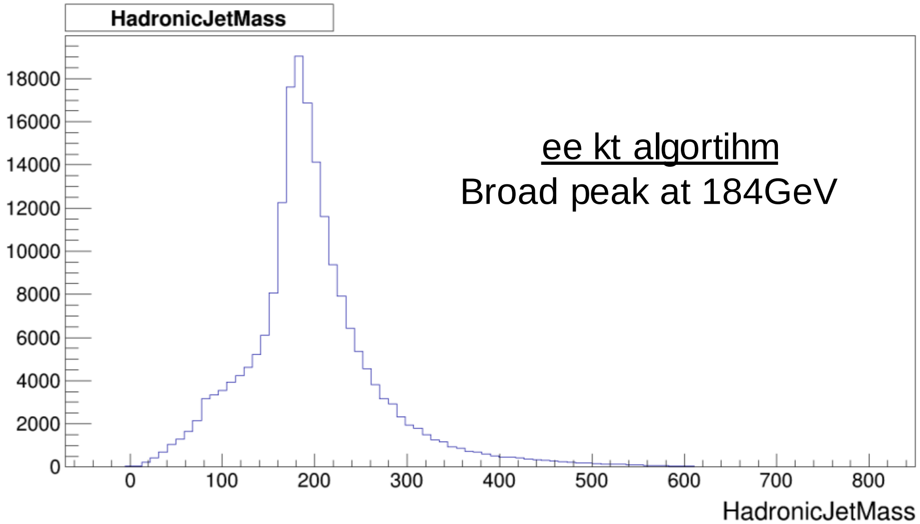
- Broad energy spectrum
- Two signal categories defined
  - >1.35TeV – “boosted Tops”
  - <1.35TeV – “low energy”

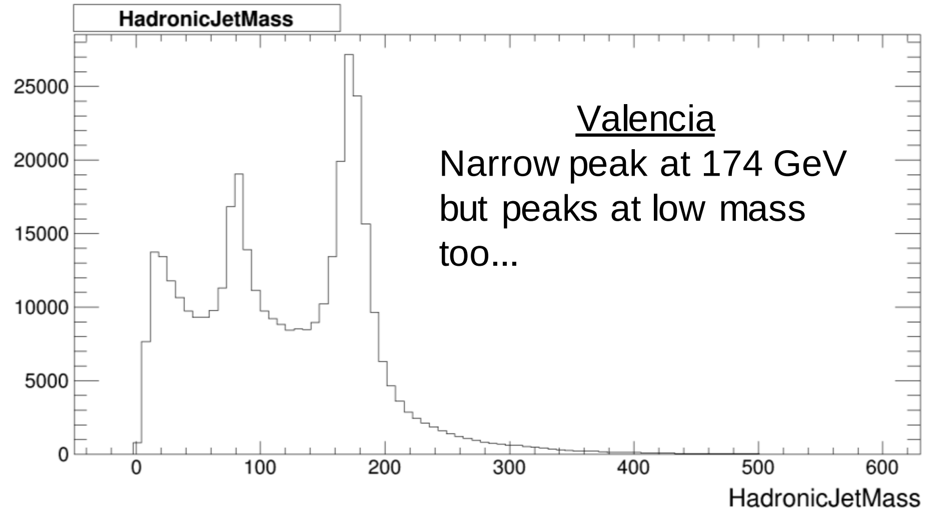
Analysis Strategy

1. Find Isolated Lepton in TightSelectedPFOs- 85% purity
2. Cluster remaining PFOs into 2 Fat Jets
3. Determine Fat Jet structure- multiplicity, nSubJettiness, angular distribution of subjets
4. Apply 2 BDTs using kinematic and jet substructure variables. Optimise BDTs separately for boosted and low-energy Tops
5. Use BDT response for each classifier to perform a final discrimination
6. Validate Top mass & angular dependence on final selection as precursor to physics studies

Fat Jet Finding

- Combine remaining PFOs into two jets
  - ee_kt algorithm and Valencia (1.5,1,1)
  - ExclusiveNJets=2
- Declare jet with the highest energy to be the hadronically decaying Top





- Valencia algorithm has slightly better background discrimination due to impact on various jet related variables used in the BDT.
- Low mass peaks when the Top decays close to the beam line, mostly disappear after event selection

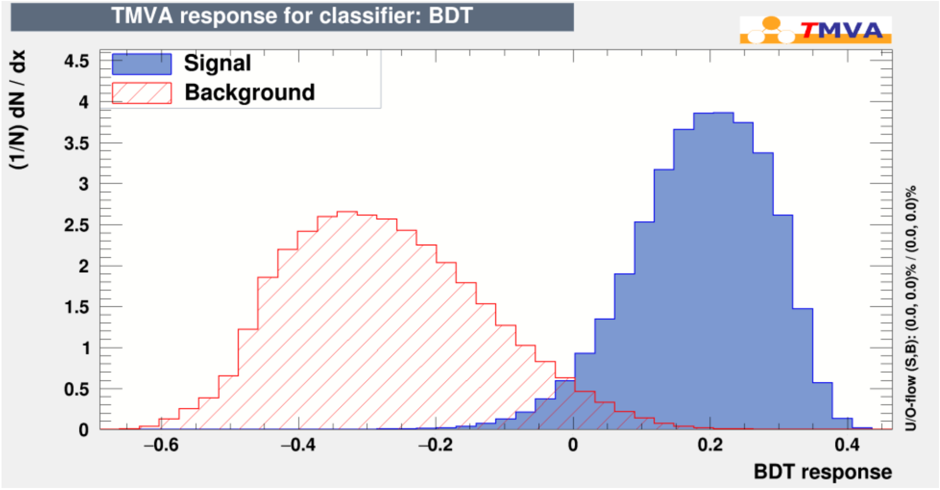
Jet Substructure

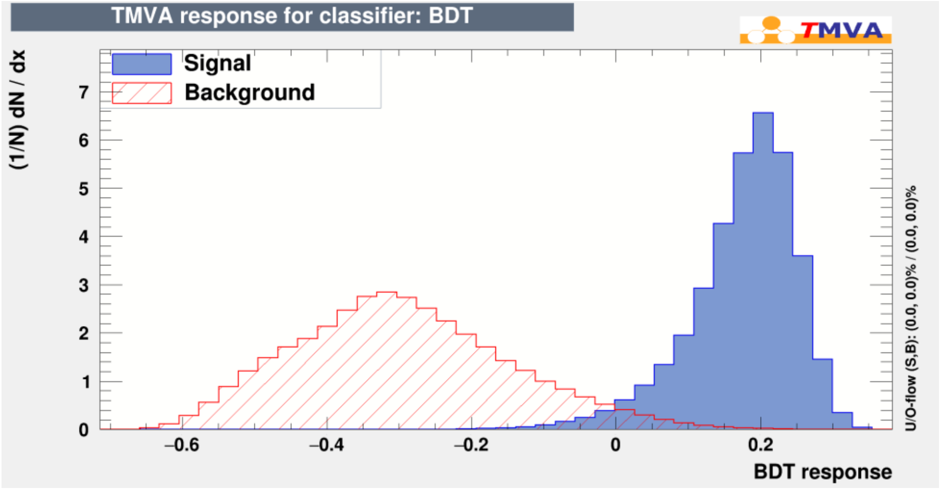
1. **Jet Multiplicity**
   - Count number of objects within a Fat Jet
   - Use jet finder with small resolution parameter (kt algorithm, R=0.05) to produce "microjets" rather than PFOs as they are better defined physics objects
   - What if we decrease R?
     - Better background discrimination
     - Increased sensitivity to hadronization modelling → systematic effects → under investigation!!
2. **N-Subjettiness**
   - Means of estimating how many subjets are within a Fat Jet
   - Already used in top searches at LHC[1]
3. **Angular Relations**
   - Recluster Fat Jet into 3 subjets using kt algorithm
   - Order jets according to energy
   - Measure relative angles between each subjet pairing
   - For backgrounds with less than three subjets, expect to see small angle between two subjets due to forced splitting of a true jet

[1] J.Thaler, K.Tilburg, "Identifying Boosted Objects with N-subjettiness", https://arxiv.org/abs/1011.2268

BDT Response

Boosted Top Classifier





Low energy Classifier

Combined Selection

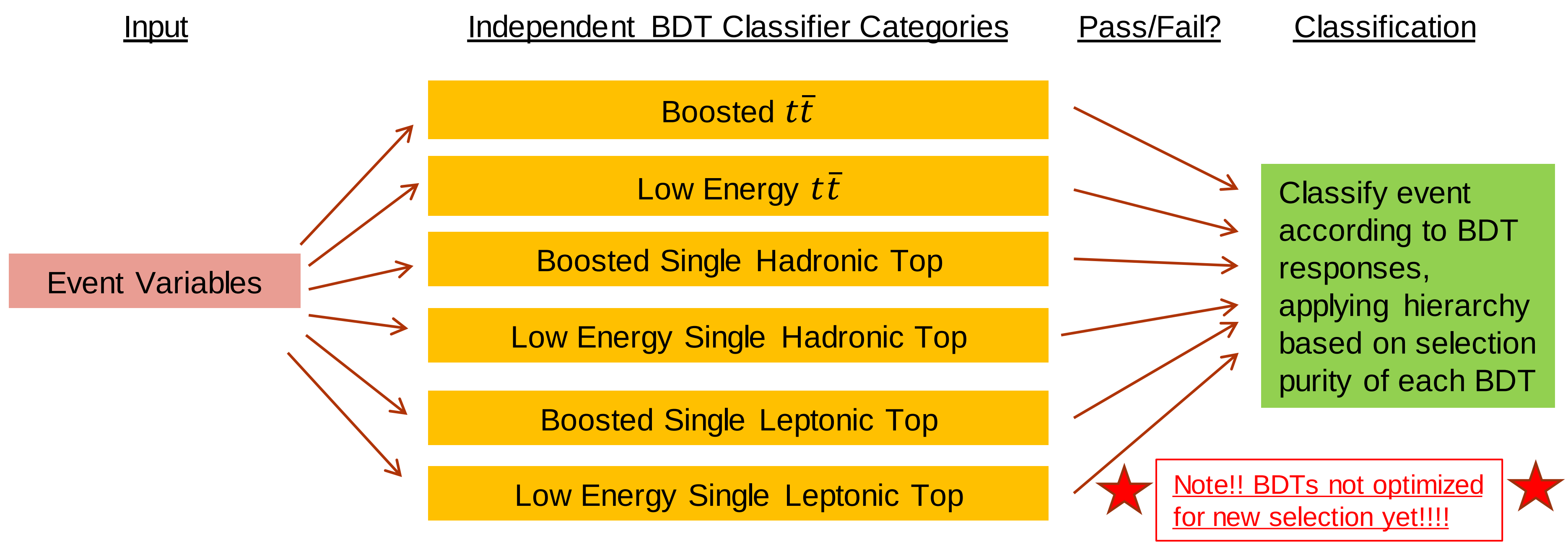
- Accept any events that pass either of the first tier BDT cuts

Process	σ(fb)	Generated Events (k)	Selection Efficiency (%)	Expected Events Passing for 1.5ab^{-1}
Signal	58.2	3,800	66	58,000
ee→qqlv	4309.7	980	0.06	3,800
ee→qqqq	1328.1	300	0.02	4,100
ee→qq	4009.5	460	0.09	5,100
ee→qqqqlv	45.4	180	20	13,300
ee→qqqqll	71.7	210	3.2	3,400
ee→qqqqvv	24.7	230	1.9	700

- Significance = 195 $(S/\sqrt{S+B})$
- Signal Efficiency = 66 %
- Background efficiency = 0.2 %
- Purity after selection = 65%

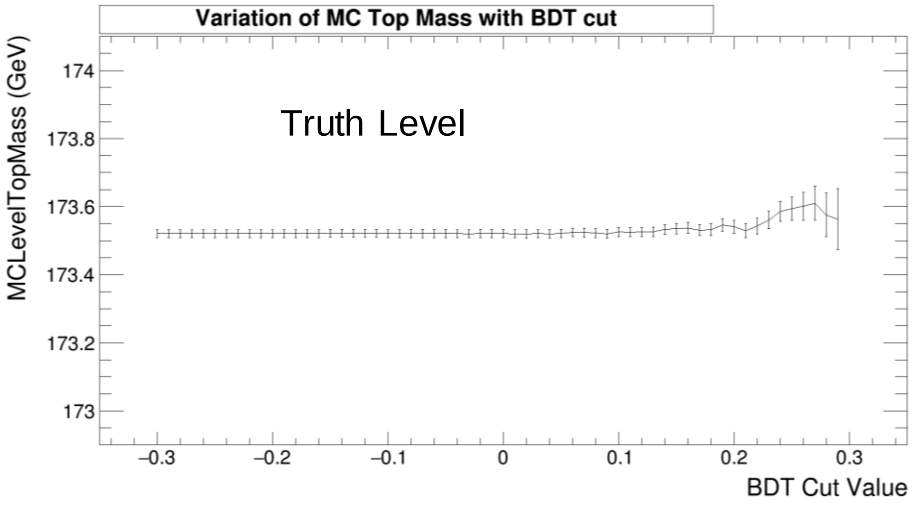
Improved efficiency using separate BDTs seen from cross-feeding effects

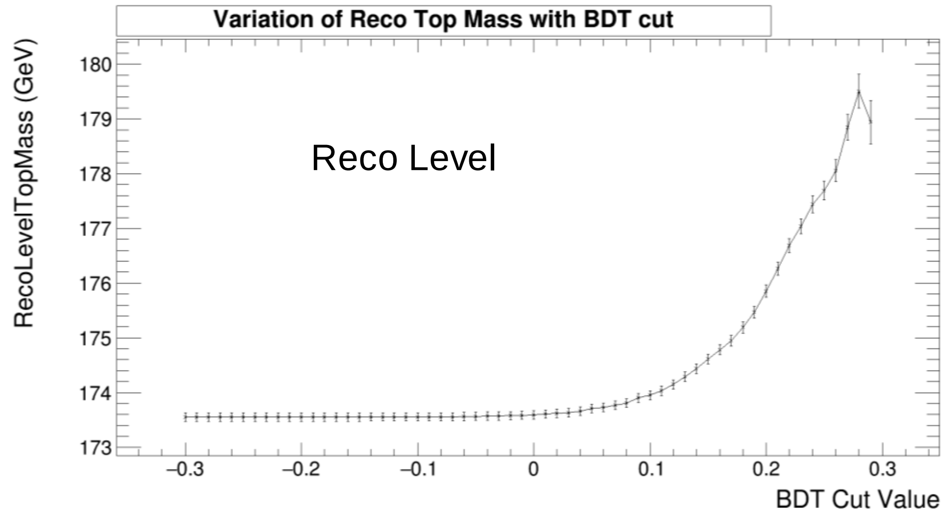
Generic Top Event Classifier



Signal Category	Efficiency (%)	Purity (%)	Signal candidates (k) for $L = 1.5ab^{-1}$	Significance $N_s/\sqrt{N_s + N_B}$
$t\bar{t}$	59	53	30	125
Single Hadronic Top	30	23	11	50
Single Leptonic Top	20	10	3	17

Top Mass Dependence

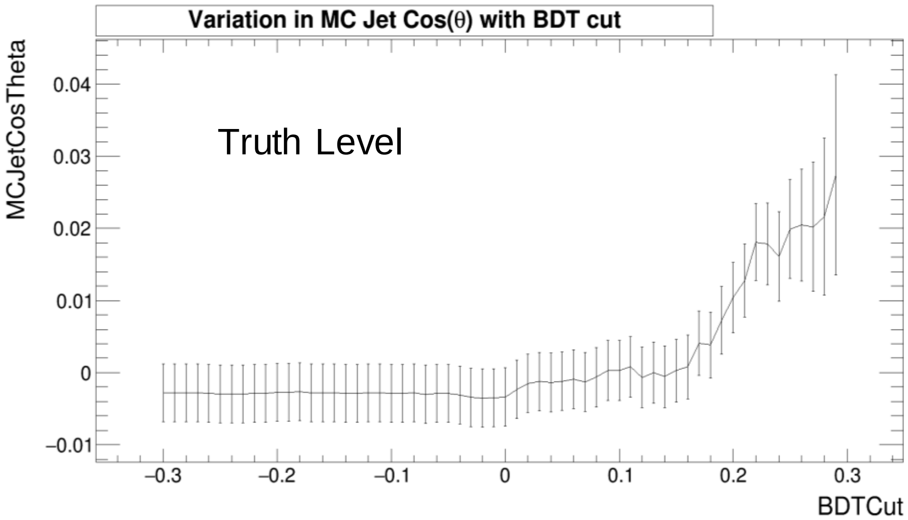


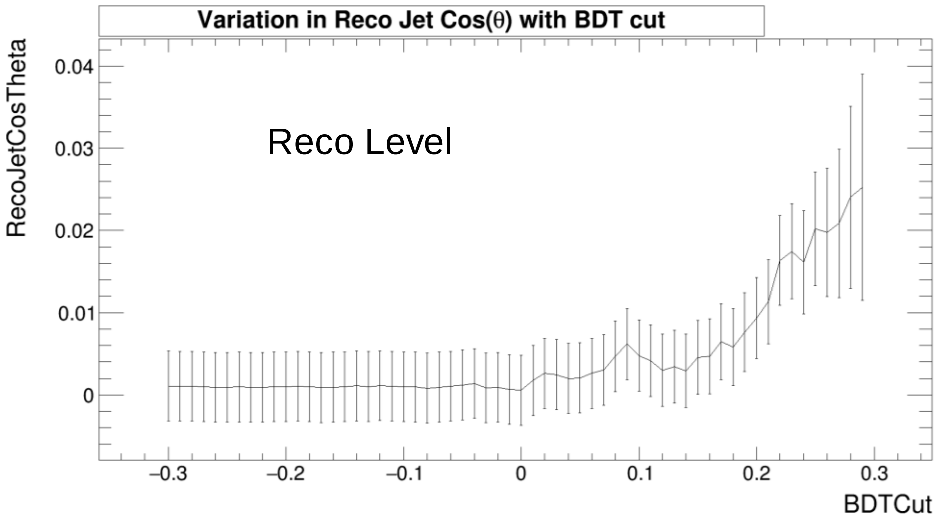


BDT cut typically placed between 0.1-0.2

- Fitting gaussian to hadronic Top jet mass in the high energy $t\bar{t}$ samples
- No significant bias seen at truth level
- Bias starts to appear at reco level in the optimal cut region- under investigation

Top Angular Dependence





- Cos(θ) defined by the jets of the hadronically decaying Top
- Minimal bias seen for both truth and reco level

Summary

- A new Top tagger has been developed using multivariate techniques and jet substructure to identify semileptonic $t\bar{t}$ decays and single Top events

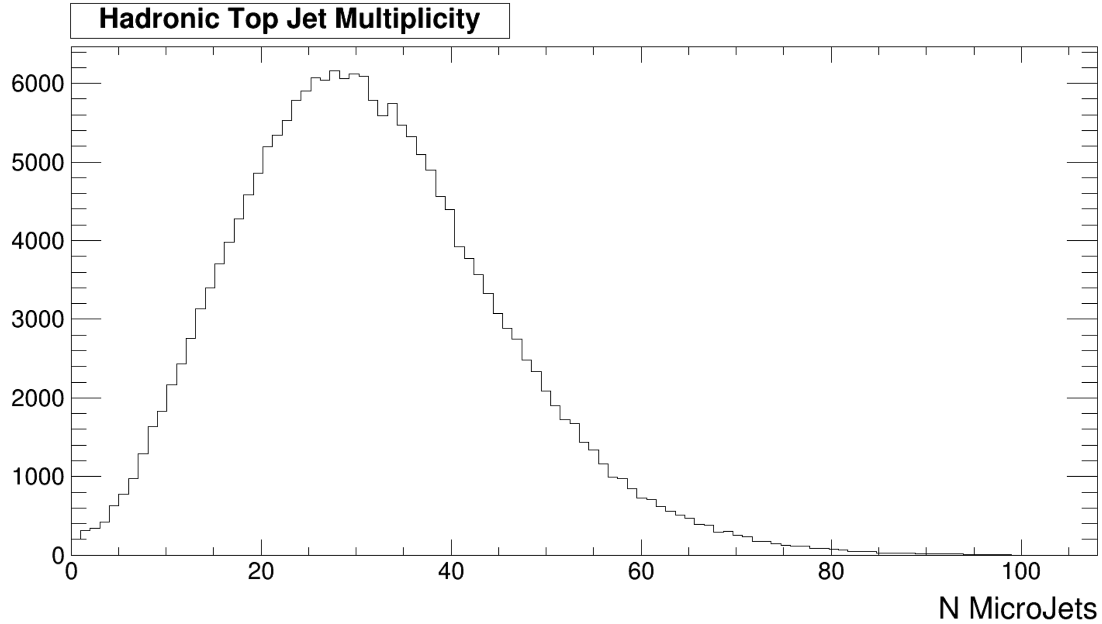
- Approach is being extended to classify the form of the Top events ($t\bar{t}$, single hadronic Top, single leptonic Top)
  - Optimization in progress

- Preliminary studies show small bias in reconstructed Top mass, and no angular bias
  - Promising for physics studies to follow

Backup Slides

Jet Multiplicity- MicroJets

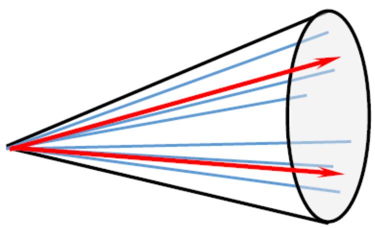
- Use jet finder with small resolution parameter ('microjets') rather than PFOs as they are better defined physics objects-
  - Become equivalent in limit of small ΔR
  - Based on old idea, subjet multiplicity (early LEP 1)
- Micro jets found by running FastJet on the Fat Jet's PFOs
  - kt_algorithm, R=0.05
- Better results seen for smaller R, but need to examine systematic effects from hadronization model



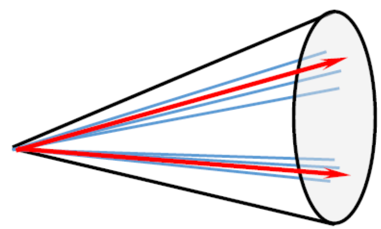
N Subjettiness

- Metric for determining how many subjets in a Fat Jet:
  - Recluster Fat Jet into N subjets: kt-algorithm R=0.3
  - Sum separation of all constituents of the Fat Jet from their nearest subjet

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min\{\Delta R_{1,k}, \Delta R_{2,k}, \ldots, \Delta R_{N,k}\}$$



High τ_2 (consitutents spread out)

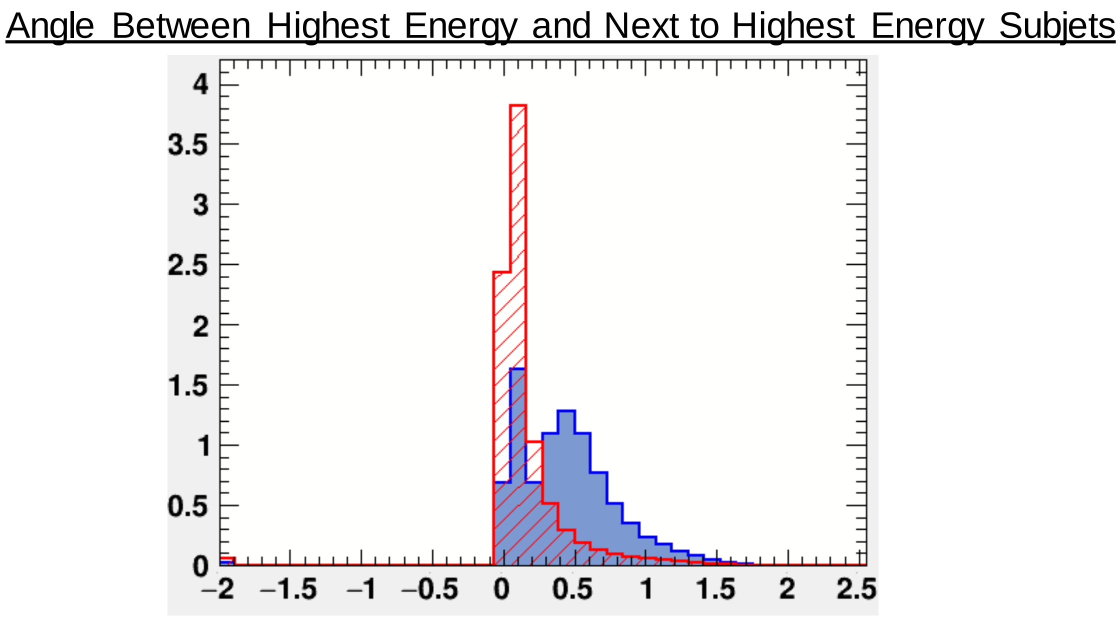


Low τ_2 (constituents close to subjet axes)

- N= number of subjets being proposed
- k= index of Fat Jet constituents
- d_0= normalization factor

Subjet Angular relations

- For signal events we expect 3 jets from the Top decay, backgrounds typically only have 1 or 2
- Study characteristics of three subjets within the Fat Jet
- Order subjets from highest to lowest energy
- Measure the angles of all subjets relative to each other
- For background events, third jet will be formed by splitting a "true" jet into two
  - Expect to see small angle of separation between at least two subjets



4) Variables Used for BDT

Leptonic Top Properties
Jet Mass
Jet Energy
Jet Multiplicity
Jet Pt
Lep Charge
Lep Energy
Lep Pz
Lep Pt
Lep PID

Hadronic Top Properties
Fat Jet Energy
Fat Jet Multiplicity
Fat Jet Pt
Angle between Fat Jet and beam axis
W Mass
W Energy
W Pt
Angle between highest and mid E subjets

General Event Properties
Visible Energy
Visible Pt
Visible Pz
Y45 (switch from 4-5 jet topology)
Angular separation of lepton and hadronic Top

Variables chosen to maximise distinction of high energy event classifier. Still need to look at re-optimization for the low energy event classifier!!

Some subjet related variables have been removed while systematics are evaluated

1) Lepton Identification

1. Cluster all TightSelectedPFOs into 5 jets using ee_kt algorithm
2. Select TightPFOs with Pandora PID=11 or 13, E>10GeV as isolated lepton candidates
3. Calculate E_{PFO}/E_{Jet} for each candidate- provides an estimate for how isolated the candidate is
4. Select the PFO with the highest E_{PFO}/E_{Jet} ratio to be the isolated lepton
5. If no candidate is selected, repeat process without the inclusion of step 2

By construction, 1 lepton identified per event

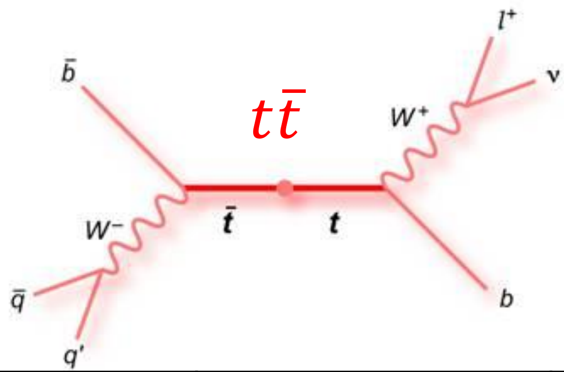
<u>Purity= 85%</u>

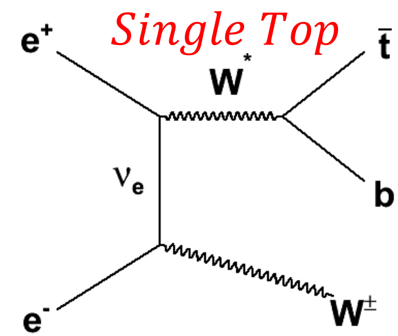
High purity <u>essential</u> for many applications

Samples used

$\sqrt{s}$=1.4TeV
P(e-) = -80% for signal

y=b,s,d
x=u,c
l=mu,tau





Process	ID	σ (fb)	Generated Events (k)
ee→yy veyx	6589	31.9	790
ee→yy xyev	6592	29.0	800
ee→yy vlyx	6634	40.7	1,140
ee→yy xylv	6637	40.7	1,130
ee→qqlv	3249	4309.7	980
ee→qqqq	2163	1328.1	300
ee→qq	2091	4009.5	460
ee→qqqqlv	2169	115.3	180
ee→qqqqll	2166	71.7	210
ee→qqqqvv	2152	24.7	230

Signal! (first four rows)