

2358

# Jet Physics and Substructure in ATLAS

......

Dresden, November 2019

Eva Hansen Lund University

## Outline



#### Introduction

- Quantum Chromo Dynamics
- Jets
- Pile-up
- Jet reconstruction
  - Measuring jets with ATLAS
  - Jet inputs
  - Jet Algorithms
- Jet calibration
  - Jet Energy Scale
  - Jet Mass Scale
- Jet Substructure
  - "Prong-like" variables
    - · Top-tagging
  - "Haze-like" variables
    - · Quark/gluon-tagging

### 11/7/19

## Outline



#### • Introduction

- Quantum Chromo Dynamics
- Jets
- Pile-up
- Jet reconstruction
  - Measuring jets with ATLAS
  - Jet inputs
  - Jet Algorithms
- Jet calibration
  - Jet Energy Scale
  - Jet Mass Scale
- Jet Substructure
  - "Prong-like" variables
    - Top-tagging
  - "Haze-like" variables
    - · Quark/gluon-tagging

#### 11/7/19



- In 1960 Particle Physics was a chaotic zoo of observations
  - Electrons, muons and neutrinos, called *leptons*
  - Protons, neutrons, and a plethora of other *hadrons*



Michael Riordan: The Hunting of the Quark



- In 1960 Particle Physics was a chaotic zoo of observations
  - Electrons, muons and neutrinos, called *leptons*
  - Protons, neutrons, and a plethora of other *hadrons*
- Murray Gell-Mann proposed that the hadrons consisted of tiny, fractionally charged subcomponents
  - Called the quirky little things quarks





- In 1960 Particle Physics was a chaotic zoo of observations
  - Electrons, muons and neutrinos, called *leptons*
  - Protons, neutrons, and a plethora of other *hadrons*
- Murray Gell-Mann proposed that the hadrons consisted of tiny, fractionally charged subcomponents
  - Called the quirky little things quarks
- The quark model had big implications:
  - Pauli exclusion principle demanded a new quantum number
    - Color charge
  - And a new force, holding the quarks together:
    - → The strong force carried by the gluon
    - Weaker at small distances (asymptotic freedom)
    - Stronger at large distances (confinement)

#### Proton:





- In 1960 Particle Physics was a chaotic zoo of observations
  - Electrons, muons and neutrinos, called *leptons*
  - Protons, neutrons, and a plethora of other *hadrons*
- Murray Gell-Mann proposed that the hadrons consisted of tiny, fractionally charged subcomponents
  - Called the quirky little things quarks
- The quark model had big implications:
  - Pauli exclusion principle demanded a new quantum number
    - Color charge
  - And a new force, holding the quarks together:
    - → The *strong force* carried by the *gluon*
    - Weaker at small distances (asymptotic freedom)
    - Stronger at large distances (confinement)
- The Standard Model of particle physics took form



# Jets: Showering and hadronisation



- QCD predicted a detectable signature that was crucial for establishing the theory: **Jets**!
- Asymptotic freedom: Quarks are ~free at small distances
  - Interact as individual particle at very high energy / short distance
  - Emit "Bremstrahlung" when accelerated in a hard scattering, forming a narrow shower of quarks and gluons
- Confinement: One can never observe a free quark
  - At distances of ~ 1 fm the quarks hadronise-
- First evidence in 1975 with the SPEAR collider at SLAC

# Jets: Showering and hadronisation



- QCD predicted a detectable signature that was crucial for establishing the theory: **Jets**!
- Asymptotic freedom: quarks are ~free at small distances
  - Interact as individual particle at very high energy / short distance
  - Emit "Bremstrahlung" like electrons when accelerated forming a narrow shower of quarks and gluons
- Confinement: one can never observe a free quark
  - At distances of  $\sim$  1 fm the quarks transform into hadrons
- First evidence in 1975 with the SPEAR collider at SLAC



## Jets: Pros and cons



- ✓ Useful probes of QCD at both soft and hard energy scales
- ✓ Probable final state for interesting processes at collider experiments
  - Higgs decay channels
  - New heavy particles in many SM extensions



## Jets: Pros and cons



- ✓ Useful probes of QCD at both soft and hard energy scales
- ✓ Probable final state for interesting processes at collider experiments
  - Higgs decay channels
  - New heavy particles in many SM extensions



# Pile-up at the LHC



- Very high energy means a ~1:1 correspondence between jet and origin particle
- Protons are collided in bunches every 25 ns to increase luminosity
  - Many collisions per bunch crossing  $\rightarrow$  (In-time) *Pile-up*
  - Energy deposits from previous/future bunch crossings → (Out-of-time) *Pile-up*
- Complicates event reconstruction and analyses

65 reconstructed vertices Tracks with  $p_{T}$  > 100 MeV



## Agenda



- Introduction
  - Quantum Chromo Dynamics
  - Jets
  - Pile-up
- Jet reconstruction
  - Measuring jets with ATLAS
  - Jet inputs
  - Jet Algorithms
- Jet calibration
  - Jet Energy Scale
  - Jet Mass Scale
- Jet Substructure
  - "Prong-like" variables
    - Top-tagging
  - "Haze-like" variables
    - · Quark/gluon-tagging

# Measuring Jets with ATLAS

- Different subdetectors allow us to identify and reconstruct most particles efficiently
- Calorimeters provide the principal signals for jet measurement
  - Full coverage and fine segmentation
- The inner detector provides precision  $p_{\rm T}$  and direction information of charged particles
  - Vertex reconstruction, pile-up mitigation, refinement of jet reconstruction





# Defining jets



Clearly 2 jets



How many do we see here?



# Inputs to Jet Algorithms



Two main input definitions used in ATLAS:



More jet inputs combining tracks and calorimeter cluster are being studied:

- Track Calo Cluster (TCC)
- Unified Flow Object (UFO)

# **Topological Clusters**

1) Clustering: Initialised by high energy seeds and expanded in two steps:



2) Origin correction: Modifies topocluster 4-momentum to point back at the primary vertex

- Improves  $\eta$ -resolution without changing the energy

3) Rescaling:

- *EM-scale*: All cell energies are weighted according to the electromagnetic scale calibration
- LCW: Topoclusters are weighted depending them being electromagnetic or hadronic due to lower response in hadronic calorimeter



## **Particle Flow**



#### Many benefits to combining information from trackers and calorimeter

- Tracking detectors:
  - Better resolution for low  $p_{T}$  particles
  - Better angular resolution
  - · Can trace particle to either the hard-scatter interaction or pile-up
- Calorimeters:
  - Better resolution for high  $p_{T}$  particles
  - Captures neutral particles

Rough sketch of the algorithm:

- 1) Select "high quality" tracks coming from the primary vertex  $p_{T} < 40 \text{ GeV}$
- 2) Match track to corresponding topocluster(s)
- 3) Subtract energy from the cluster depending on position and track  $p_{T}$
- 4) Selected tracks and remaining topoclusters constitute PFlow objects passed to the jet algorithm

# More PFlow

- Improved  $p_{\rm T}$  resolution
- Improved angular resolution
- Less pile-up contribution







11/7/19

### Eva Hansen

Better E<sup>miss</sup>!

# Jet Finding Algorithms

- Intuitive way: Define a cone of fixed size and sum up all momenta inside
- NB! Jet algorithms must be insensitive to arbitrarily soft and collinear splittings in order to make theoretical predictions we can compare to data!
- Sequential algorithms to the rescue!



• Generalised definition:

1) Define the two distances,  $p=\{-1, 0, 1\}$ :

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta_{ij}^2}{R^2}, \quad d_{iB} = k_{ti}^{2p}$$

2) If  $d_{ii}$  is smallest, combine *i* and *j* 

- 3) Else, declare *i* a jet and remove it
- 4) Repeat until no more particles remain
- Most popular is p = -1: the Anti-kt algorithm
  - Clusters hardest constituents first
  - Gives nearly conical jets
- *R* is the radius parameter
  - Typically R = 0.2, 0.4, 0.6, 1.0, 1.2





### 11/7/19

# Jet Finding Algorithms

- Intuitive way: Define a cone of fixed size and sum up all momenta inside
- NB! Jet algorithms must be insensitive to arbitrarily soft and collinear splittings in order to make theoretical predictions we can compare to data!

Eva Hansen

• Sequential algorithms to the rescue!



Generalised definition, p={-1, 0, 1}:
1) Define the two distances:

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta_{ij}^2}{R^2}, \quad d_{iB} = k_{ti}^{2p}$$

2) If  $d_{ii}$  is smallest, combine *i* and *j* 

3) Else, declare *i* a jet and remove it

- 4) Repeat until no more particles remain
- Most popular is p = -1: the Anti-kt algorithm
  - Clusters hardest constituents first
  - Gives nearly conical jets
- *R* is the radius parameter
  - Typically R = 0.2, 0.4, 0.6, 1.0, 1.2



## Outline



- Introduction
  - QCD
  - Jets
  - LHC
- Jet reconstruction
  - Measuring jets with ATLAS
  - Jet inputs
  - Jet Algorithms
- Jet calibration
  - Jet Energy Scale
  - Jet Mass Scale
- Jet Substructure
  - "Prong-like" variables
    - Top-tagging
  - "Haze-like" variables
    - · Quark/gluon-tagging

### 11/7/19

# Why calibrate?





To correct the translation of calorimeter signal to original parton for detector effects:

- Dead material
  - Energy deposited in non-sensitive regions of the detector
- Calorimeter non-compensation
  - Partial measurement of the energy deposited by hadrons
- Punch-through
  - Showers extending beyond the calorimeters

- Pile-up
  - Additional energy deposits from other particles
- Out-of-cone
  - Part of the particle shower not included in the jet cone
  - Worse for low  $p_{\rm T}$  jets because of magnetic field
- Energy deposits below noise threshold

### 11/7/19

# Jet Energy Scale



Focus of

- Calibrations are provided for several jet definitions
  - "Small-R jets": Anti-kt R=0.4, based on Particle Flow
  - "Large-R jets": Anti-kt R=1.0, based on Local Cell Weighting f this talk
  - "R-Scan jets": Anti-kt R=0.2 and 0.6 LCW jets
  - Heavy Ion Jets
- Calibration differs slightly for the different definition, but principles are the same:





- Pile-up subtraction done in two steps
  - Area based subtraction of the per-event pile-up contribution to the  $p_{T}$  of each jet
  - Residual  $N_{PV}$  and  $\mu$  based subtraction

$$p_{T}^{corr} = p_{T}^{reco} - \rho \times A + \alpha (N_{PV} - 1) - \beta \langle \mu \rangle$$



11/7/19

# Jet Energy Scale



- Grooming" techniques reduce the contribution of pile-up and soft/wide-angle emissions
- Improves the  $p_{T}$  and mass resolution
- Makes substructure variables less dependent on fragmentation
- Full calibration provided for trimmed jets
- Reclusters the R=1.0 jet into constituent subjets with  $R_{sub} = 0.2$



Recalculates the jet four-momentum from the remaining constituents •

#### 11/7/19

#### Jet Energy Scale Pile-up Mitigation MC-based Calibration Global Sequential Calibration Lunversity

- Energy response differs across η
  - Especially at boundaries between different calorimeter technologies and granularities
- Isolated reco jets are matched to truth jets and compared



Two/three corrections are applied

#### 1) Absolute JES correction

 Response: Mean of a Gaussian fit to E<sup>reco</sup>/E<sup>truth</sup>

#### 2) Jet η correction

- Response: η<sup>reco</sup>-η<sup>truth</sup>
- 3) Jet mass correction (just for large-R)
  - Response: m<sup>reco</sup>/m<sup>truth</sup>

#### Jet Energy Scale Pile-up MC-based Calibration Global Sequential

Calibration

- Only done for Small-R jets!
- The GSC is applied to adjust for:
  - Non-compensation: Difference in response to hadrons, leptons and photons
  - Flavor dependence: Difference in response to quarks and gluon
  - **Punch-through:** Jets extending beyond the calorimeters
- Calibration is done in five/six steps (LCW/PFlow)
- Uses observables related to
  - Energy deposits in the calorimeter
  - Track information of jets
  - Activity in the muon segments
- For each observable a 4-momentum correction is derived as a function of p<sub>T</sub><sup>truth</sup> and |η|
- Does not change the average energy



### 11/7/19

Eva Hansen

28

UNIVERSITY



- Last step is the residual in-situ calibration
  - Corrects for potential differences between data and MC
  - Applied only to data
- The in-situ methods rely on a well-calibrated reference object in the event to constrain the true jet  $p_{\rm T}$

$$Response = R = \left\langle p_T^{jet} / p_T^{ref} \right\rangle$$



• Consists of a set of sub-steps:





# Jet Mass Scale

Pile-up

Correction

LUNUVER

In-situ

Validation

- Two in-situ methods are employed to correct the calorimeter mass response
- Forward folding
  - + Uses  $t\bar{t}$  events with hadronically decaying boosted Ws and tops
  - Fits the mass peaks and jet mass response of the W and top
- The R<sub>trk</sub> method
  - · Uses track jets to provide an independent measurement of the jet mass scale

Global

Sequential

Calibration

• The combination is done separately for each mass bin



## Outline

- Introduction
  - Quantum Chromo Dynamics
  - Jets
  - Pile-up
- Jet reconstruction
  - Measuring jets with ATLAS
  - Jet inputs
  - Jet Algorithms
- Jet calibration
  - Jet Energy Scale
  - Jet Mass Scale

#### • Jet Substructure

- Jet mass
- "Prong-like" variables
  - · Top-tagging
- "Haze-like" variables
  - · Quark/gluon-tagging





# Why study substructure?



- To identify what kind of particle initiated the jet
  - Light quark, gluon, or something heavy?
- Measuring heavy SM particles (W/Z/top/H) as well as potential new heavy resonances is central for big parts of the ATLAS physics program
- At LHC energies, heavy particles are often produced with a large Lorentz boost
  - Leads to collimated decay products
  - Visible by the internal structure of jets
- Three main substructure variables: Mass, "Prong-ness" and "Hazy-ness"





Particle decaying at rest

Boosted particle decay



### 11/7/19

# Combined Jet Mass

- Mass is the ID-card of particles
- Measuring jet mass requires granularity finer than the size of the jet
  - Depend on both energy and opening angle between decay products
- Two definitions are used
  - Calorimeter Mass:

$$m^{\text{calo}} = \sqrt{\left(\sum_{i \in J} E_i\right)^2 - \left(\sum_{i \in J} \vec{p_i}\right)^2}$$

Track-Assisted Mass:

$$m^{\mathrm{TA}} = \frac{p_{\mathrm{T}}^{\mathrm{calo}}}{p_{\mathrm{T}}^{\mathrm{track}}} \times m^{\mathrm{track}}$$

 Best performance is obtained from a linear combination:

$$m^{\text{comb}} = w^{\text{calo}} \times m^{\text{calo}} + w^{\text{TA}} \times m^{\text{TA}}$$

the formulation preliminary  

$$\sqrt{s} = 13 \text{ TeV}, \text{WZ} \rightarrow qqqq$$
  
 $0.25$   
 $1000 \text{ Trimmed} (f_{cut} = 0.05, \text{R}_{sub} = 0.2)$   
 $1000 \text{ LCW} + \text{JES} + \text{JMS calibrated}$   
 $0.15$   
 $0.15$   
 $0.15$   
 $0.15$   
 $0.05$   
 $0.05$   
 $0.05$   
 $0.00$   
 $1000$   
 $1500$   
 $1000$   
 $1500$   
 $2000$   
 $2500$   
Truth jet p\_ [GeV]



ATLAS-CONF-2016-035

#### Eva Hansen

34

# "Prong-like" Variables

- Several options out there
- N-subjettiness:
  - Define a variable that quantifies how well the jet is described by N subjets:

$$\tau_N = \frac{1}{d_0} \sum_k p_{\mathrm{T}k} \times \min(\delta R_{1k}, \delta R_{2k}, ..., \delta R_{Nk}), \text{ with } d_0 \equiv \sum_k p_{\mathrm{T}k} \times R$$

- Typically use the ratio  $\tau_{N,N-1} = \tau_N / \tau_{N-1}$  for tagging a jet as "*N*-prong"
- +  $\tau_{_{32}}$  found to perform best for top tagging
- Energy correlation ratios:
  - Takes ratios and double ratios of energy correlation functions:

$$\operatorname{ECF}(N,\beta) = \sum_{i_1 < i_2 < \dots < i_N \in J} \left(\prod_{a=1}^N p_{T_{i_a}}\right) \left(\prod_{b=1}^{N-1} \prod_{c=b+1}^N R_{i_b i_c}\right)^{\beta}$$

Found to perform best for W tagging

35





#### Eva Hansen

#### 11/7/19

# Top tagging



- Wishes: Discrimination, stability against pile-up, and understood systematics
- Simple cut on tau32 and combined mass give good overall performance
- Still be something to gain with more complex multivariate techniques
- ATLAS now has a new Neural Network-based tagger

Observable	Variable	Used for	Defense	1 <u> </u>	
		Used Ioi	References		-+ 2-var optimised <b>AILAS</b> Simulation
Calibrated jet kinematics	$p_{\mathrm{T}}, m^{\mathrm{comb}}$	top,W	[44]		tagger $s = 13 \text{ IeV}$ Trimmed anti-k, R = 1.0 jets
Energy correlation ratios	$e_3, C_2, D_2$	top,W	[50, 54]	<u> </u>	$50 \rightarrow BDT \text{ top}$ $ \eta^{\text{true}}  < 2.0$
N-subjettiness	$\tau_1, \tau_2, \tau_{21}$	top, W	[55, 56]	ion	I op tagging at $\epsilon_{sig} = 80\%$
W-Subjettimess	$ au_{3},  au_{32}$	top		ect	40
Fox–Wolfram moment	$R_2^{\rm FW}$	W	[57, 58]	Lej.	-
Splitting measures	Z <sub>cut</sub>	W	[59, 60]	ק	30
	$\sqrt{d_{12}}$	top, W		Ino	
	$\sqrt{d_{23}}$	top		J	20-
Planar flow	${\cal P}$	W	[ <mark>61</mark> ]	ach	
Angularity	<i>a</i> <sub>3</sub>	W	[62]	Ĕ	
Aplanarity	A	W	[58]		
KtDR	KtDR	W	[63]		
Qw	$Q_w$	top	[59]		

 $p_{\tau}^{\text{true}}$  [GeV]

11/7/19

Eva Hansen

CERN-EP-2018-192

## Example of top-tagging use: tt resonance search





- Search for new heavy particles decaying to top pair
- Looking for deviations in the invariant mass spectrum of the tt system
- Using events where both tops decay hadronically  $(t \rightarrow Wb \rightarrow qqb)$
- Different search strategies used to target different resonance mass ranges
  - M < 1.2 TeV: Top decay products are resolved</li>
  - M > 1.2 TeV: Top is boosted and the decays merge into a single jet
- For the "boosted" analysis tops are tagged with straight cuts on the jet mass and  $\tau_{_{32}}$

CERN-EP-2018-350 4.5 4.5 7 = 13 TeV 2 = 13 TeV  $2 = 13 \text$ 

## "Haze-like" Variables



- Used to characterise radiation pattern when not interested in the number of prongs
- Popular haze-variables include
  - Number of constituents
    - Often approximated by the track multiplicity  $n_{trk}$
  - Width of the jet
    - Often defined by the sum of distances between tracks and jet axis weighted by  $p_{_{\rm T}}$

$$w_{track} = \frac{\sum_{i} p_{T}^{i} \Delta R(i, jet)}{\sum_{i} p_{T}^{i}}$$

# Quark/gluon tagging





- Gluon jets tend to be broader and have more constituents
- Track multiplicity  $n_{\rm trk}$  is strongest discriminating variable
- Challenges to quark/gluon tagging:
  - 1) No universal way of truth labeling in Monte Carlo
  - 2)  $n_{\rm trk}$  is sensitive to fragmentation modeling
  - 3) Quark and gluon jets are rather alike...

### Eva Hansen

n<sub>track</sub>

# Quark/gluon tagging



Current recommendation based only on n<sub>trk</sub>

11/7/19

- Data-driven technique used to estimate uncertainty
- For a given  $p_{\rm T}$ ,  $n_{\rm trk}$  does not depend on eta, but the probability of a jet being q or g does

$$\langle n_{\text{charged}}^{\text{f}} \rangle = f_{\text{q}}^{\text{f}} \langle n_{\text{charged}}^{q} \rangle + f_{\text{g}}^{\text{f}} \langle n_{\text{charged}}^{g} \rangle$$

$$\langle n_{\text{charged}}^{\text{c}} \rangle = f_{\text{q}}^{\text{c}} \langle n_{\text{charged}}^{q} \rangle + f_{\text{g}}^{\text{c}} \langle n_{\text{charged}}^{g} \rangle,$$



## Example of quark/gluon-tagging use: Vector-boson fusion Higgs



- A Higgs produced via VBF is accompanied by two light-flavor quarks
- Background processes are more rich on gluon jets
- Select events with four jets of which two are b-tagged
- *N*<sub>trk</sub> is used as an input variable in a BDT to discriminate signal from background events
- The uncertainty on  $n_{trk}$  is propagated through to the limit setting





## One more example: Dark QCD-like sectors

- QCD-like hidden sector models can lead to jets with substructure than SM jets
- Composition of **visible** and **invisible** partons in the jet dependent on parameter choice:
  - Exotic I: Displaced vertices, emerging jets
  - Exotic II: Semi-visible jets
  - We target SM QCD-like models
    - With s-channel mediator decaying to two dark quarks
- Four models implemented in Pythia Hidden Valley process
  - All have larger confinement scales than SM QCD!
    - → Many more constituents!
  - Based on <u>arXiv:1712.09279</u>
- Strategy:
  - Select dijet events using substructure variables
  - Look for a bump in the dijet invariant mass spectrum





#### 11/7/19

#### Eva Hansen

42

## Conclusions



- Jets are
  - abundant in LHC experiments
  - interesting for both QCD studies and new physics
  - challenging because of large backgrounds and pile-up
- In the high-pile-up era we are entering, there is a lot to gain from combining track information with calorimeter signal
- Though the topic of jet substructure has existed for a long time, it is still a very vigorous field of study, which will only be more important as colliders go to higher energies

# Backup





## **Calibration Chain: Global Sequential Calibration**



- Observables are related to
  - Energy deposits in the calorimeter / Non-compensation
  - Track information of jets / Flavor dependence
  - Activity in the muon segments / Punch-through



### **Calibration Chain: Global Sequential Calibration**



### Small-R JER

• The Jet Energy Resolution (JER) is the width of the response distribution in a given bin



Parameterised as



- N: Pile-up and electronic noise
- S: Statistical Poisson fluctuations
- **C**: Signal loss in passive material
- Noise term constrained using Random Cone Method and a μ=0 MC sample
- Other terms obtained by fitting in-situ measurements from dijet (and potentially Z/γ+jet) events with N held fixed

### Small-R JER

• Noise term includes pile-up and electronic noise



- N<sub>pile-up</sub> is derived with the Random Cone Method:
  - Construct two random cones in zero-bias data sample
  - Sum energy clusters within the two cones
  - Fluctuations due to pile-up are taken as the width of the  $p_{\rm T}$  difference distribution
- $N_{\mu=0}$  is derived from a MC sample with no pile-up
- Dijet method:
  - Similar to the  $\eta$ -intercalibration
  - JER is the width of the asymmetry distribution





#### Sep 11, 2018

#### Eva Hansen, HCW 2018

### Small-R JER

- Fit performed to dijet measurements with constraint on noise term from Random Cones method
- JER measurements in Z/γ+jet events may be included to span more phase space
- Brand new recommendations out now



