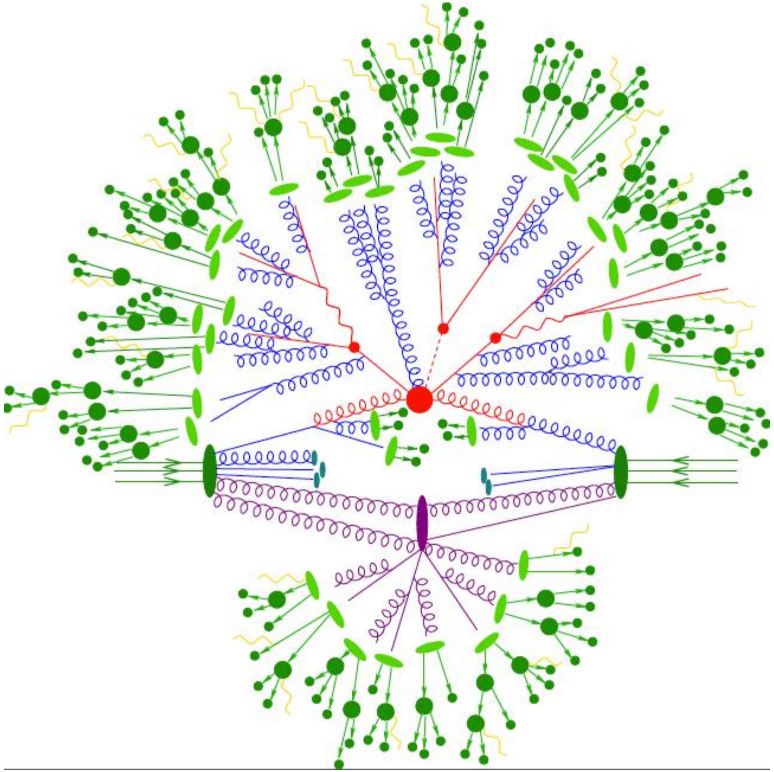
Introduction to Jet Substructure

Frédéric Dreyer and Steven Schramm

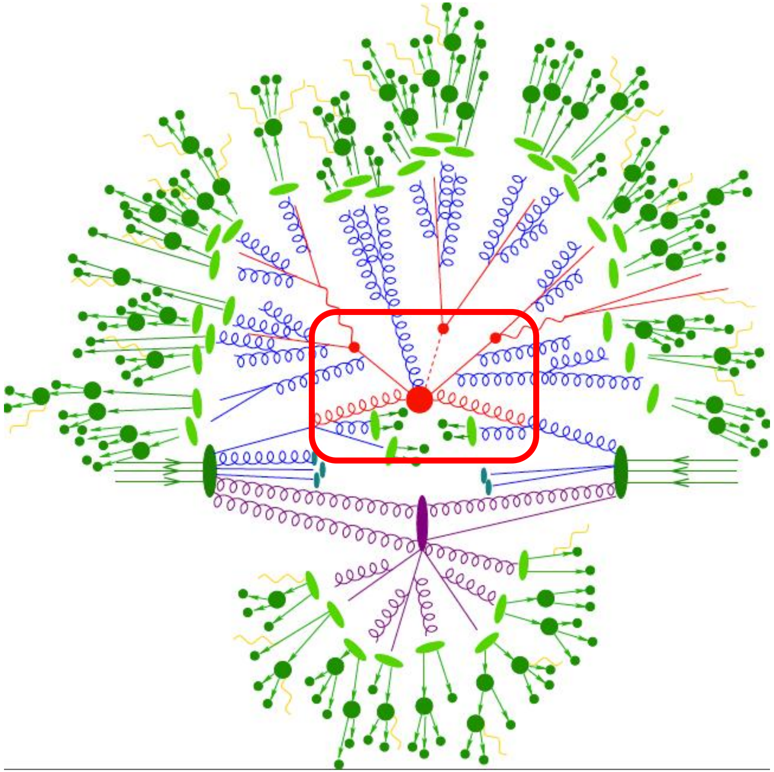
October 22, 2018

What is a jet?

Schematics on underlying event, partons, etc

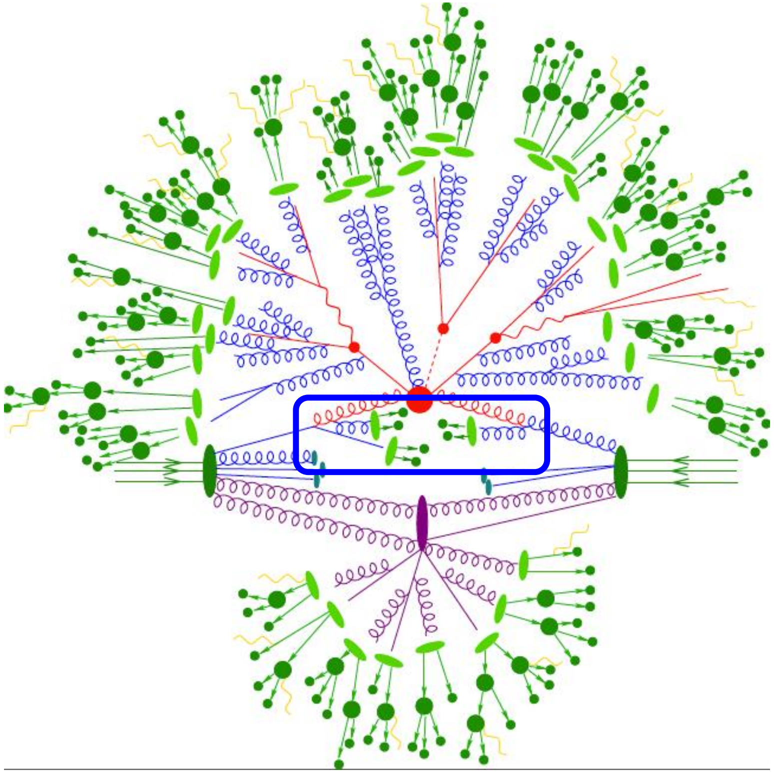


Schematics on underlying event, partons, etc



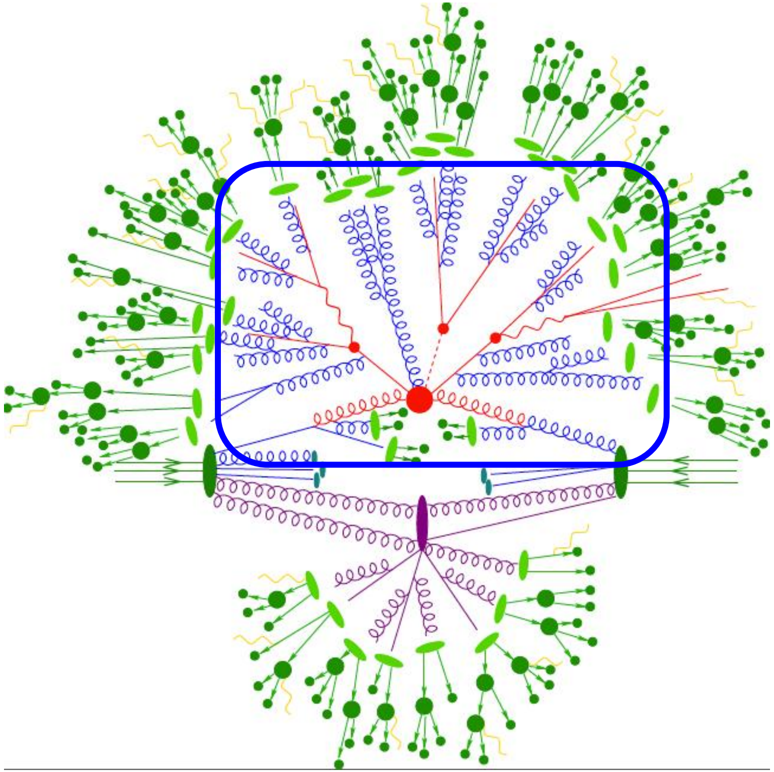
- Hard scattering : g g →W q q

Schematics on underlying event, partons, etc



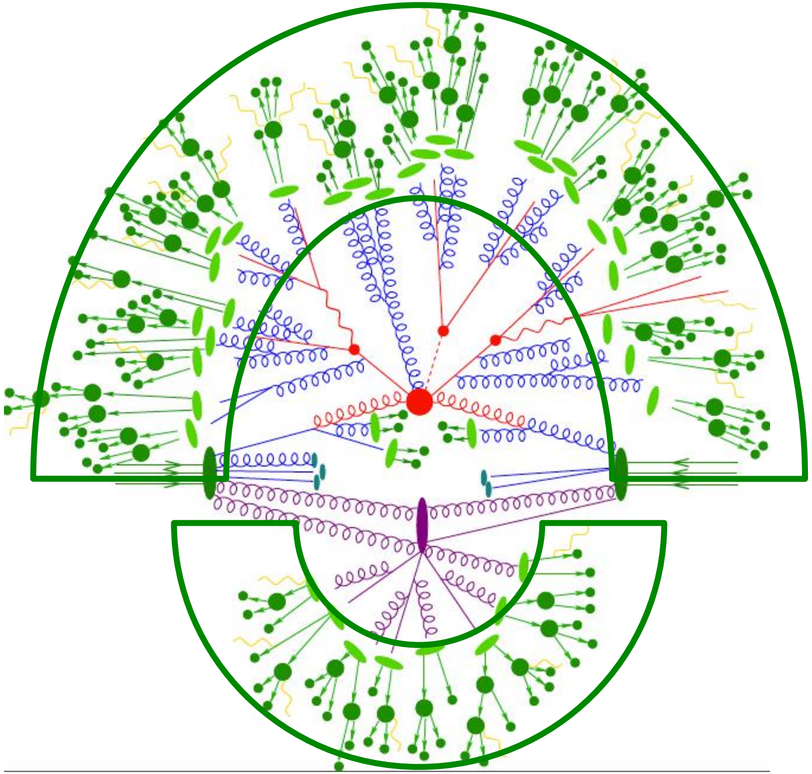
- Hard scattering
- Initial state radiation : emitted by incoming partons before hard scattering

Schematics on underlying event, partons, etc



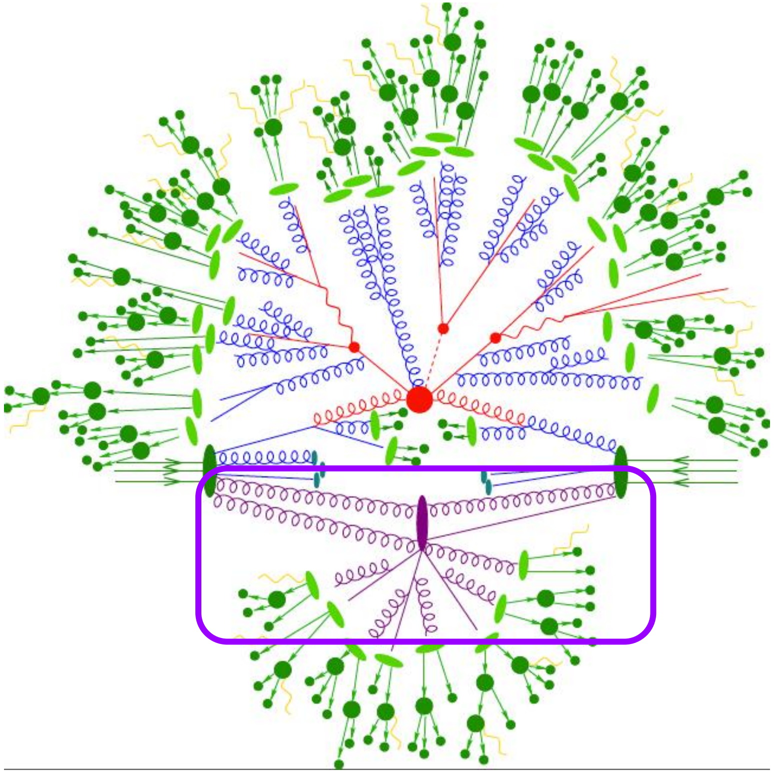
- Hard scattering
- Initial state radiation
- Parton shower : cascades of radiation from QCD decays

Schematics on underlying event, partons, etc



- Hard scattering
- Initial state radiation
- Parton shower
- Hadronization : partons form hadrons

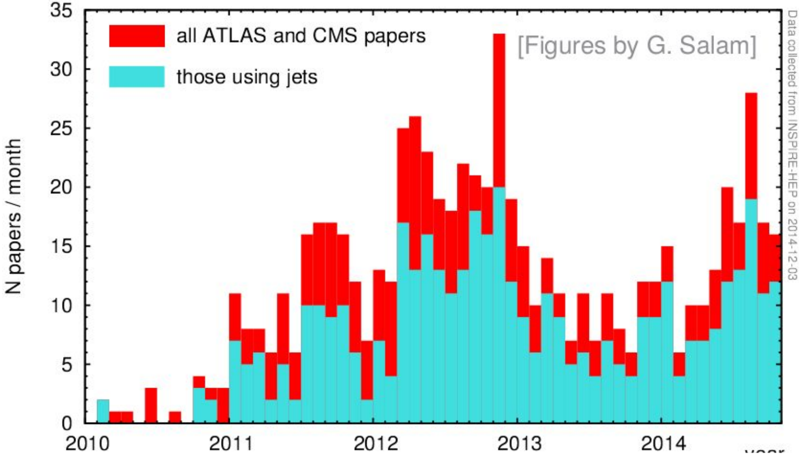
Schematics on underlying event, partons, etc



- Hard scattering
- Initial state radiation
- Parton shower
- Hadronization
- Multi-parton interactions (MPI) : additional interactions from the other partons forming the proton

Jets

- Because of color confinement, quarks and gluons shower and hadronise immediately into collimated bunches of particles.
- Hadronic jets emerge from a number of processes:
    - Scattering of partons inside colliding protons
    - Hadronic decay of heavy particles
    - Radiative gluon emission from partons
- Jets are prevalent at hadron colliders, and used in 2/3 of ATLAS and CMS analyses



Jet algorithm

A jet algorithm maps final state **particle momenta** to **jet momenta**.

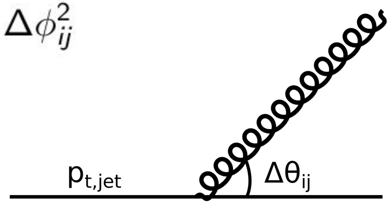
$$\underbrace{\{p_i\}}_{\text{particles}} \Longrightarrow \underbrace{\{j_k\}}_{\text{jets}}$$

Requires an external parameter, the **jet radius R**, which specifies up to which angle separate partons are recombined into a single jet.

Recursively cluster particles that are closest in a metric defined by

$$d_{ij} = \min(p_{t,i}^{2p}, p_{t,j}^{2p}) \frac{\Delta\theta_{ij}^2}{R^2}$$

$$\Delta\theta_{ij}^2 = \Delta y_{ij}^2 + \Delta\phi_{ij}^2$$

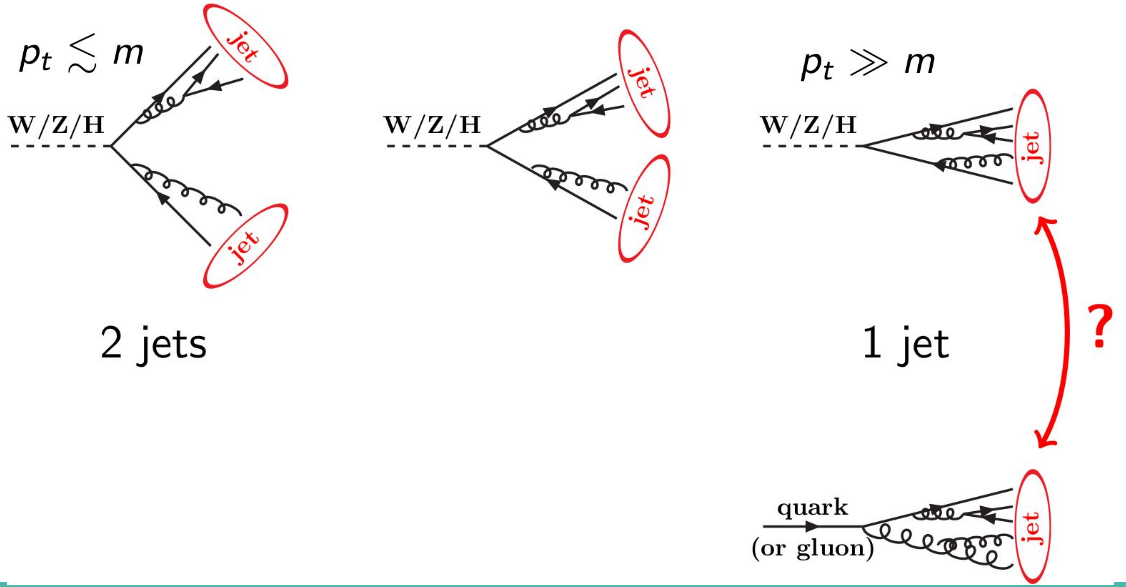


Jet radius

- The radius parameter in the equation roughly controls the size of the jet
  - Complementary information to the algorithm
- Typical choice for small-R jets: R=0.4 (ATLAS and CMS)
  - Used for “standard” QCD jets by most experimental analyses
  - Idea is roughly to contain a light quark or gluon in such a jet
- Typical choice for large-R jets: R=0.8 (CMS) or R=1.0 (ATLAS)
  - Used for “boosted” jets by most experimental analyses
  - Idea is roughly to contain a hadronically decaying particle (W, Z, H, top, etc)

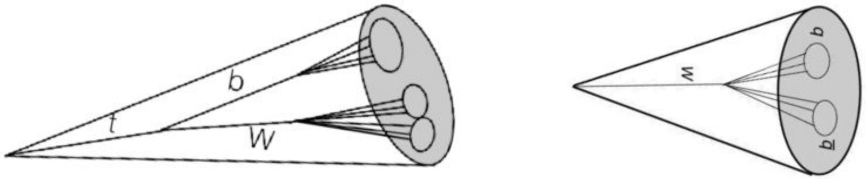
Boosted objects at the LHC

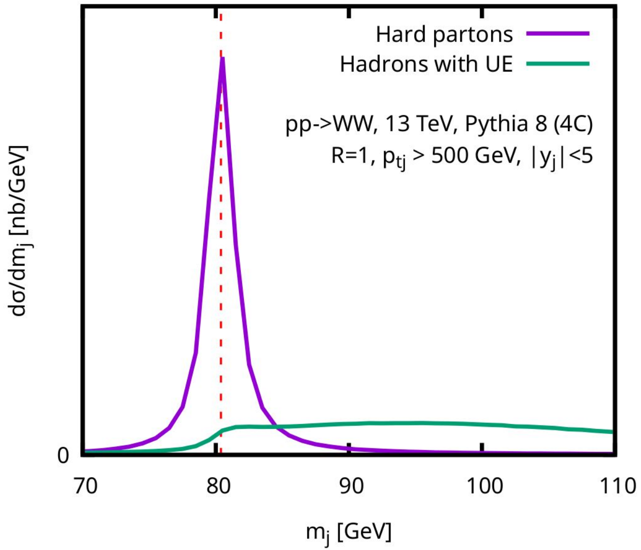
- At LHC energies, EW-scale particles (W/Z/t. . .) are often produced with $p_t \gg m$, leading to collimated decays.
- Hadronic decay products are thus often reconstructed into single jets.



Boosted objects at the LHC

- Many techniques developed to identify hard structure of a jet based on radiation patterns.
- In principle, simplest way to identify these boosted objects is by looking at the mass of the jet.
- But jet mass distribution is highly distorted by QCD radiation and pileup.

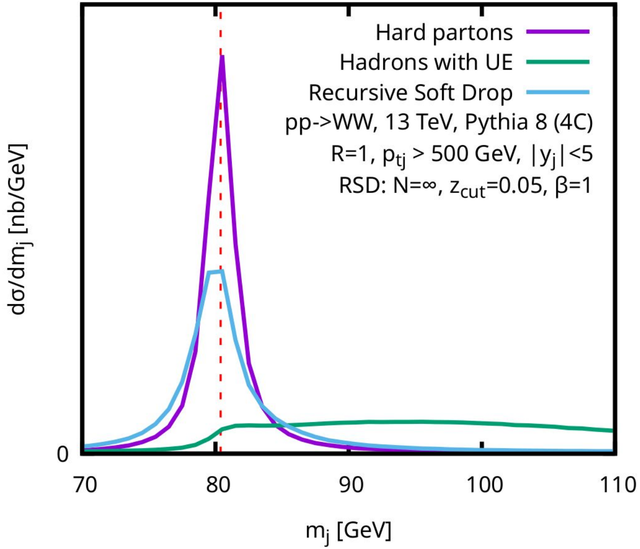




Jet grooming: (Recursive) Soft Drop/mMDT

- Mass peak can be partly reconstructed by removing unassociated soft wide-angle radiation (grooming).
- Recurse through clustering tree and remove soft branch if

$$\frac{\min(p_{t,1}, p_{t,2})}{p_{t,1} + p_{t,2}} > z_{\rm cut} \left(\frac{\Delta R_{12}}{R_0} \right)^{\beta}$$



[Dasgupta, Fregoso, Marzani, Salam JHEP 1309 (2013) 029]
[Larkoski, Marzani, Soyez, Thaler JHEP 1405 (2014) 146]
[Dreyer, Necib, Soyez, Thaler JHEP 1806 (2018) 093]

Jet grooming: common tools

Trimming

- Take jet with radius R
- Reclusters components into smaller subjets with radius $R_{sub} < R$
- Keep subjets that satisfy $p_{t,\,sub} > z_{cut}\, p_{t,\,jet}$

[Krohn, Thaler, Wang, JHEP 1002 (2010) 084]

Pruning

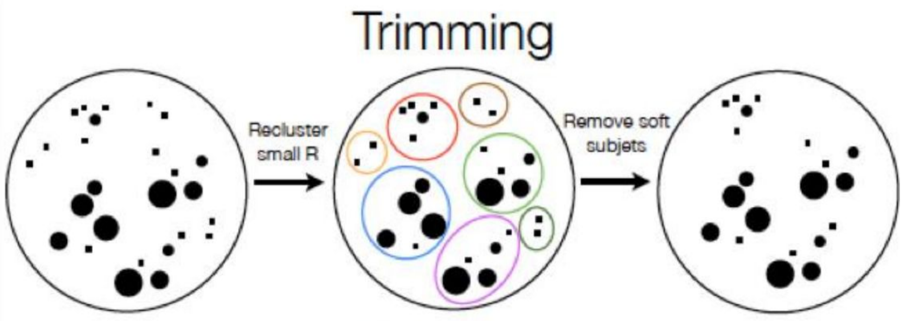
- Define pruning radius $R_{prun} = R_{cut}\;\; 2\,m\,/\,p_t$
- For every step of clustering $j_1 + j_2 \rightarrow j_{12}$, check:
  - Wide-angle: $\Delta R_{12} > R_{prun}$
  - Soft: $\min(p_{t1}, p_{t2}) < z_{cut}\, p_{t,\,jet}$
- If either condition fails, eliminates softer subjet
- If both pass, continue clustering

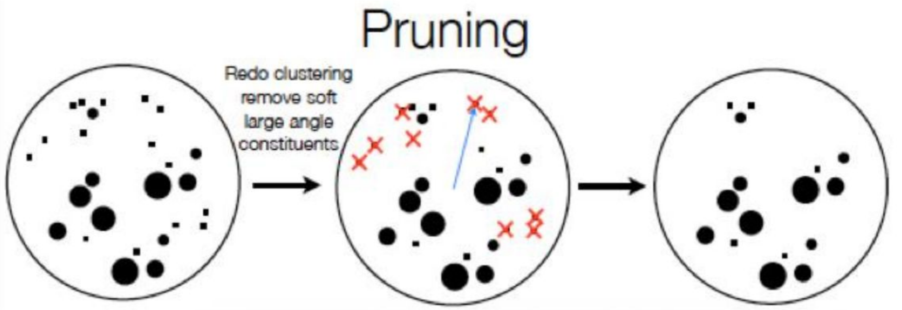
[Ellis, Vermilion, Walsh, PRD81 (2010) 094023]

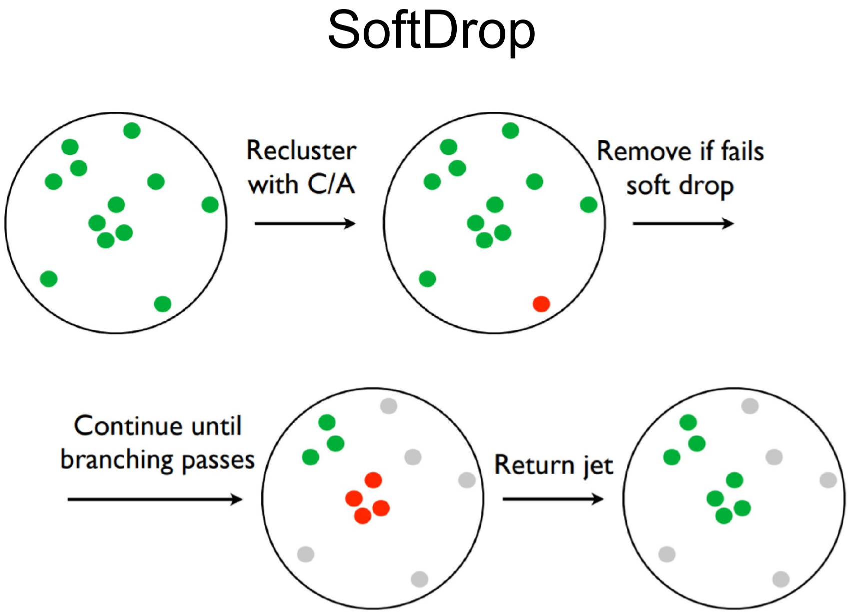
(Recursive) Soft Drop / mMDT

- Decluster jet $j_{12} \rightarrow j_1 + j_2$
- Check condition $\min(p_{t1}, p_{t2})/p_{t,jet} > z_{cut}(\Delta R_{12}/R)^{\beta}$
  - z_{cut}, β: tunable values
- If condition fails, the softer subjet is removed
- If passes, stops recursion
- For β=0, it is mMDT

[Dasgupta, Fregoso, Marzani, Salam JHEP 1309 (2013) 029]
[Larkoski, Marzani, Soyez, Thaler JHEP 1405 (2014) 146]
[Dreyer, Necib, Soyez, Thaler JHEP 1806 (2018) 093]

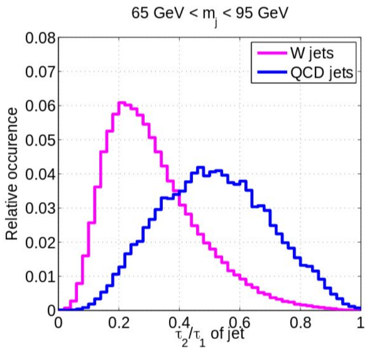

Trimming
Recluster small R
Remove soft subjets

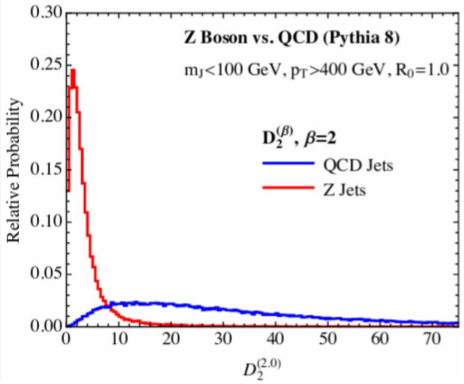

Pruning
Redo clustering remove soft large angle constituents


SoftDrop
Recluster with C/A
Remove if fails soft drop
Continue until branching passes
Return jet

Identifying jets with substructure observables

- Variety of observables have been constructed to probe the hard substructure of a jet (V/H/t decay lead to jets with multiple hard cores).
- Radiation patterns of colourless objects (W/Z/H) differs from quark or gluon jets.
- Efficient discriminators can be obtained e.g. from ratio of N -subjettiness or energy correlation functions.



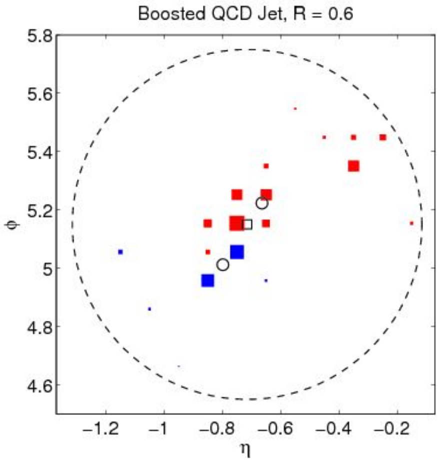


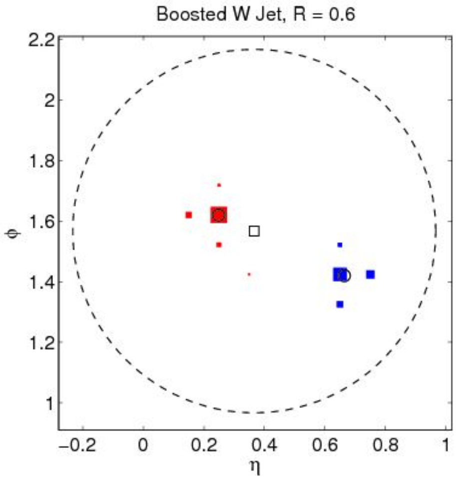
Jet shapes : N-subjettiness

- Measures radiation around N pre-defined axis

$$\tau_N^{(2)} = \frac{1}{p_{t,jet} R^2} \sum_{i \in jet} p_{t,i} \min_{a_i \dots a_N} \left(\theta_{ia_1}^2, \dots, \theta_{ia_N}^2\right).$$

- Use $\tau_{21} = \tau_2/\tau_1$ for 2-pronged jets and $\tau_{32} = \tau_3/\tau_2$ for 3-pronged jets





[Larkoski, Salam, Thaler, JHEP 1306 (2013) 108]
[Larkoski, Moult, Neill, JHEP 1412 (2014) 009]

Jet shapes : Energy correlation functions

- Similar to N-subjettiness (measures “dispersion”)
- Advantage of not needing pre-defined axis ($z_i = p_{t,i} / p_t$, the momentum fraction)

$$e_2^{(\beta)} = \sum_{1 \leq i < j \leq n_J} z_i z_j \theta_{ij}^{\beta}$$

$$e_3^{(\beta)} = \sum_{1 \leq i < j < k \leq n_J} z_i z_j z_k \theta_{ij}^{\beta} \theta_{ik}^{\beta} \theta_{jk}^{\beta}$$

- For 2-pronged jets

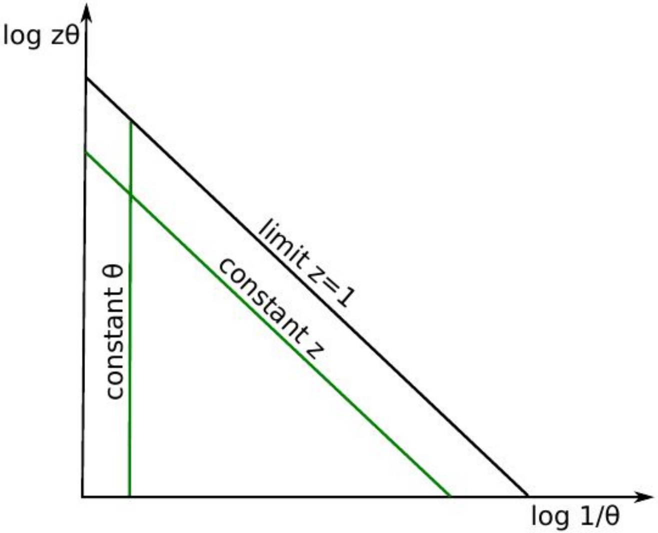
$$D_2^{(\beta)} = \frac{e_3^{(\beta)}}{(e_2^{(\beta)})^3}$$

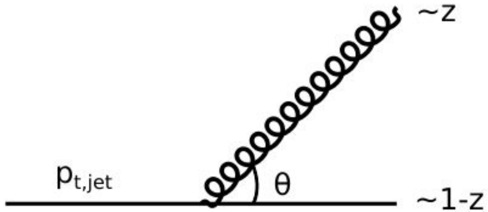
- For 3-pronged jets

$$C_3^{(\beta)} = \frac{e_4^{(\beta)} e_2^{(\beta)}}{(e_3^{(\beta)})^2}$$

Lund diagram

- Graphical representation of emissions in zθ vs. 1/θ coordinates.
- Used to illustrate branching phase space in parton shower Monte Carlo simulations and in perturbative QCD resummations.

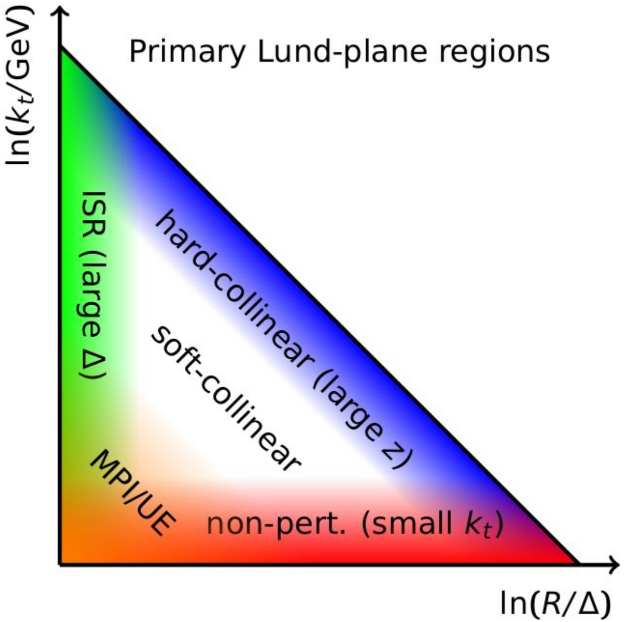


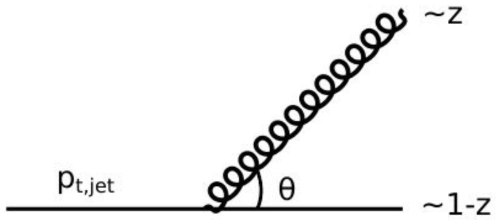


$$dP_{\mathrm{em}} \sim \frac{\alpha_s C_R}{\pi} \frac{d\theta^2}{\theta^2} dz\, p_i(z)$$

Lund diagram

- Soft-collinear emissions are emitted uniformly in the Lund plane
- Different kinematic regimes are clearly separated

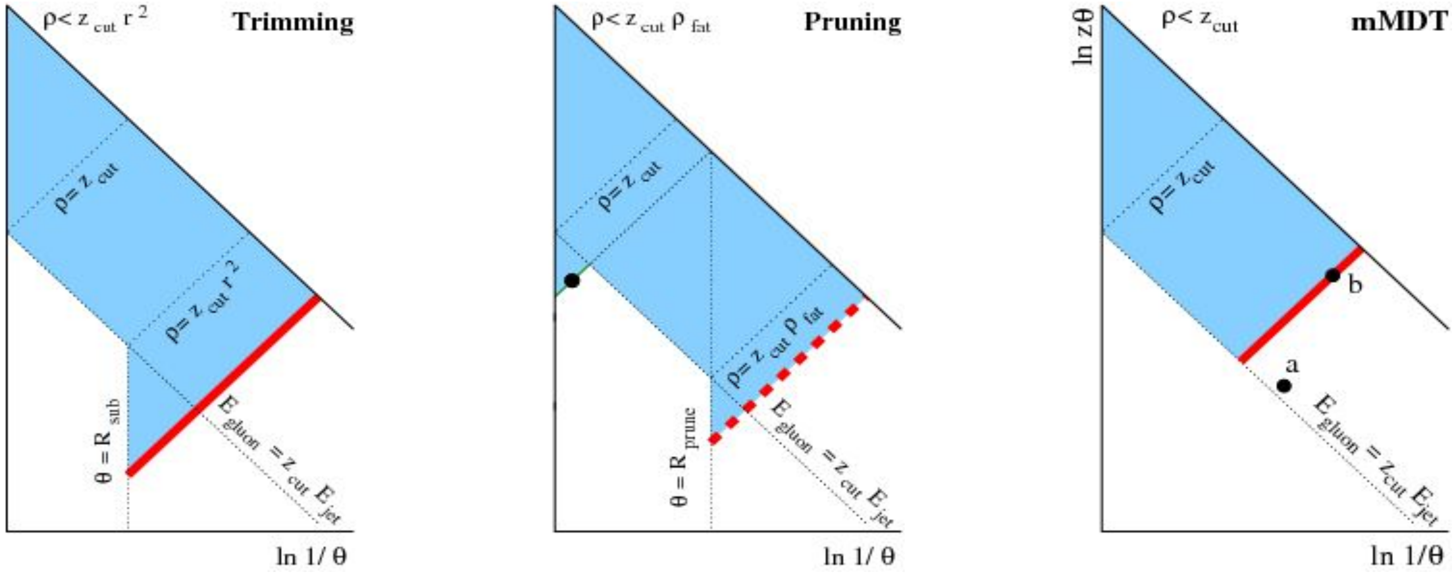




$$dP_{\rm em} \sim \frac{\alpha_s C_R}{\pi} \frac{d\theta^2}{\theta^2} dz\, p_i(z)$$

How grooming works

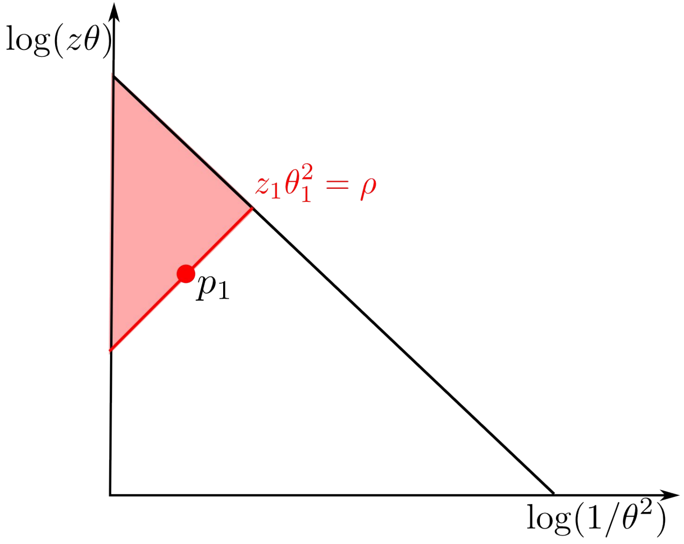
- Grooming eliminates kinematic region dominated by NP effects



[Dasgupta, Fregoso, Marzani, Salam JHEP 1309 (2013) 029]

Jet mass in the Lund plane

- For a jet with mass m, no real emissions allowed with $z\theta^2 > \rho$
- Virtual and real emissions cancel out, except in shaded region
- All-order resummation gives an exponential factor at LL



$$\Sigma(\rho) = 1 - e^{-R(\rho)}$$

$$\Sigma(\rho) = 1 - \frac{\alpha_s C_F}{\pi} \exp\left[-\frac{\alpha_s C_F}{2\pi} \log\left(\frac{1}{\rho}\right)^2\right]$$

Groomed jet mass

- Connection between **measurements** and **calculations**
- **Jet mass** is one of the simplest observables
- **Grooming** eliminates part of UE contamination
- Studied modified MassDrop Tagger and SoftDrop
- **Needs to be resummed at all orders, matched to fixed-order**

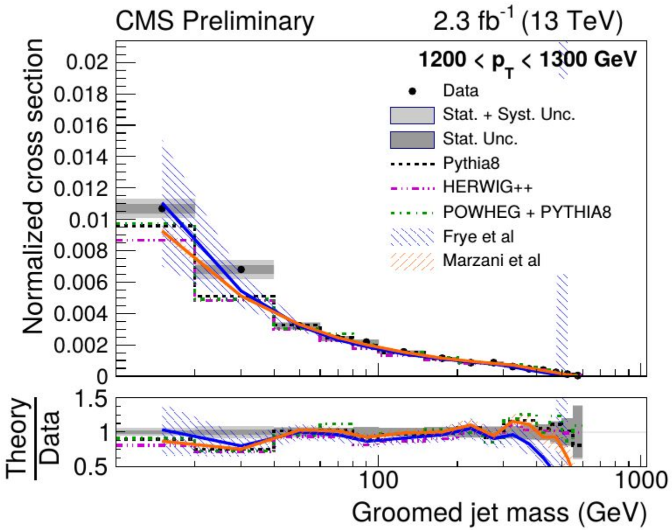
Groomed jet mass

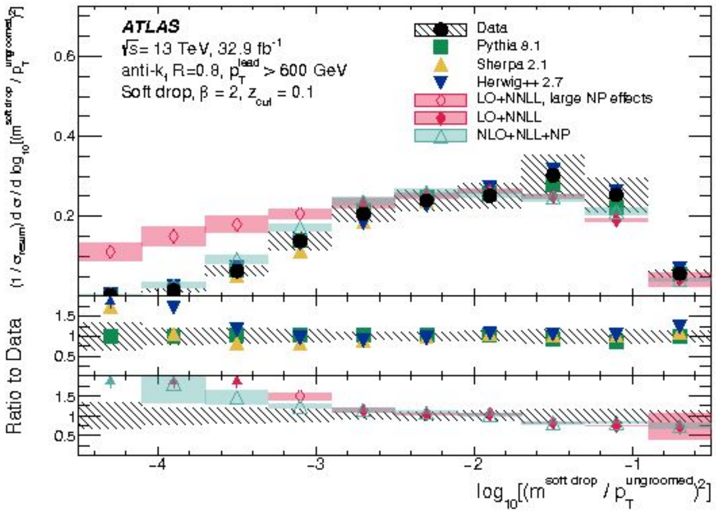
- Various interesting QCD structures emerging
  - For mMDT it becomes $[\alpha_s\, f\,(z_{cut})\, \log(1/\rho)]^n$ at leading-log
  - Finite z_{cut} introduce a flavour changing matrix structure
- Compare with experiment → needs a matching procedure:

Resummation of large logs ⟵ $\underbrace{N^kLL}_{\text{small } \rho} + \underbrace{N^mLO}_{\text{large } \rho}$ ⟶ **Fixed order** at ~$O(\alpha_s)$

- Calculations done with different theoretical approaches
  - NLL + LO for $z_{cut} \ll 1$ — Frye, Larkoski, Schwartz, Yan (2016)
  - LL + NLO for all z_{cut} — Marzani, Soyez, Schunk (2017)
  - Inclusive jets version — Kang, Lee, Liu, Ringer (2018)

Groomed jet mass





CMS measurements with mMDT
CMS-PAS-SMP-16-010

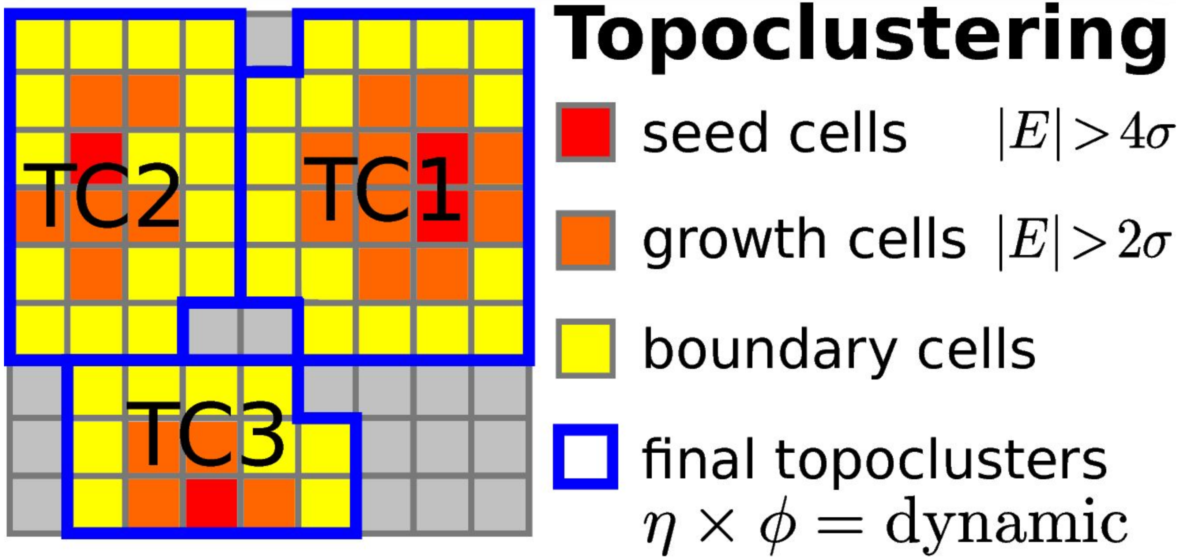
ATLAS measurements with SoftDrop
CERN-EP-2017-231

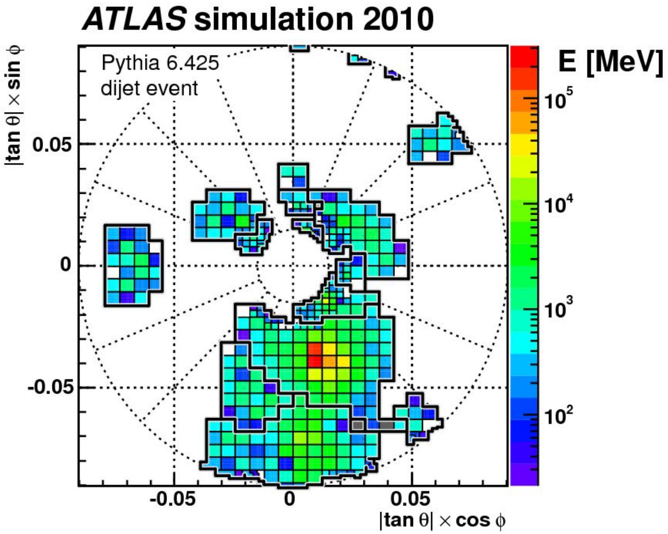
Experimental considerations

Experimental inputs to jet reconstruction

- Calorimeter objects
  - Calorimeters measure the energy of hadronic particles, both charged and neutral
  - **Typical input to ATLAS jet reconstruction**
- Tracks
  - Tracking detectors measure the momentum of charged particles (not neutral)
  - Tracks can typically be traced back to their independent vertex
  - **Typically used to augment jets, not for the jet four-vector**
- Particle flow objects
  - Combinations of calorimeter and tracking information to benefit from both detectors
  - Ideally the best aspects of both tracks and calorimeter objects
  - **Typical input to CMS jet reconstruction, increasingly used in ATLAS**

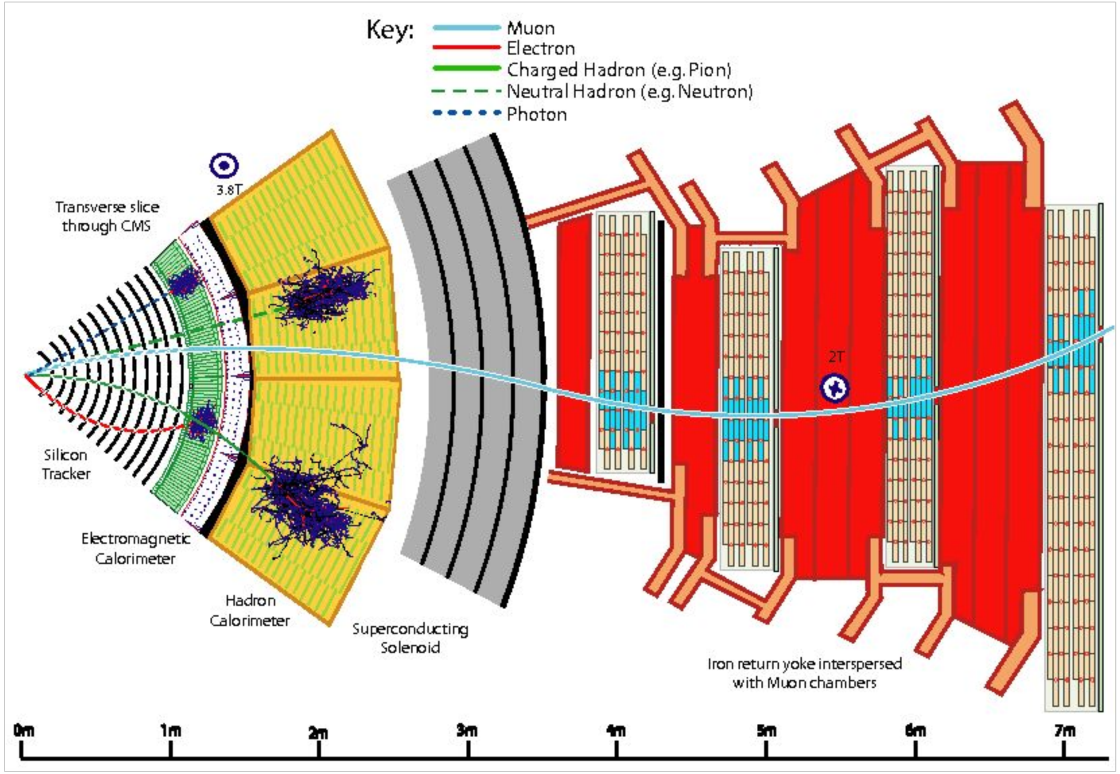
ATLAS: topological calorimeter-cell clusters





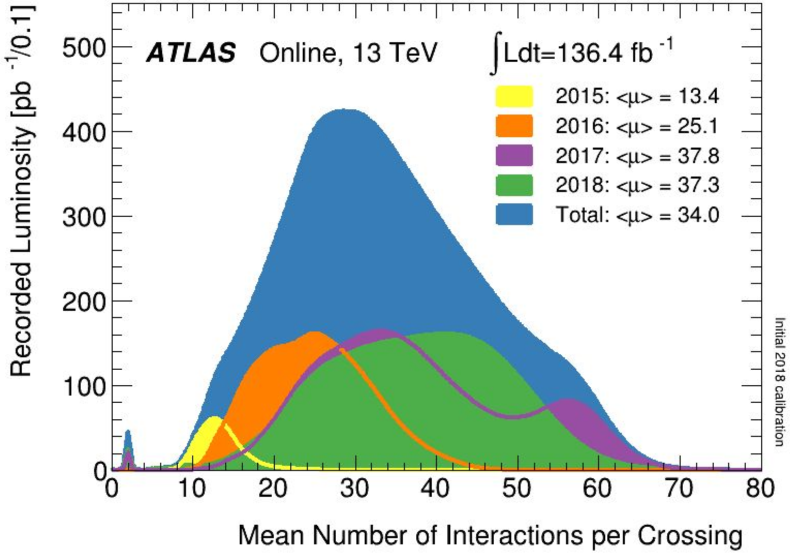
CMS: particle flow

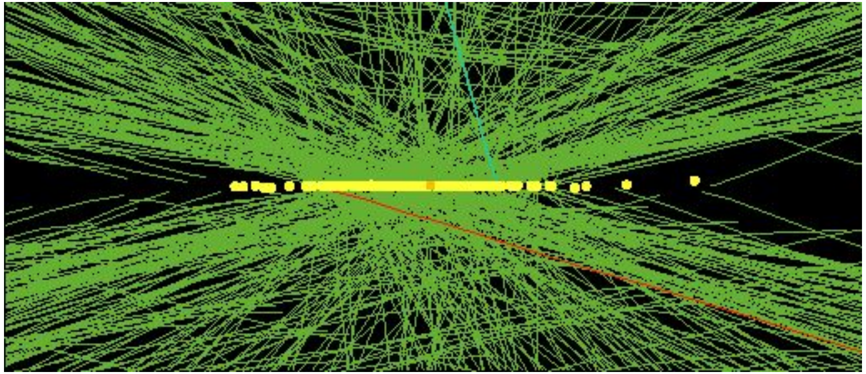
- Match inputs from different detectors and build a coherent single object



Pileup

- At the LHC, we do not (usually) see a single proton-proton collision
  - In Run 2, there are an average of roughly 34 collisions per bunch crossing
  - This produces a lot of “spurious” energy in the detector
  - This is referred to as pileup: a major experimental consideration/challenge





In-time vs out-of-time pileup

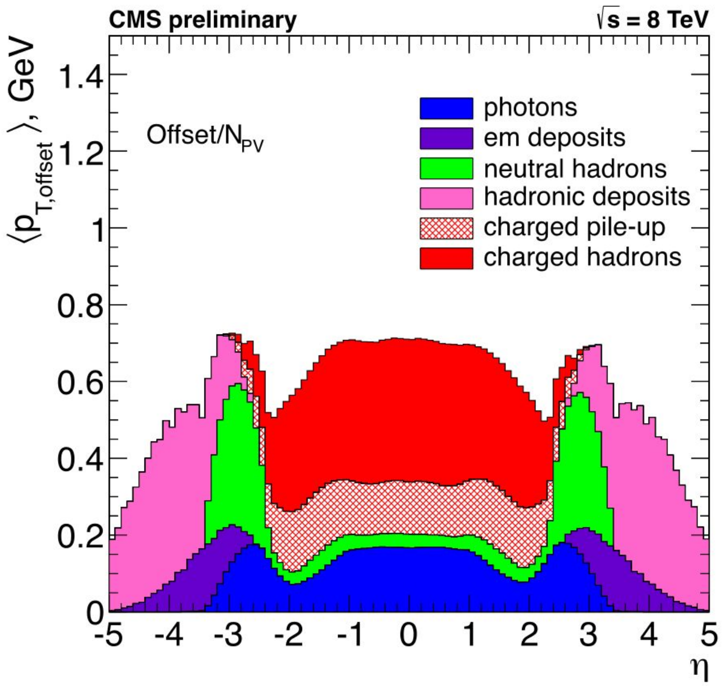
- The last slide was primarily “in-time pileup”
  - The impact of many proton-proton interactions in the same bunch crossing
- Depending on the detector, may need to consider “out-of-time” pileup
  - The LHC bunch spacing is currently 25 ns
  - What happens if your detector takes longer than 25 ns to read out?
    - Your interpretation of a given event may depend on the previous/next collision
    - Shaping functions are used to minimize such effects, but they are not perfect
  - Different detectors are more or less susceptible to such effects

Impacts of pileup

- This additional spurious energy impacts jets in many ways
- Additional jets in the event, referred to as pileup jets
  - Ideally, we should veto such jets and focus on the hard scatter event
- Additional energy in hard scatter jets
  - Bias the energy of the jet (extra energy, not only the hard scatter)
  - Increase the jet energy resolution (additional noise in the jet)
  - Distort jet substructure (extra energy with random angles within the jet)
- Many other impacts on other observables
  - Missing transverse momentum resolution, lepton isolation, etc

Mitigating pileup

- Tracks are excellent for pileup mitigation
  - Extrapolate to calorimeter with high precision
  - Uniquely identified vertex allows for removing energy from other collisions
  - Labelled “charged hadrons”
- However, there are limitations
  - Tracker does not cover full detector
  - Neutral in-time pileup in calorimeters does not have corresponding tracks
  - Charged or neutral out-of-time pileup in calorimeters has no corresponding tracks
  - Pileup may overlap with hard energy
- Additional means of pileup suppression are required to handle these cases

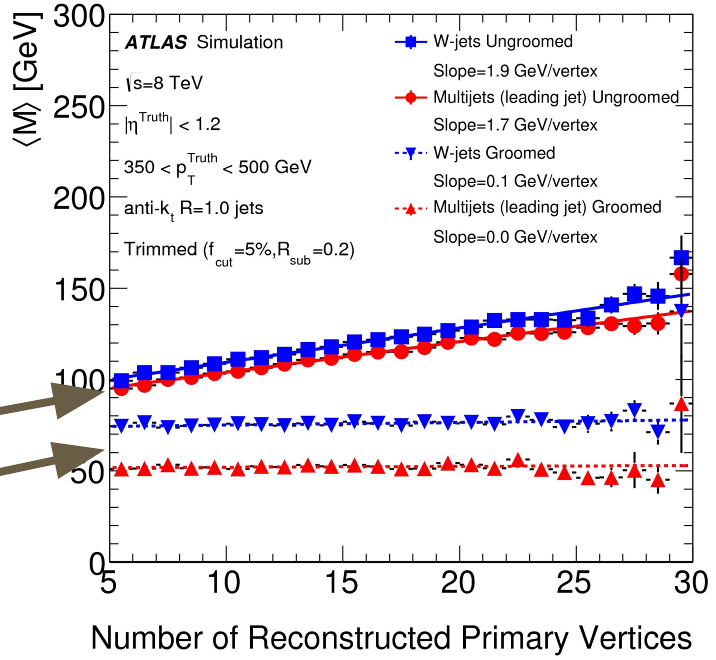


Pileup-suppressing jets

- There are different approaches to removing pileup beyond tracks
- Traditional: jet-areas based pileup suppression
  - Calculate the average pileup energy density event-by-event
  - Calculate the jet area jet-by-jet
  - Subtract (energy density)x(area) from each jet to correct the energy scale
- More advanced methods are becoming common
  - Defining new jet inputs: particle flow, etc
  - Pileup-suppressing jet inputs: constituent subtraction, SoftKiller, etc
  - Jet-level: jet-areas based correction as explained above
  - Jet modification: grooming is often used for pileup suppression, next slide

Jet substructure, grooming, and pileup

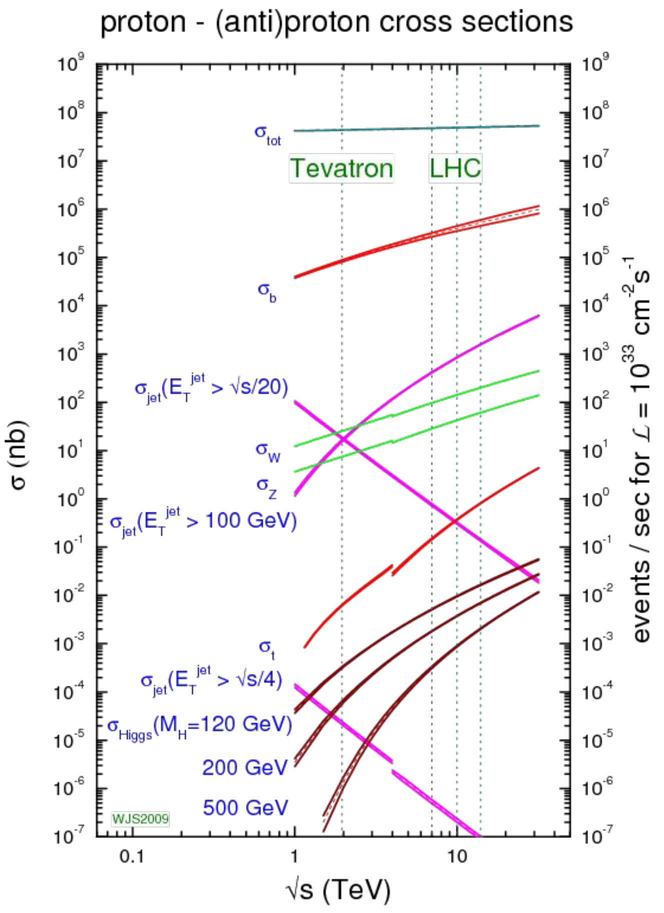
- Substructure vars are pileup-sensitive
  - Quantify the energy distribution within the jet
  - Pileup overlapping with a hard scatter jet = randomly distributed energy deposits
- Jet grooming originally envisioned to suppress the underlying event
  - However, performs very well against pileup!
- The jet mass is very pileup-unstable
  - After grooming, it becomes pileup robust
  - Also additional QCD vs signal separation
- Similar story for other substructure vars
  - Grooming is important to jet substructure



How can we “tag” (identify) a jet?

Jet identification

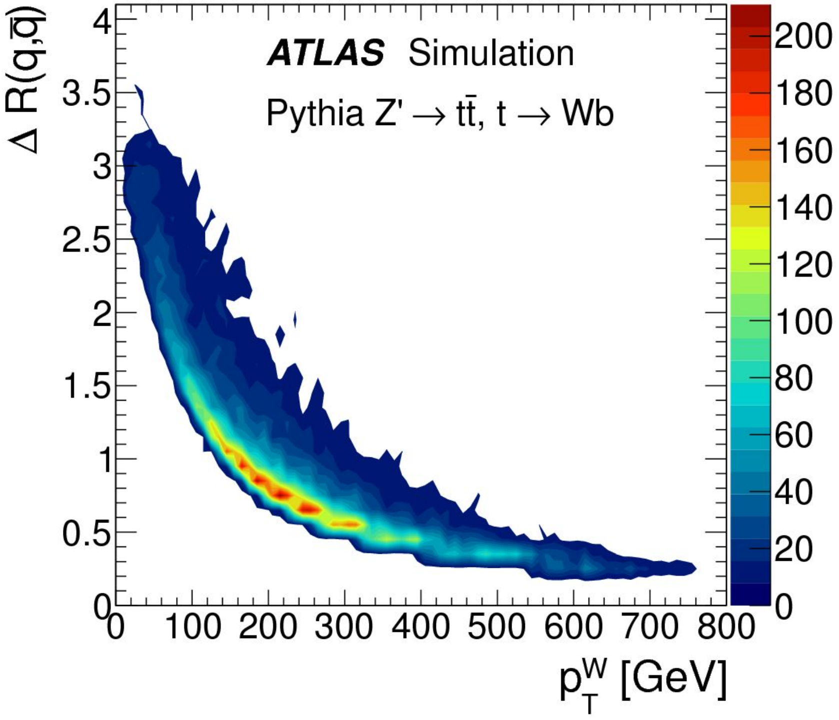
- Common types of jet identification:
  - Massive hadronic particle decays (W/Z/H/top) vs QCD
  - Light-quark vs gluon discrimination
  - Pileup vs hard-scatter jet identification
  - Flavour tagging (b-jets, c-jets, etc)
  - Other types exist, these are just the most common
- We will focus on hadronic decay tagging
  - Fighting the enormous QCD background
  - Need many orders of magnitude QCD rejection



arxiv:1204.0952

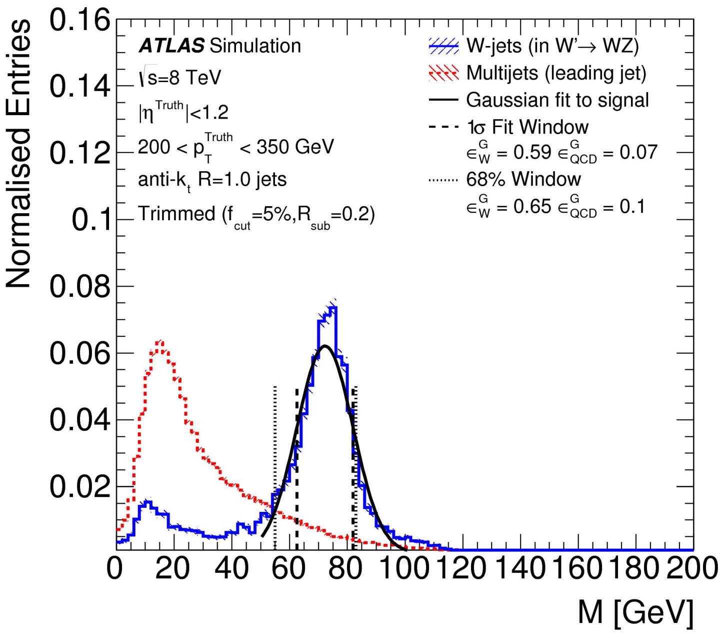
Hadronic decays and collimation

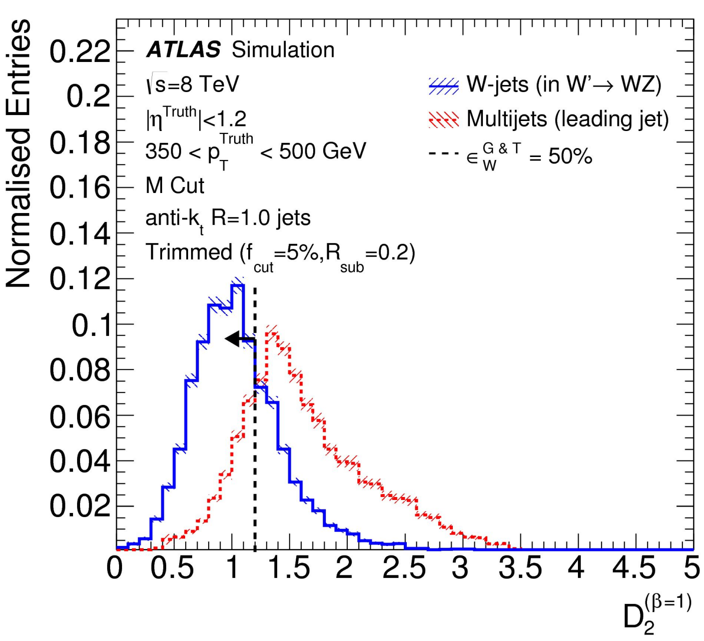
- Hadronic decays back-to-back at rest
  - Boost to experiment frame → collimation
- Two body decay, a → b c
  - $\Delta R_{bc} \approx m_a / \mathbf{x_b}(1-\mathbf{x_b})$
  - $\mathbf{x_b}$ is the fraction of a's momentum attributed to decay product b
    - On average, equal split between b and c, so $\mathbf{x_b} \approx 0.5\, p_{T,a}$
  - Result: $\Delta R_{bc} \approx 2\, m_a / p_{T,a}$
  - Dependence visible in W → qq' decays
- This is for two body decays
  - More complex for many-body decays
  - However, same trend: high $p_{T,a}$ → collimated



Simple cut-based taggers

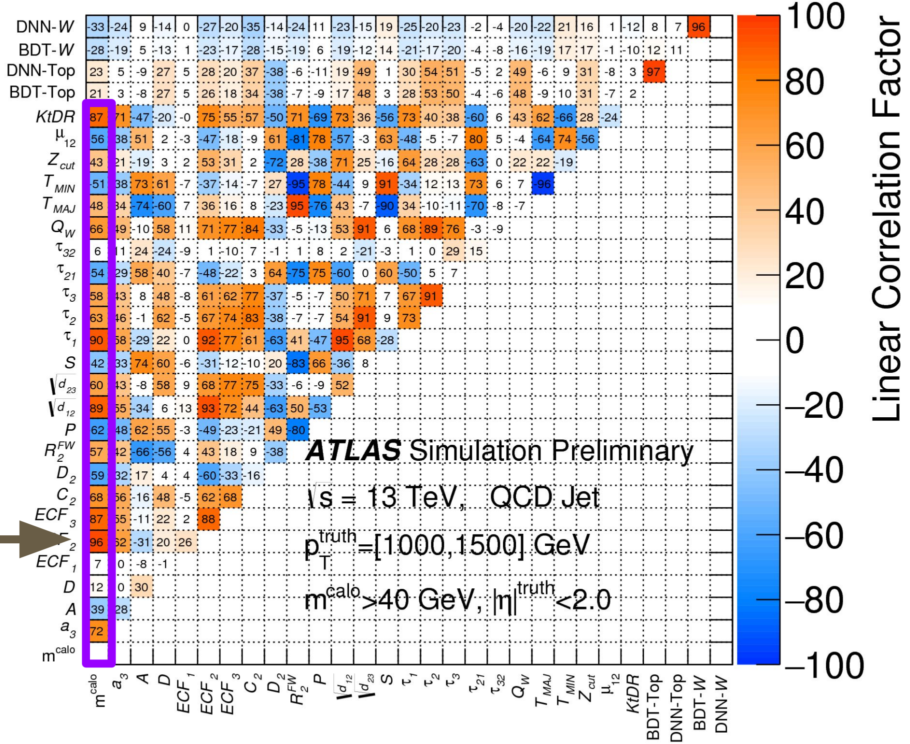
- Cut-based selections are already powerful (e.g. mass+D2 for W/Z-tagging)
  - ~1.5% QCD efficiency for 50% signal efficiency





Substructure variable correlations

- Two-variables are great, why not keep adding more information?
  - There is only so much information to exploit → diminishing returns
- Substructure variables are mostly correlated with each other
  - In particular, most variables are correlated with jet mass
  - Very few light-coloured entries



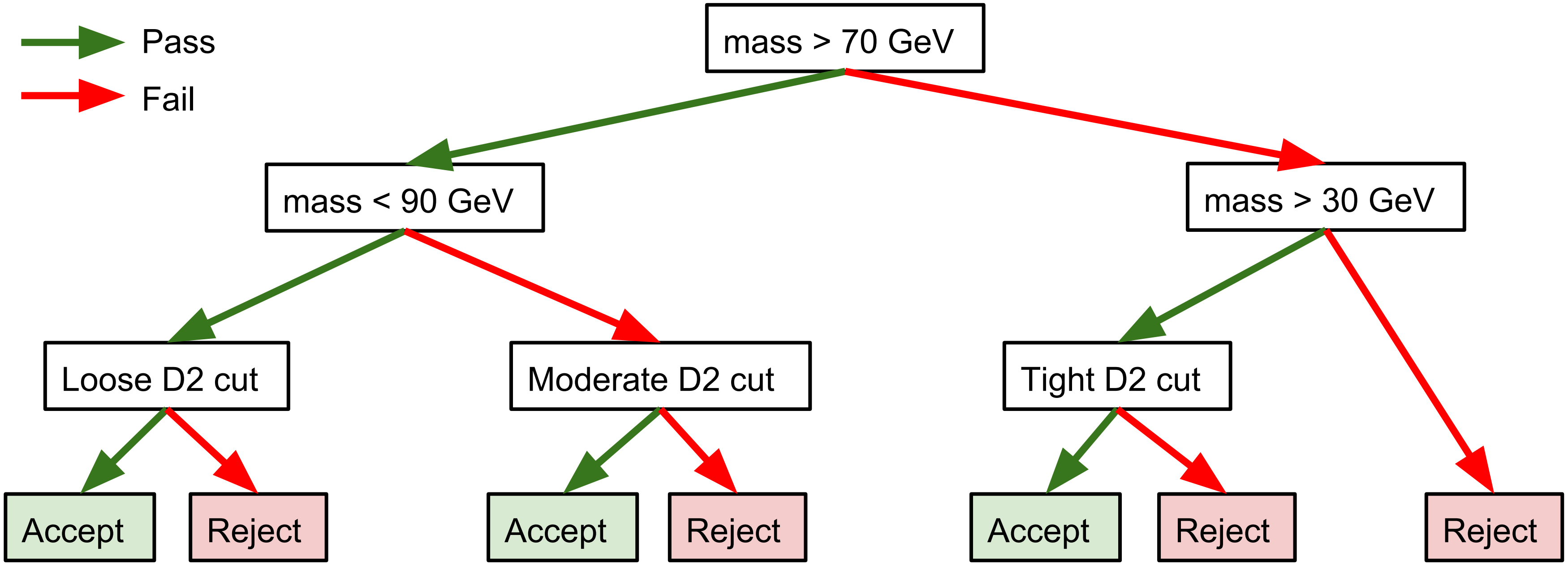
Machine Learning (ML)

- Powerful technique to extract information from a dataset
  - Can leverage non-linear correlations, not just linear correlations (last slide)
- ML has both positives and negatives
  - Usually better jet tagging power than non-ML approaches
    - Gains for using ML can be quite substantial, depending on the task and inputs
  - However, less understanding of what is being done
    - Traditionally a “black-box” approach
    - Need to be careful to ensure the ML is learning real features, not MC artifacts
- There are many, many types of ML techniques
  - We will focus on the two most commonly used techniques in modern HEP

Boosted Decision Trees (BDTs)

- BDTs are excellent for combining variables in taggers
  - Usually fast to train and nearly optimal performance
  - Also easier to “understand” what they are doing
- Idea: build a “decision tree”, or a set of sequential cuts on variables
  - One-dimensional decision tree is just a single variable cut
  - Output is typically a binary decision: “signal” or “background”
- This decision tree will have some rate of misclassifying events
  - Build another decision tree which focuses on these misclassified events
  - This is the “boosting” step in BDT
- Repeat this many times to train your BDT for optimal performance

BDT W-tagging example, inputs = (mass, D2)

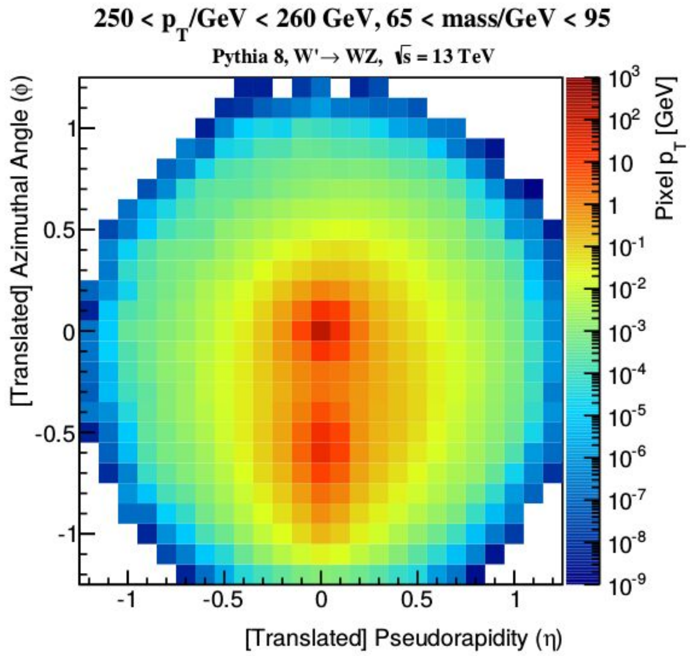


Neural Networks (NNs)

- Neural networks were originally inspired by biology and brains
- Comprised of inputs, hidden layer(s), and output(s)
    - Each variable is typically connected with all others in the previous and next layer(s)
        - Weights between nodes control the flow of information
    - If there is more than one hidden layer, it is a “deep network” → “deep learning”
    - One output is a binary discriminant, many outputs is a multi-class discriminant
- There are now many types of neural networks
    - NNs, Deep (DNNs), Convolutional (CNNs), Recurrent+Recursive (RNNs), and more
    - All except for simple NNs are typically “deep”
- Very powerful, but very complex, and difficult to interpret
    - An *infinitely deep* network can represent *any* non-linear function

Examples of modern usage of neural networks

CNN: Jet image of a W boson

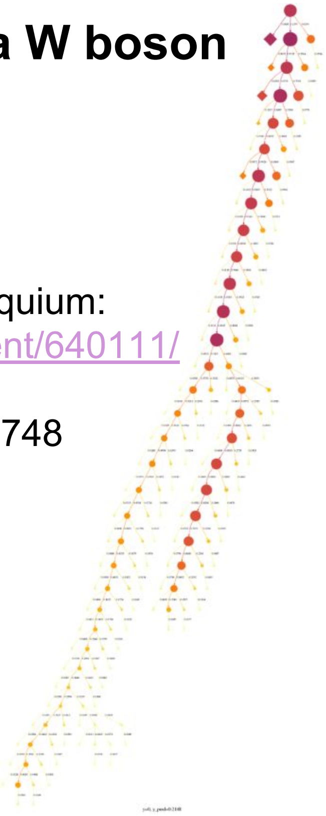


arXiv:1511.05190

RNN: tree view of a W boson

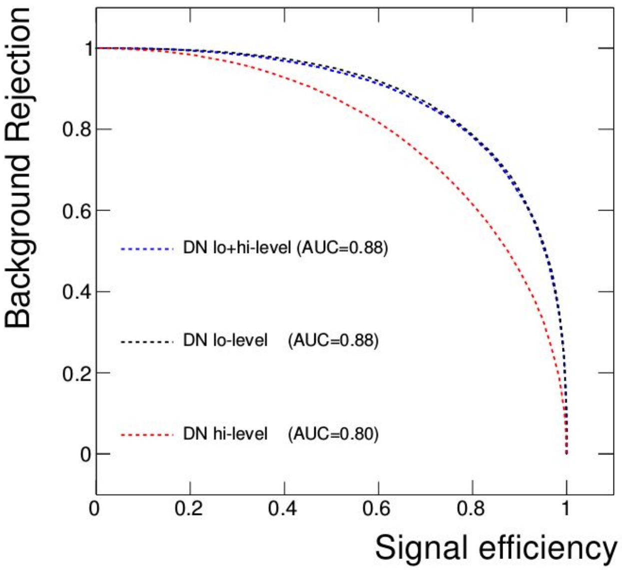
From CERN theory colloquium:
https://indico.cern.ch/event/640111/

Related to arXiv:1702.00748



Machine Learning taggers

- How well do these ML taggers actually work?
  - Short answer: quite well, especially for complex objects
- Performance depends on the inputs used for ML
  - “High-level” variables are properties of the jet itself
    - Jet four-vector, substructure variables, etc
  - “Low-level” variables are properties of inputs to jets
    - Constituent four-vectors, constituent properties, etc
- Typically, BDT/DNN perform similarly on high-level inputs
  - However, DNNs typically benefit further when using low-level inputs
- We will see more examples in the next couple of days



arxiv:1402.4735

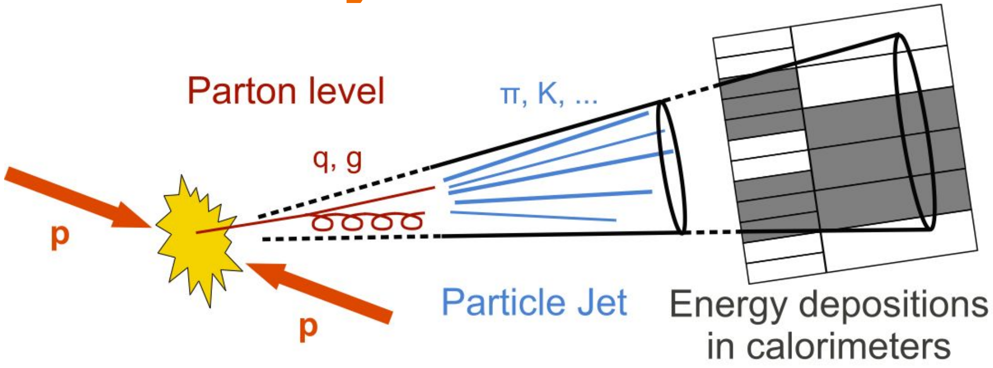
Summary

Summary

- Jet substructure is a rapidly growing and increasingly important field
  - Used throughout the experimental physics programs
  - New searches and measurements of jet substructure come out every year
- Machine learning has a growing role in jet tagging
  - Some modern examples were shown, you will see many more in the talks
- This introduction should have provided you with the key pieces that you need to follow the talks in the next couple of days
  - Don’t be shy! Ask questions and participate in discussions

BACKUP

What goes into the jet?



- Parton level : before hadronization
  - Generally limited to fixed-order calculations
- Particle level : after hadronization
  - Good proxy for the “truth” definition of a jet (particles generate detector signal)
- Calorimeter / track / particle flow objects
  - Detector-specific views of an event

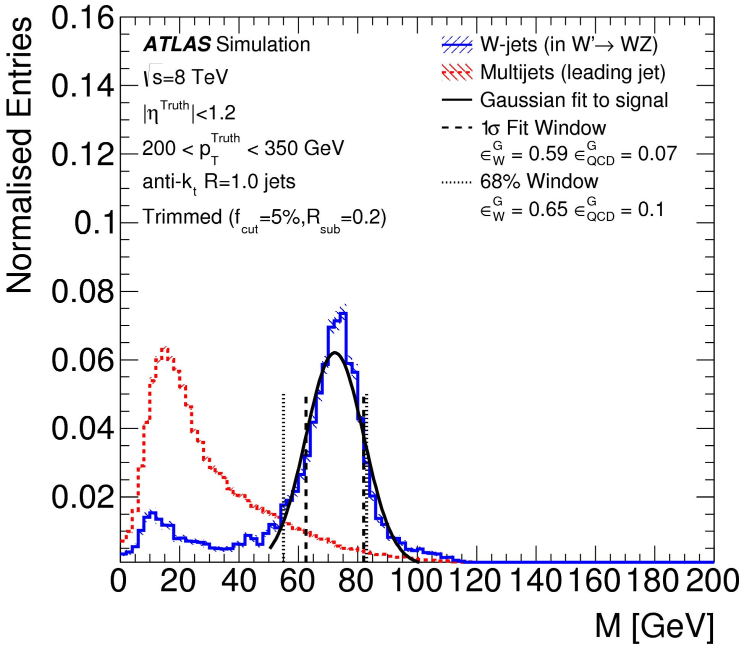
Theory / Pheno

Theory / Pheno / Experiment

Experiment

Jet mass

- The most intuitive substructure variable: four-vector invariant mass
- Well-defined expectation for decays
  - Mass of a hadronic decay should correspond to the parent particle (such as the W mass)
  - Assumes full decay contained in the jet
- Ill-defined expectation for QCD jets
  - “Mass” of light quarks and gluons depends on detector granularity, algorithms, etc
  - After grooming, QCD mass peak at low values
- Excellent means of discriminating QCD from hadronic resonance decays



Analytical approach to jet substructure

Why do we need it?

- We can acquire **insight from analytical expressions**
  - Better understand a phenomenon
  - Develop better tools (e.g. boson and top taggers)
- Obtain **more precise results**
  - Parton Shower only provide the lowest logarithm accuracy
  - Resummation can achieve higher accuracies
  - Results are systematically improvable
- Compute robust **uncertainty bands**
  - Correct assessment of the higher orders corrections we are neglecting

Resummation techniques

- Suppose one wants to compute a QCD observable
- We first try a fixed-order expansion in the strong coupling α_s

$$\langle O \rangle = \sum_n \alpha_s^n c_n$$

Fixed-order $\longrightarrow$ truncating this series at a given n

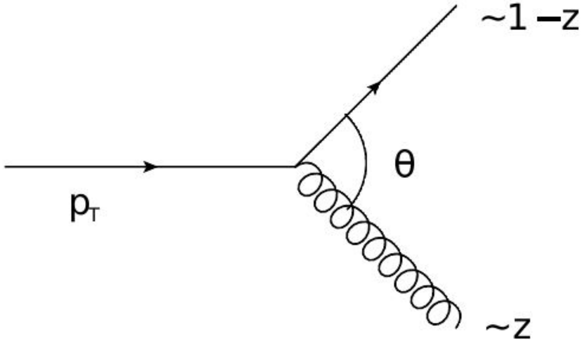
- Problem: when c_n is not “well behaved” $\longrightarrow$ **FO expansion does not converge**
- Cases with a strong hierarchy between scales e.g. boosted regimes $p_t \gg m$
- **Need an all-order (in α_s) resummed calculation**

Resummation techniques

- Resummation rarely describe all phase space
  - needs to be **matched to a fixed-order (FO) calculation**
  - one needs to avoid double-counting of the terms
  - this is a non-trivial process
- Limitation of the analytical approach is the poor understanding of non-perturbative effects (underlying event and hadronization)
- There are different techniques used to so resummation (“perturbative” QCD or SCET -- Soft Collinear Effective Theory), **they are equivalent**

Simple example : jet mass

- Integrated distribution for the jet mass m
- Simplest case: jet with only one emission q → q + g



$$m^2 \simeq p_T^2 z(1-z)\theta^2$$

- Boosted jets → $\rho = m^2/(p_t R)^2 \ll 1$

Simple example : jet mass

- For boosted jets :

$$\Sigma(\rho) \simeq \frac{\alpha_s C_F}{\pi}\left[\frac{1}{2}\boxed{\log\left(\frac{1}{\rho}\right)^2} - \frac{3}{4}\log\left(\frac{1}{\rho}\right) + \mathcal{O}(1)\right]$$

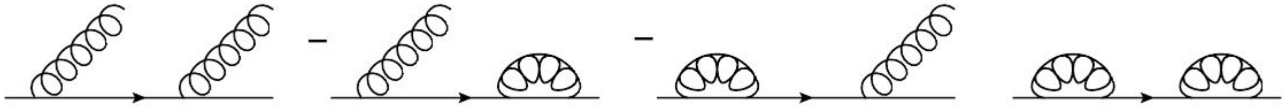
- For higher orders in α_s → terms like $[\alpha_s \log(1/\rho)^2]^n$
- If $\rho \ll 1$ → $\alpha_s \log(1/\rho)^2 \sim 1$

→ **fixed-order** expansion **does not converge**

- **Need resummation at all orders**

Simple example : jet mass

- Only interested at the dominating term ~ $\alpha_s \log(1/\rho)^2 \rightarrow$ **Leading Logarithm (LL)**
- Virtual emissions → cancel out soft and collinear divergences.



$$\Sigma(\rho) = 1 - \frac{\alpha_s C_F}{\pi} \exp \underbrace{\left[-\frac{\alpha_s C_F}{2\pi} \log \left(\frac{1}{\rho} \right)^2 \right]}_{\textbf{Sudakov exponent}}$$

New angles on energy correlation functions

Moult, Necib,Thaler (2016) -- Boost 2016

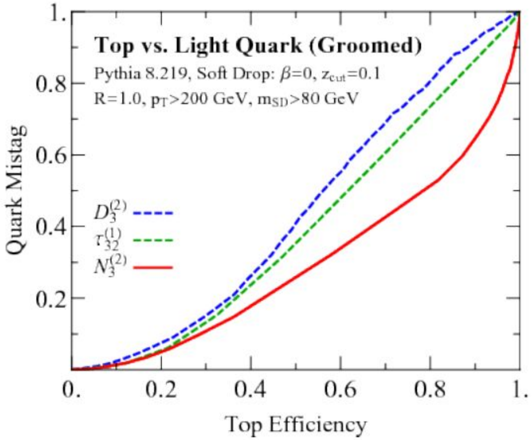
- Generalization of the energy correlation functions

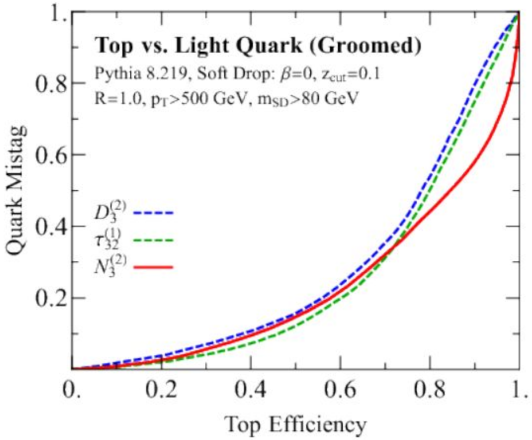
$$_v e_n^{(\beta)} = \sum_{1 \le i_1 < \cdots < i_n \le n_J} z_{i_1} \ldots z_{i_n} \prod_{m=1}^{v} \overset{(m)}{\min_{s<t \in \{i_1 \ldots i_n\}}} \{\theta_{st}^{\beta}\}$$

- Defined the series:
  - M_i : identify jets with hard prongs $\quad M_i^{(\beta)} = \frac{_1e_{i+1}^{(\beta)}}{_1e_i^{(\beta)}}$
  - N_i : mimics the behavior of N-subjettiness $\quad N_i^{(\beta)} = \frac{_2e_{i+1}^{(\beta)}}{(_1e_i^{(\beta)})^2}$
  - U_i : probe multiple emissions within 1-pronged jets $\quad U_i^{(\beta)} = {_1e_{i+1}^{(\beta)}}$

Example : top tagging with N_3

- Observables proposed: $N_3 = {}_2e_4^{(\beta)}/({}_1e^{(\beta)})^2 \sim \tau_3/\tau_2$
- Grooming matters! Affects efficiency of observable.





- These shapes perform slightly better than standard $D_2^{(2)}$

Dichroic Jet Shapes

Salam, LS, Soyez (2016) -- Boost2016

- Explore the interplay between groomers / prong finders and jet shapes;
- Example : N-subjetiness -- the usual τ_{21} measure is

$$\tau_{21} = \frac{\tau_2(\mathrm{mMDT})}{\tau_1(\mathrm{mMDT})} \quad \text{or} \quad \frac{\tau_2(\mathrm{SD})}{\tau_1(\mathrm{SD})} \quad \text{or} \quad \frac{\tau_2(\mathrm{plain})}{\tau_1(\mathrm{plain})}$$

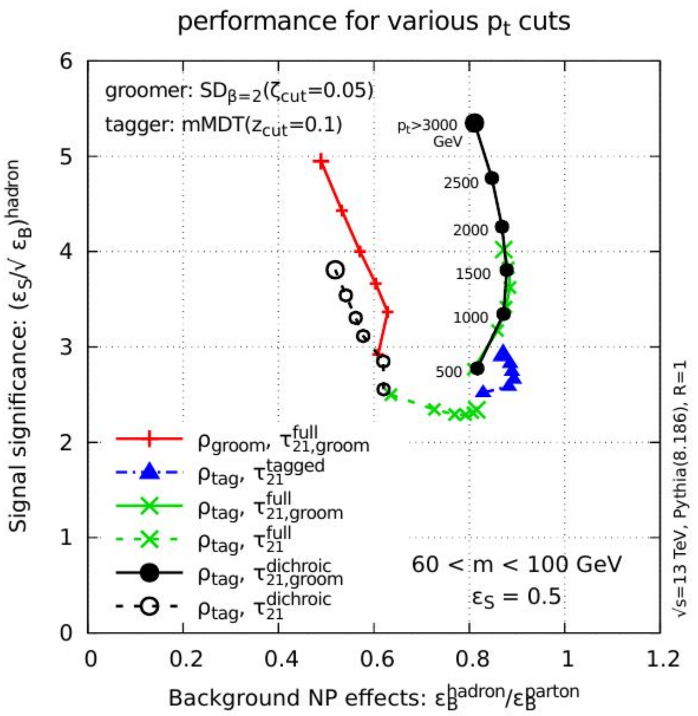
- Dichroic : different subjets for numerator / denominator in τ 21 ratios;

$$\tau_{21}^{\mathrm{dichroic}} \equiv \frac{\tau_2^{\mathrm{full}\ /\ \mathrm{SD}}}{\tau_1^{\mathrm{tagged}}}$$

- τ_2 on large jet → sensitivity to different color structures
- τ_1 on small jet → only sensitive to the invariant mass

  → smaller influence of non-perturbative effects.

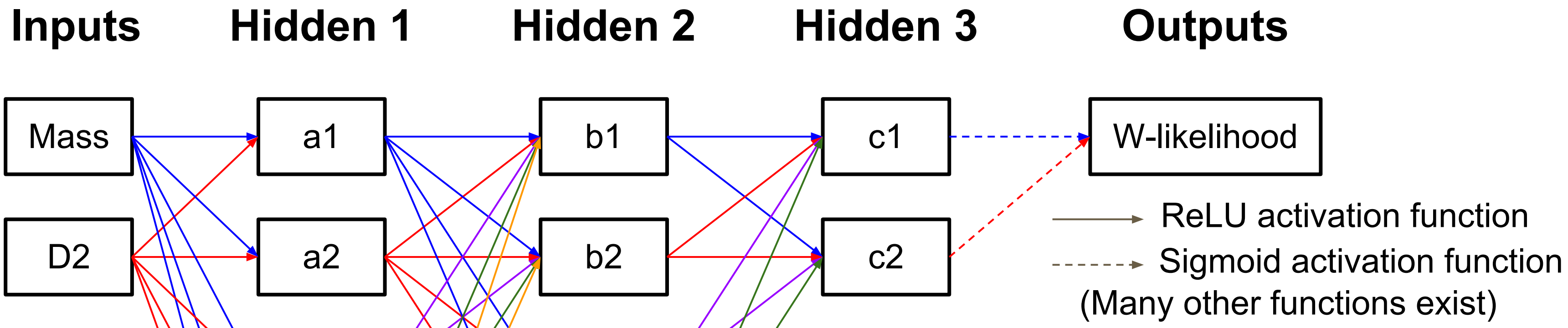
Dichroic Jet Shapes



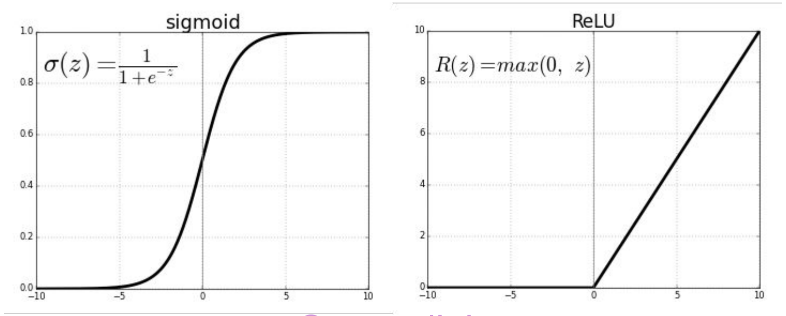
- Dichroic τ_{21} variation→ increases discriminating power
- With **pre-grooming step**→ **reduction of NP effects** and still has a **better performance** (black solid dots)
- Performance gain increases as p_t increases

More ML content

DNN W-tagging example, inputs = (mass, D2)



(Many other functions exist)



Source link

Tagging efficiency scale factors and uncertainties

- With ML taggers, it is very important to evaluate performance in data
  - ML may be learning features of MC which do not represent data
  - Propagating uncertainties on ML inputs is not sufficient
    - ML exploits correlations between variables, which are typically poorly known
- Derive tagging efficiency scale factors and uncertainties *in situ*
  - Use a “standard candle” were we have high purity to measure data/MC differences
  - Examples of standard candles on the next slide
- With such techniques, ML becomes a well-defined and useful tool
  - However, we still don’t necessarily know *what it has learned*
  - This remains an open question of great interest for the future

Tagging efficiency standard candles

- W peak in semi-leptonic ttbar selection (left) and W/Z peak in fully hadronic V+jets selection (right) are two examples of standard candles
  - Fit the peak and extract the data/MC difference for V-tagging efficiencies