Jet substructure for everyone

Gregory Soyez

IPhT, CNRS, CEA Saclay

Jets and their substructure from LHC data, May 31-June 4 2021, CERN (online)

The growing scope of jet substructure

In the early 2000-2010's: Jet substructure was a niche

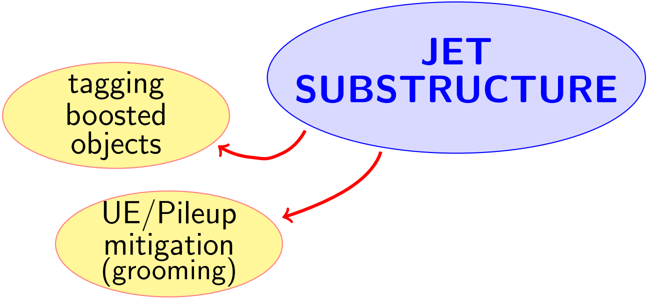
(I would have had to explain what it is and why it is interesting)

The growing scope of jet substructure

In the early 2000-2010's: Jet substructure was a niche

(I would have had to explain what it is and why it is interesting)

Over ∼10 years it has become a standard tool at colliders

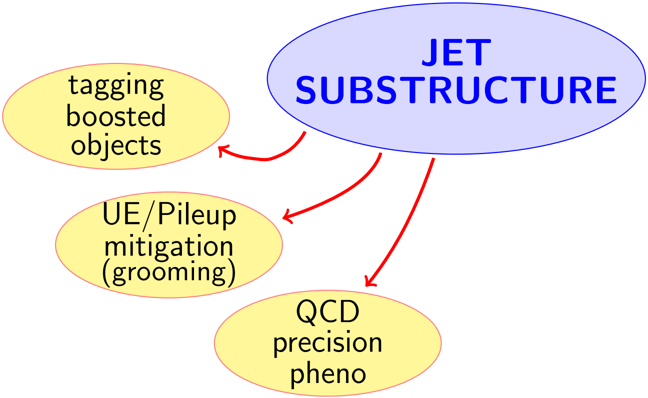


The growing scope of jet substructure

In the early 2000-2010's: Jet substructure was a niche

(I would have had to explain what it is and why it is interesting)

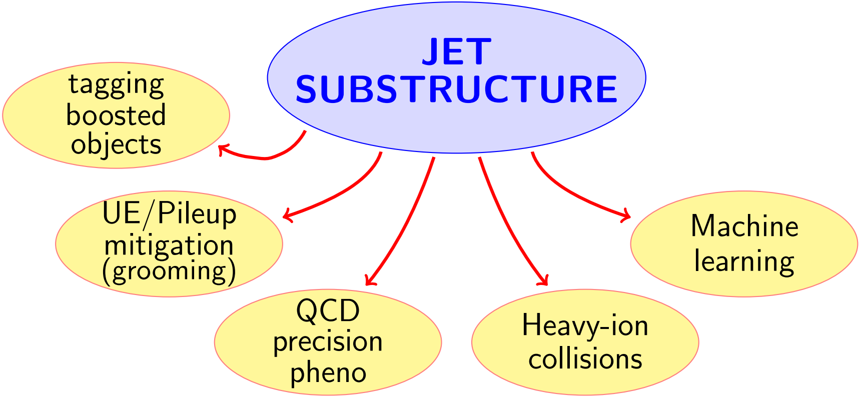
Over $\sim$10 years it has become a standard tool at colliders



In the early 2000-2010's: Jet substructure was a niche

(I would have had to explain what it is and why it is interesting)

Over $\sim$10 years it has become a standard tool at colliders

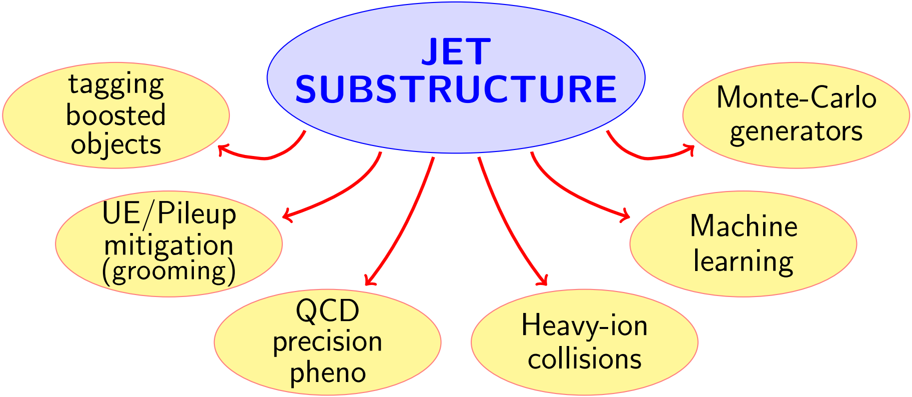


The growing scope of jet substructure

In the early 2000-2010's: Jet substructure was a niche

(I would have had to explain what it is and why it is interesting)

Over $\sim$10 years it has become a standard tool at colliders

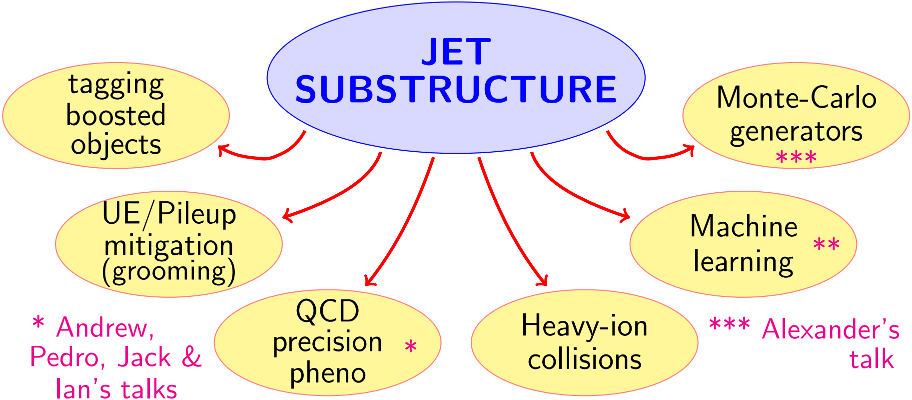


The growing scope of jet substructure

In the early 2000-2010's: Jet substructure was a niche

(I would have had to explain what it is and why it is interesting)

Over ∼10 years it has become a standard tool at colliders



* Andrew, Pedro, Jack & Ian's talks

*** Alexander's talk

** Jesse & Gregor's talk

+ Andreas/Ben/Ezra's talk

substructure $\equiv$ internal dynamics of jets

Before getting into applications, a few words about HOW it is done

(excluding ML)

Two main classes of approaches for jet clustering:

1. Jets as particle branchings $\rightarrow$ pairwise clustering ((anti-)k_t, C/A, ...)
2. Jets as energy flows $\rightarrow$ cone algorithms (SISCone, MidPoint, ...)

Two main classes of approaches for jet clustering:

1. Jets as particle branchings $\rightarrow$ pairwise clustering ((anti-)k_t, C/A, ...)
2. Jets as energy flows $\rightarrow$ cone algorithms (SISCone, MidPoint, ...)

The same broad classes apply to substructure:

1. Use particle branchings $\rightarrow$ techniques based on de-clustering the jet
2. Use energy flows $\rightarrow$ techniques based on (sub)jet shapes

A decade of substructure tools

(modified) MassDrop Tagger

(generalised) angularities

(recursive) SoftDrop

N-subjettiness

Trimming

Energy Correlation Functions

Pruning

Shower deconstruct^{n}

Energy flow Polynomials

JH Top tagger

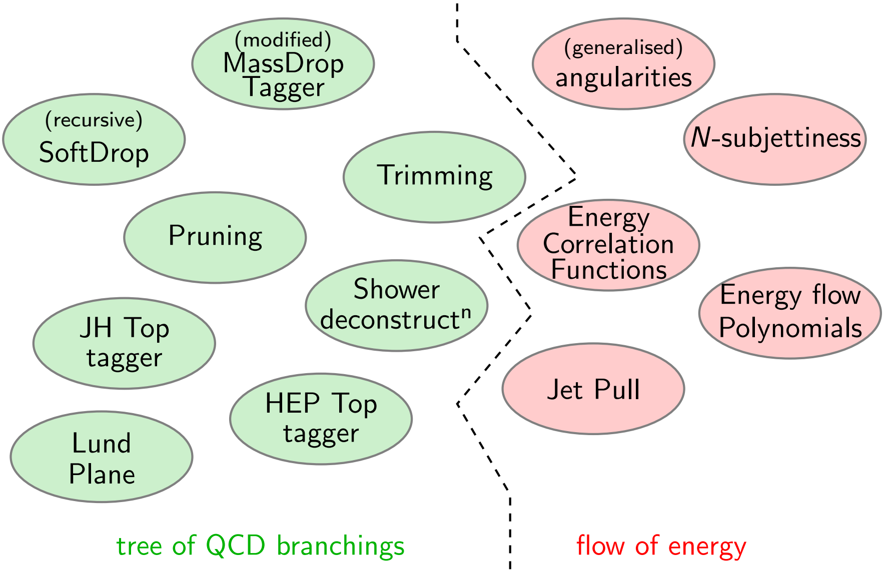
Jet Pull

HEP Top tagger

Lund Plane

* Non-exhaustive/biased/… list

A decade of substructure tools



A decade of substructure tools

(modified) MassDrop Tagger

(recursive) SoftDrop

Trimming

Pruning

Shower deconstruct^{n}

JH Top tagger

HEP Top tagger

Lund Plane

(generalised) angularities

N-subjettiness

Energy Correlation Functions

Energy flow Polynomials

Jet Pull

Take these two which offer a complete view of the jet

tree of QCD branchings

flow of energy

Energy flows

[P.komiske,E.Metodiev,J.Thaler,1712.07124]

Energy-flow polynomials

$$\mathrm{EFP}_G(\text{jet } j) = \sum_{i_1,\dots,i_N \in j} z_{i_1} \dots z_{i_N} \sum_{(k,l)\in G} \theta_{i_k i_l}$$

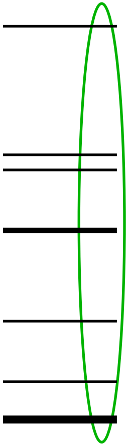
with G a (multi)graph with N vertices and some edges (k,l)

Main properties:

- linear basis for IRC-safe substructure observables
- includes energy-correlation function: $\sum_{i,j} z_i z_j \theta_{ij}^{\beta}$ (and higher-orders) widely used for tagging
- Can be used for various ML tagging applications
- Interesting underlying computational structures

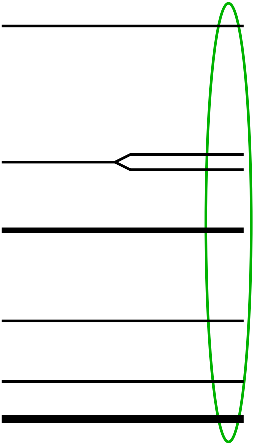
Clustering approach: Cambridge/Aachen

Cambridge/Aachen: iteratively recombine the closest pair



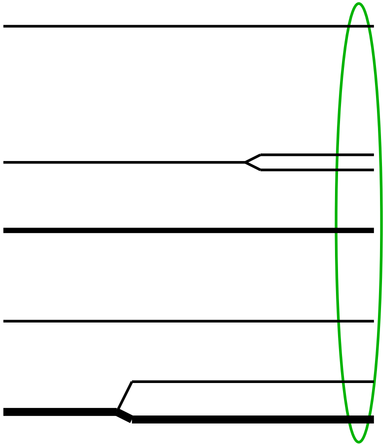
Clustering approach: Cambridge/Aachen

Cambridge/Aachen: iteratively recombine the closest pair



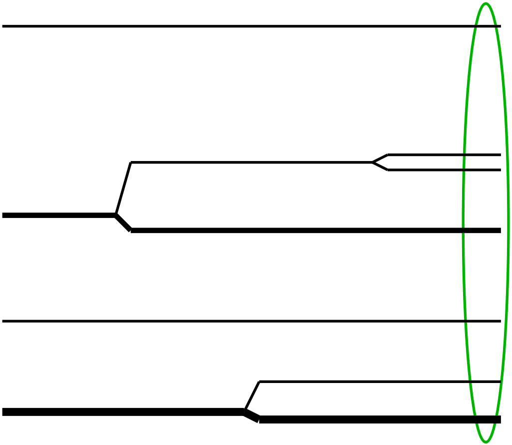
Clustering approach: Cambridge/Aachen

Cambridge/Aachen: iteratively recombine the closest pair

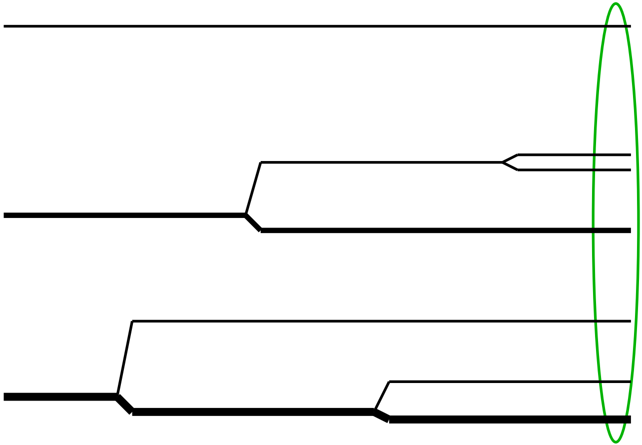


Clustering approach: Cambridge/Aachen

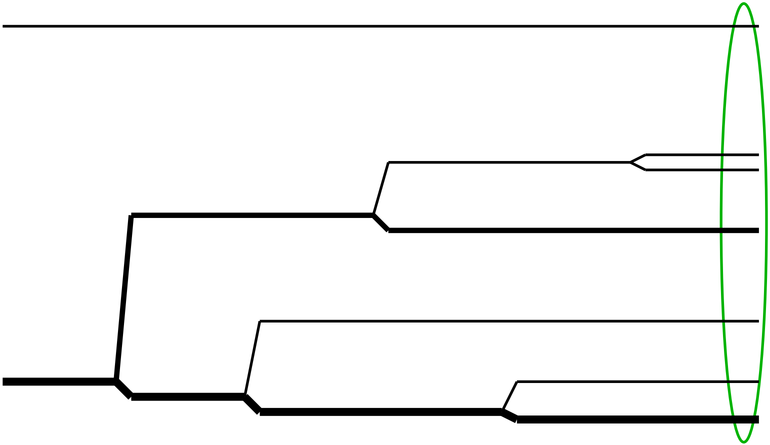
Cambridge/Aachen: iteratively recombine the closest pair



Cambridge/Aachen: iteratively recombine the closest pair

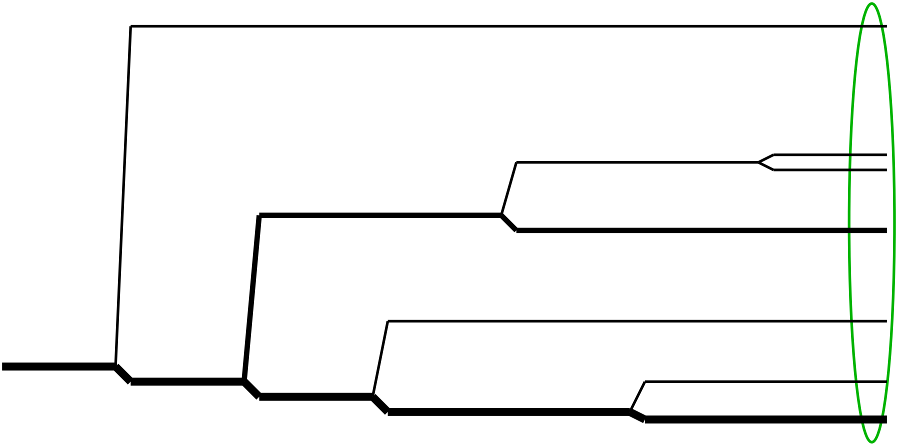


Cambridge/Aachen: iteratively recombine the closest pair



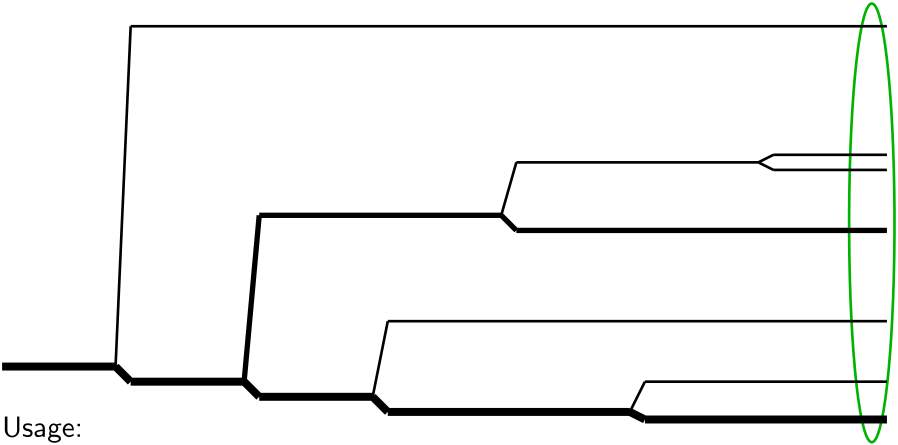
Clustering approach: Cambridge/Aachen

Cambridge/Aachen: iteratively recombine the closest pair



Clustering approach: Cambridge/Aachen

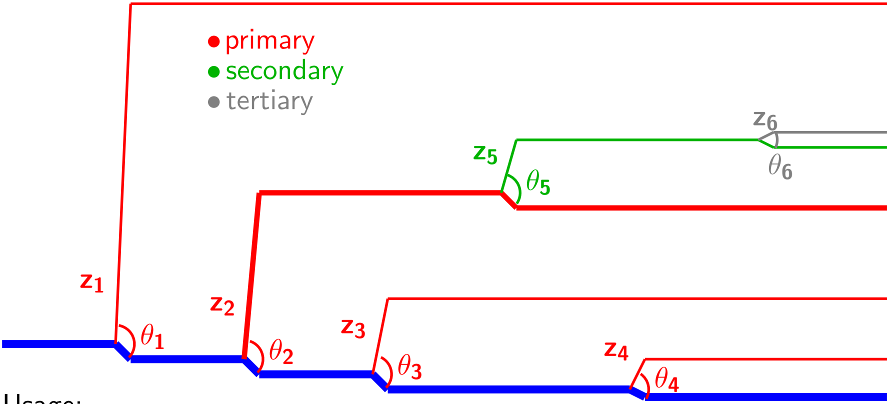
Cambridge/Aachen: iteratively recombine the closest pair



Usage:

- (iteratively) undo clusterings (following hard branch) to find structure (SoftDrop,...)
- study kinematic properties of the branchings (LundPlane)

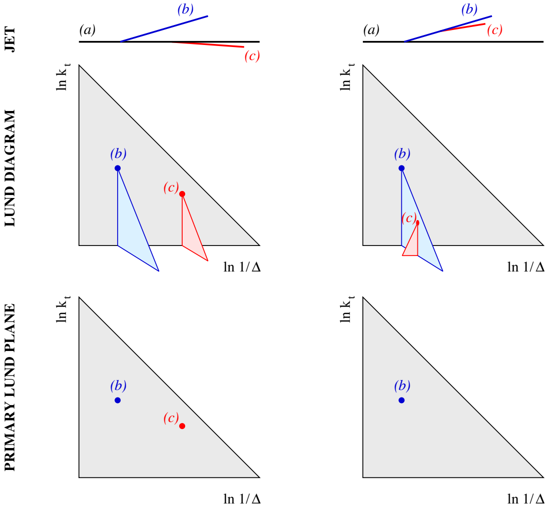
Cambridge/Aachen: iteratively recombine the closest pair



Usage:

- (iteratively) undo clusterings (following hard branch) to find structure (SoftDrop,...)
- study kinematic properties of the branchings (LundPlane)

[F.Dreyer,G.P.Salam,GS,1807.04758]



- C/A mimics angular ordering
- structure close to the Lund diagrams used in resummation/MC

Properties at each vertex

$$\mathcal{T}_i = \{\theta_i \equiv \Delta_i, k_{t,i}, z_i, \varphi_i, \dots\}$$

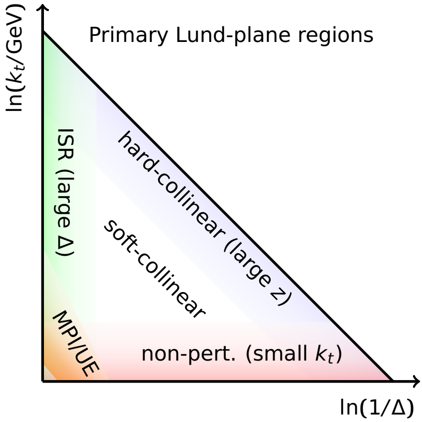
Either primary only:

$$\mathcal{L}_{\text{prim}}(j) = [\mathcal{T}_1, \dots, \mathcal{T}_n]_{\text{primary}}$$

Or full tree

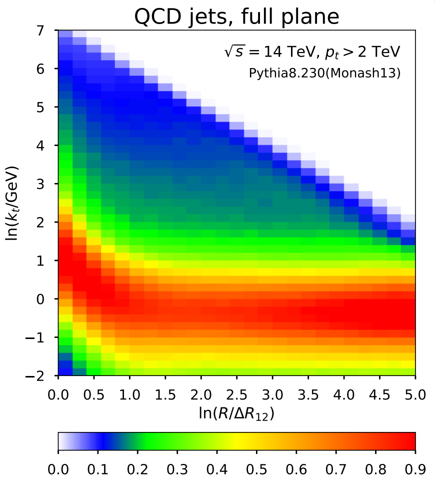
$$\mathcal{L}_{\text{tree}}(j) = [\mathcal{T}_{j\to j_1,j_2}, \mathcal{L}_{\text{tree}}(j_1), \mathcal{L}_{\text{tree}}(j_2)]$$

Lund plane physics regions



Different physics contributions in different regions of the (primary) Lund plane

In particular, selecting $k_t \geq k_{t,\mathrm{cut}}$ selects a pQCD region (IRC safety)



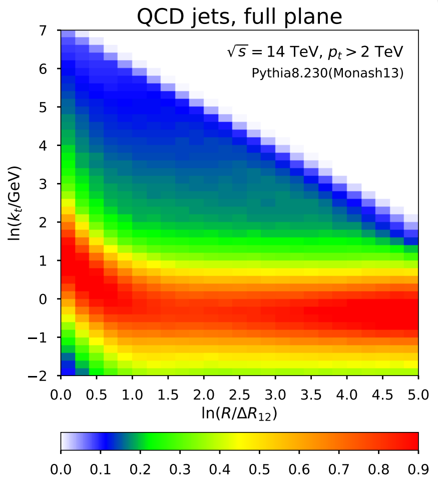
Different physics contributions in different regions of the (primary) Lund plane

In particular, selecting $k_t \geq k_{t,\text{cut}}$ selects a pQCD region (IRC safety)

Simple observable

Lund plane density:

$$\rho = \frac{1}{N_{\text{jet}}} \frac{d^2 N}{d \ln 1/\Delta \, d \ln k_t}$$



Different physics contributions in different regions of the (primary) Lund plane

In particular, selecting $k_t \geq k_{t,\mathrm{cut}}$ selects a pQCD region (IRC safety)

Simple observable

Lund plane density:

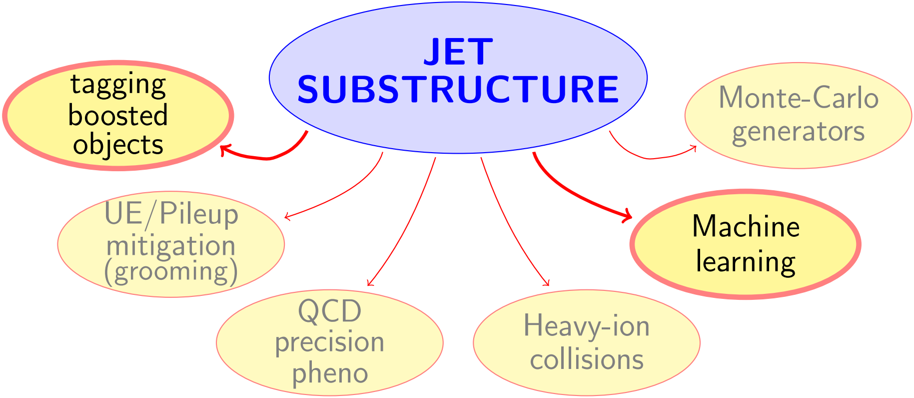
$$\rho = \frac{1}{N_{\mathrm{jet}}} \frac{d^2 N}{d \ln 1/\Delta \, d \ln k_t}$$

Prospect: trees & E flows are not exclusive: we can define flow/shape observables on tree constructions

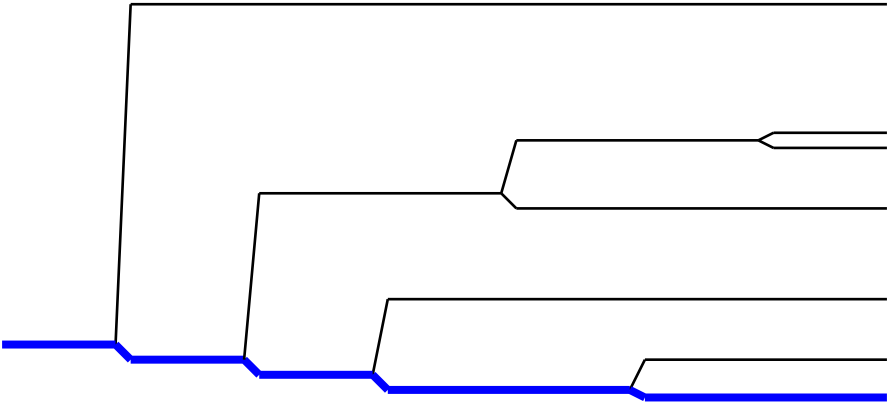
New and exciting progress

In a nutshell: for a few big applications of substructure I will

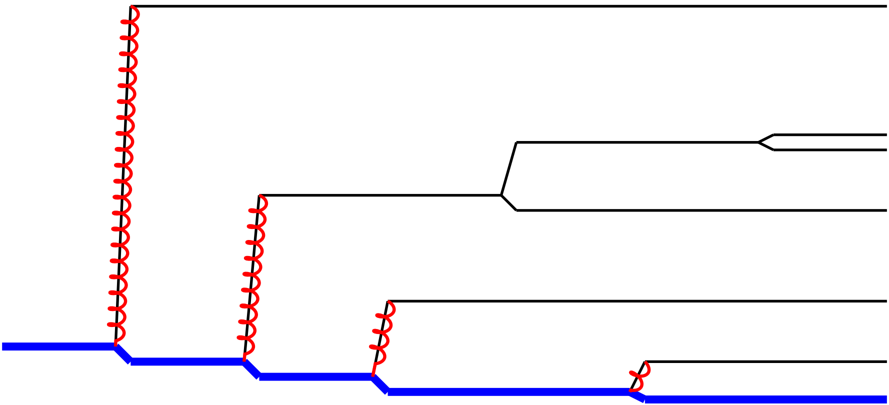
1. Describe the physical idea (phrasing it in the Lund Plane picture)
2. Give (at least one) interesting recent development
3. Discuss potential new developments


JET SUBSTRUCTURE
tagging boosted objects
Monte-Carlo generators
UE/Pileup mitigation (grooming)
Machine learning
QCD precision pheno
Heavy-ion collisions

Physics ideas

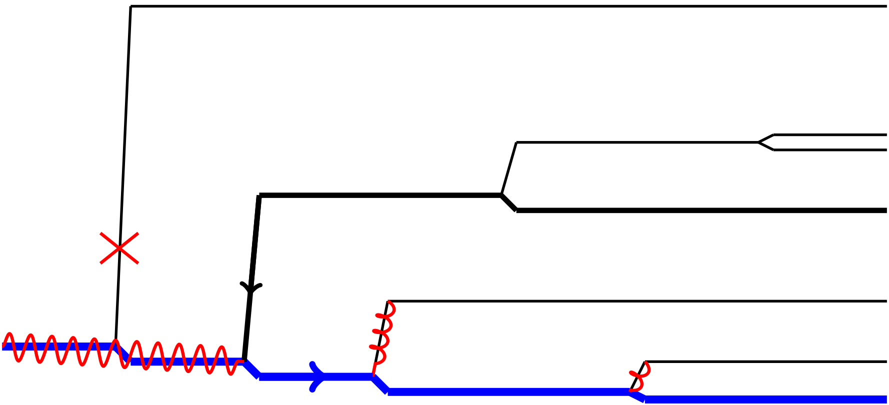


Physics ideas



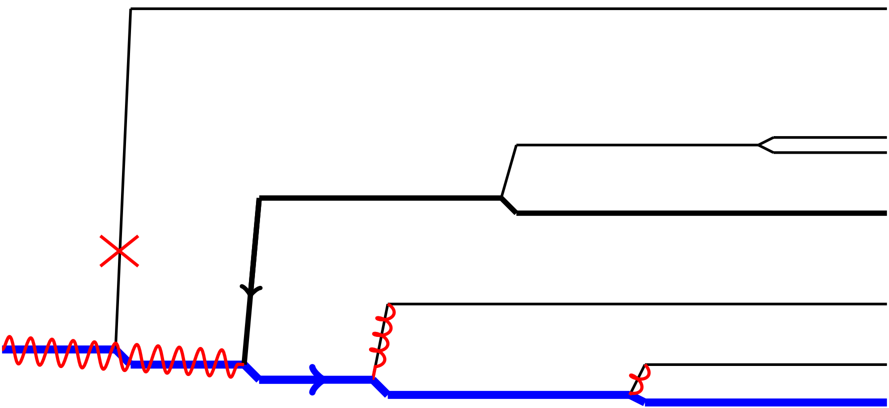
QCD jet: mostly soft gluon emissions

Physics ideas



QCD jet: mostly soft gluon emissions

W jet: hard $W \to q\bar{q}$ branching + no large-angle emissions



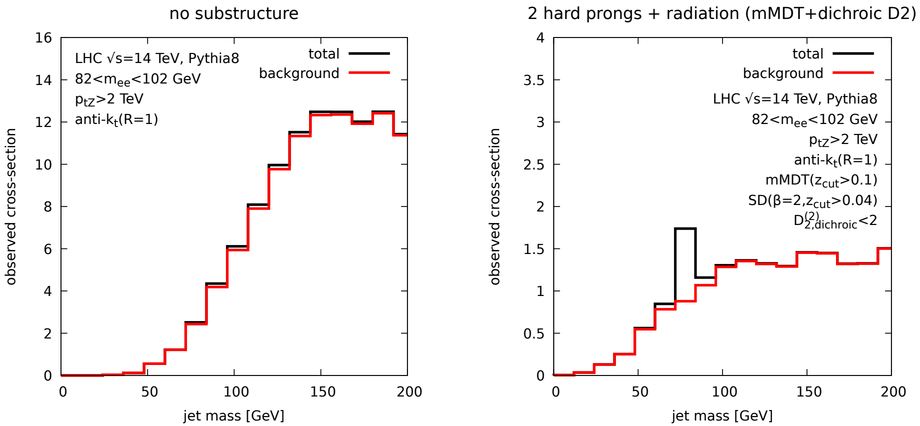
QCD jet: mostly soft gluon emissions

W jet: hard $W \to q\bar{q}$ branching + no large-angle emissions

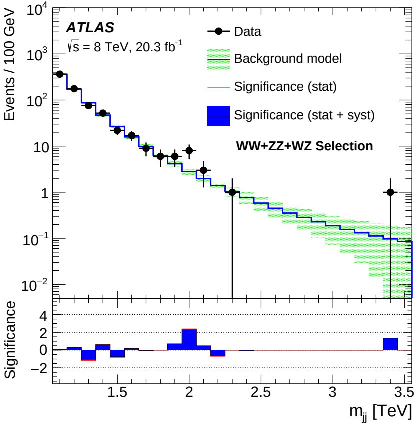
typically tree   typically shape

Tagging a boosted W boson (using $Z(\to \mu\mu) + W(\text{jet})$ vs. $Z(\to \mu\mu) + \text{jet}$)

no substructure — with substructure



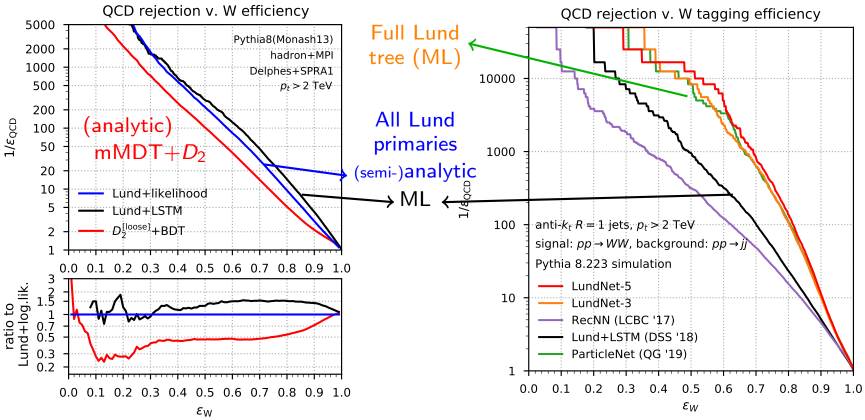
mMDT: 1307.0007; D_2: 1305.0007; dichroic: 1612.03917



(now-gone) di-boson "excess" (end of Run-I)

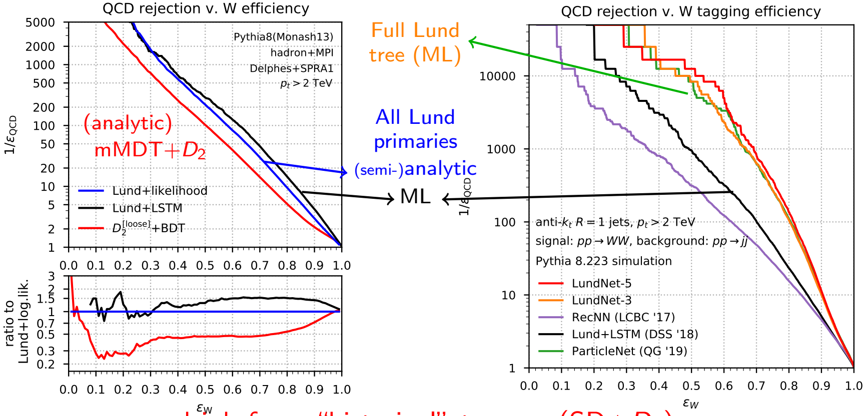
analytic pQCD and ML for W tagging

Maximise background rejection for given signal efficiency



analytic pQCD and ML for W tagging

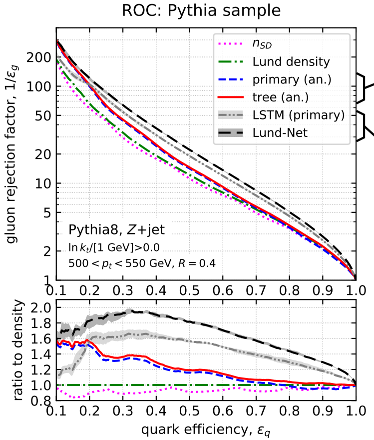
Maximise background rejection for given signal efficiency



Messages: - good job from "historical" taggers (SD+D_2)

- Including all (primary) declusterings help [F.Dreyer,G.Salam,GS,1807.04758]
- ML can extract some additional information (∼ ParticleNet w smaller cost [Gregor's])
- clear gain including the full tree [F.Dreyer,H.Qu,2012.08526]

[F.Dreyer,GS,A.Takacs,soon]



Use Lund declusterings
(primary or full tree)

Either compute analytically: $^{(*)}$

$$\mathbb{L} = \frac{p(q|\mathcal{L})}{p(g|\mathcal{L})}$$

Or use as ML input

Main messages:

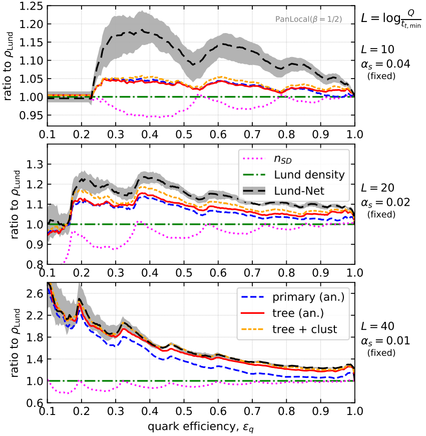
✓ Nice analytic performance

✓ Better gain for ML

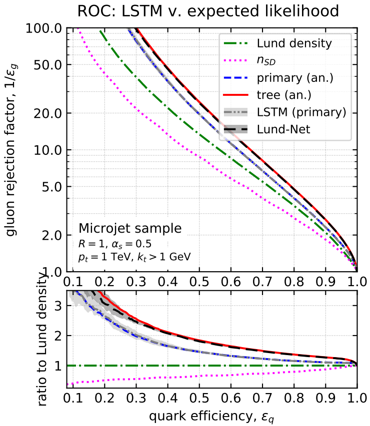
$^{(*)}$ $\sim$ NLL generalisation of $n_{\rm SD}$

If we increase the Boost (fixing $\alpha_s \log(p_t/k_{t,\min})$) the gap between analytic and ML closes

- Great validation
- Question: extra information used by ML?
- Question: MC accuracy? is that extra information well-described in MC?



Fun with taggers: validate ML with QCD

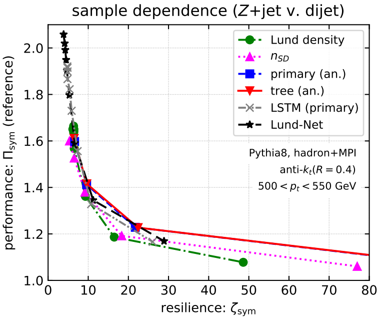


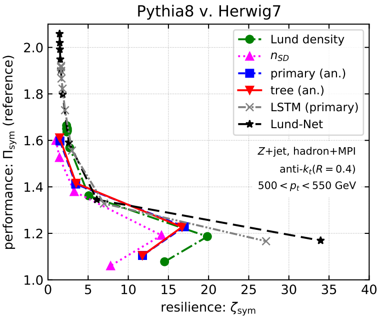
Use a sample with exact LL angular ordering (*) and compare analytics to ML

Shows very good agreement

(*) [F.Dreyer,M.Dasgupta,G.Salam,GS,1411.5182]

Performance v. resilience



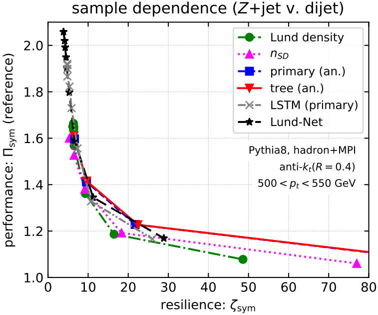


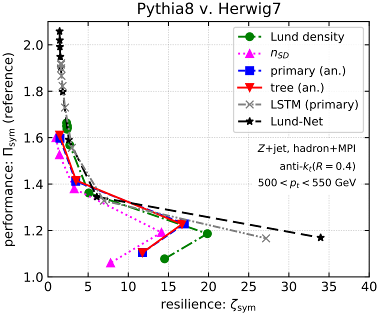
Performance v. resilience

Less sensitive to process (better)

Less sensitive to MC choice (better)

Better discrimination





- Generic trade-off between discriminating power and resilience
- ML gain mostly at small k_t where modelling more important
- some ML gain at larger k_t. Accessible in pQCD?

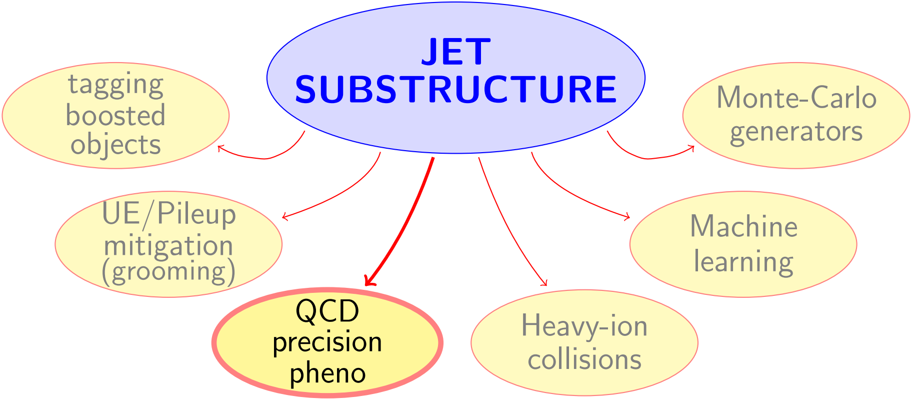
What to look for in the future

Tagging progress is (mostly) ML-oriented today

There are nonetheless some interesting directions where "QCD" can help

- Improved inputs/architecture
  Better control, faster convergence,...
- QCD analytics can be used as a validation of ML in known limits
- Trade-off between performance and "resilience"
- Room for "simple" QCD-based taggers (like the above Lund-plane, or arXiv:2006.10480)

More in Gregor&Jesse'talks (ML) and in Jack's talk (analytics)

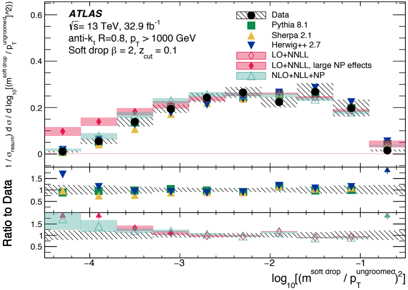

JET
SUBSTRUCTURE
tagging
boosted
objects
Monte-Carlo
generators
UE/Pileup
mitigation
(grooming)
Machine
learning
QCD
precision
pheno
Heavy-ion
collisions

Analytics for precision: basics

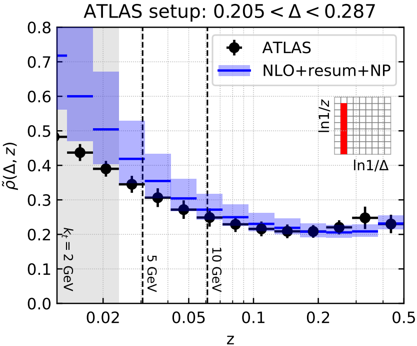
- What: compute substructure observables in pQCD
- Why: gain understanding on tools, compare to data, ...
- How: substructure often probes physics at scales $\ll$ jet p_t
  $\rightarrow$ need for resummation
  - from "standard" QCD and SCET
  - State-of-the-art: NLL (or NNLL) matched to NLO
  - intimate link with Parton Showers (more on this later)
- A nice benefit: "groom" soft & large-angle radiation
  - makes calculations somewhat easier (mostly collinear physics)
  - reduced non-perturbative effects, pushed them to lower scales

Analytics for precision

SoftDrop jet mass



Lund-plane density

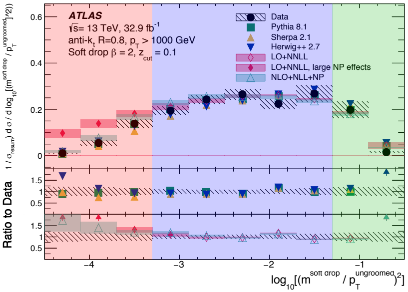


- Regions dominated by either NP, resummation or fixed-order
- Unique framework for new probes of QCD over several scales
- Question: does substructure help for α_s or PDF extraction?
- See e.g. 1711.08341, 1603.09338, 1704.02210, 1712.05105, 1807.05974, 2104.06920, CMS-PAS-SMP-20-010

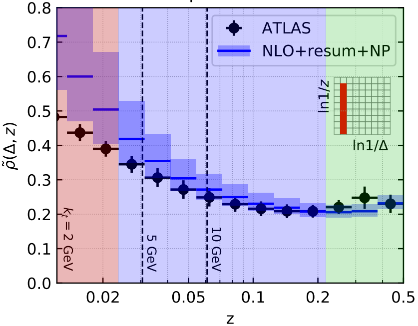
More in Andrew, Ian and Jack's talks

Analytics for precision

SoftDrop jet mass



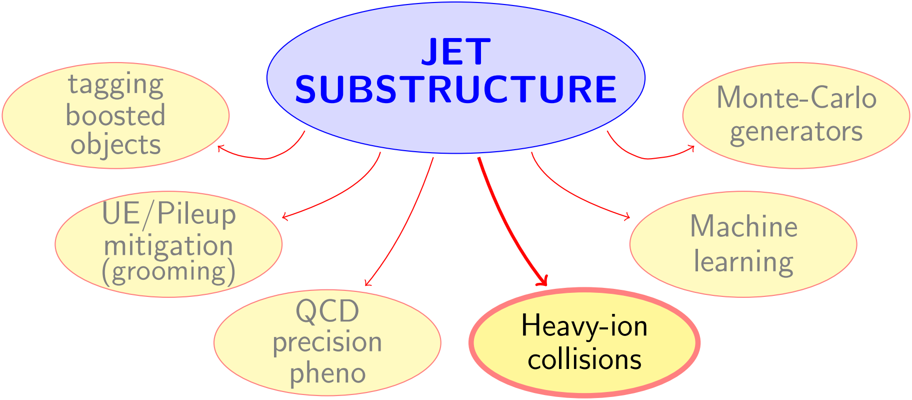
Lund-plane density



- Regions dominated by either NP, resummation or fixed-order
- Unique framework for new probes of QCD over several scales
- Question: does substructure help for α_s or PDF extraction?
- Prospect: study more complicated/differential quantities

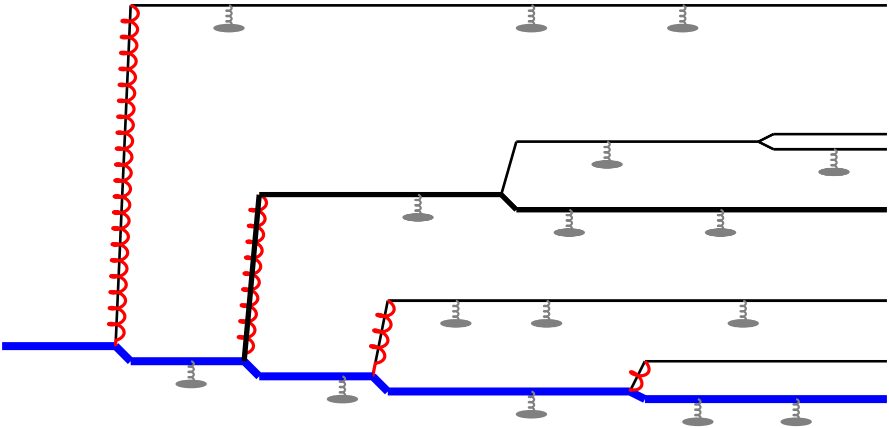
More in Andrew, Ian and Jack's talks

New prospects at the LHC



Jet quenching and substructure

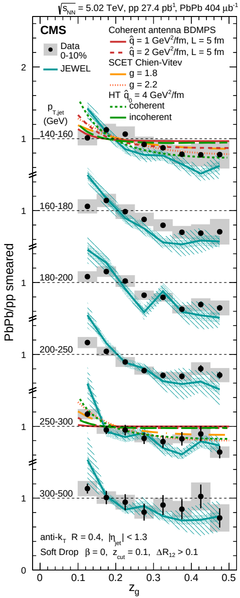
Idea: interaction with the quark-gluon plasma



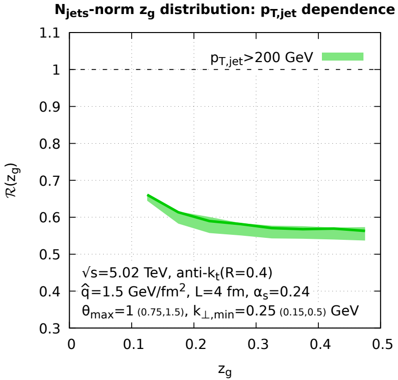
QGP affects the jet dynamics $\Rightarrow$ probe with substructure

Measuring the splitting function

CMS (CMS-HIN-16-006)



[P.Caucal,E.Iancu,GS,1801.09703,1907.0486617]



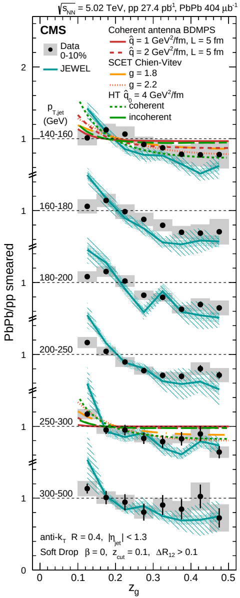
Based on perturbative QCD

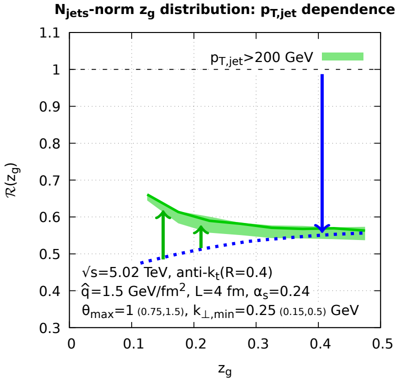
At LL: factorisation between "vacuum" (standard) shower and medium effects.

Measuring the splitting function

CMS (CMS-HIN-16-006)

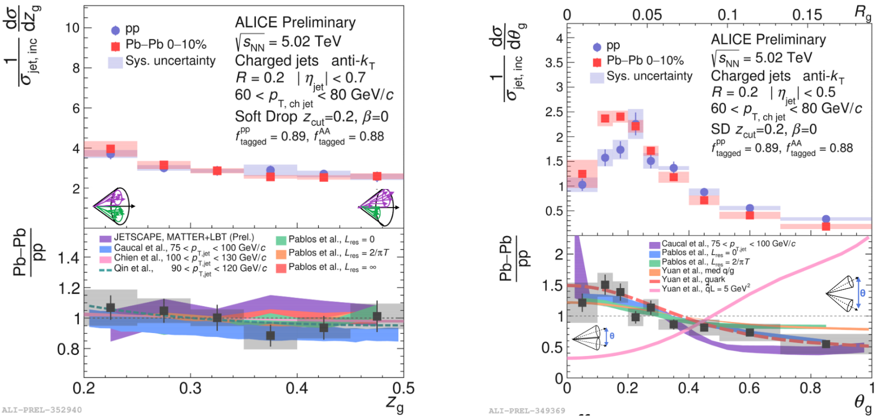
[P.Caucal,E.Iancu,GS,1801.09703,1907.0486617]





- Reduction from E loss
- Peak from extra emissions

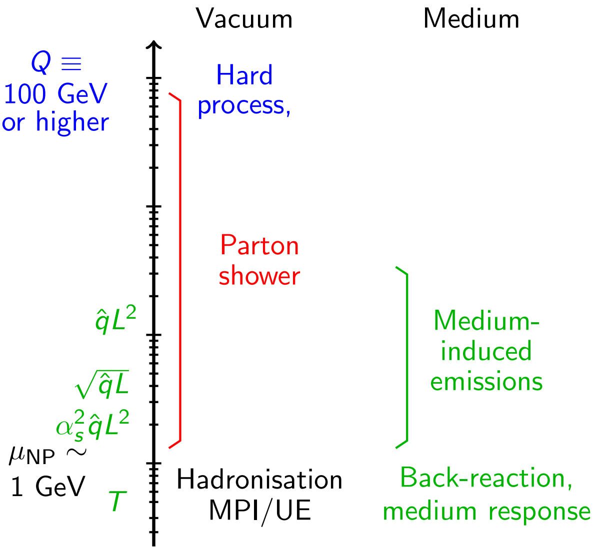
Recent measurement by the Alice collaboration



- Increasing number of substructure measurements at the LHC
- Comparisons to QCD calculations and MC simulations

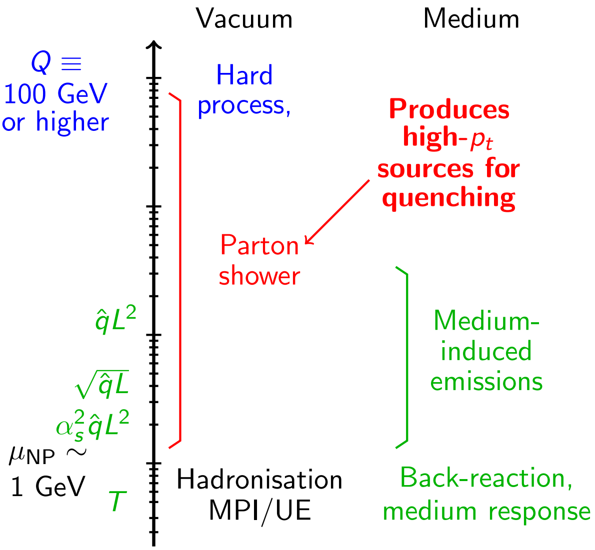
see also Pedro's talk

One aspect (I think) is key for the future



- The QGP brings new physics scales

One aspect (I think) is key for the future

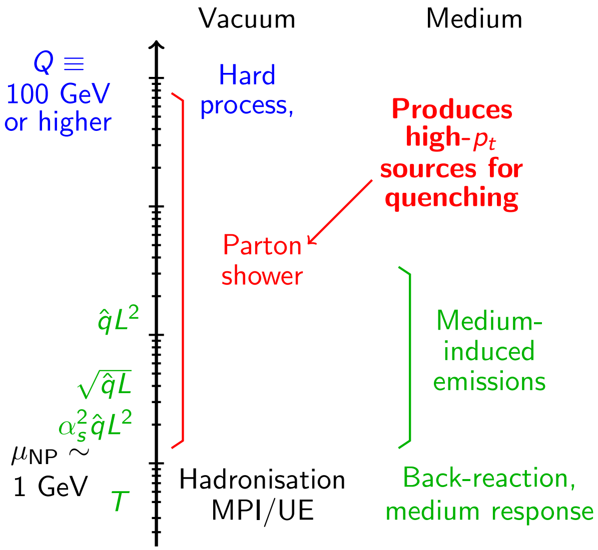


- The QGP brings new physics scales
- Large k_t dominated by vacuum physics
- Use substructure to "select" the structure of the jet propagating through the medium

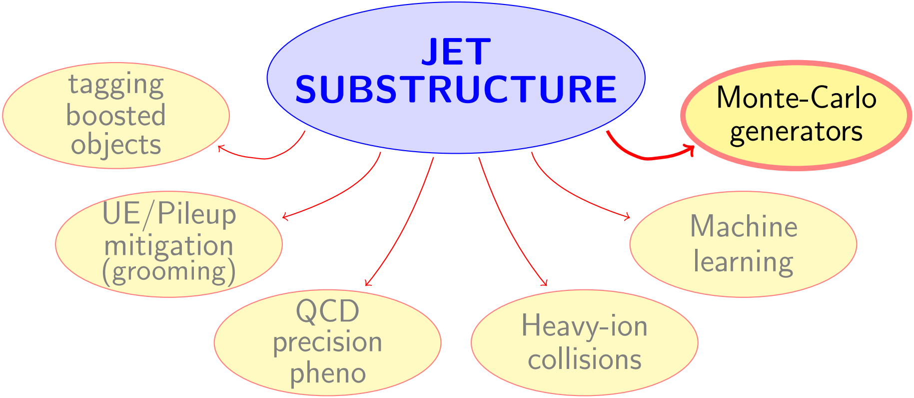
  $\Downarrow$

  fine-grained/differential study of the QGP.

One aspect (I think) is key for the future

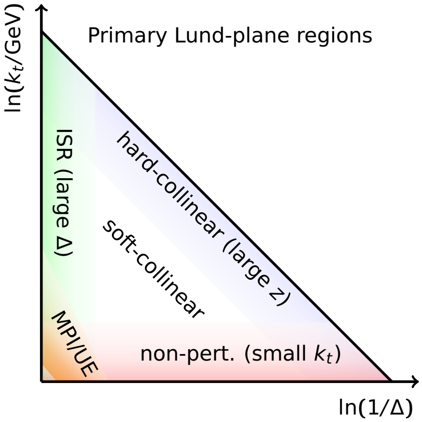


- The QGP brings new physics scales
- Large k_t dominated by vacuum physics
- Use substructure to "select" the structure of the jet propagating through the medium
  $\Downarrow$
  fine-grained/differential study of the QGP.
- Exp challenge: requires high-p_t


JET SUBSTRUCTURE
tagging boosted objects
UE/Pileup mitigation (grooming)
QCD precision pheno
Heavy-ion collisions
Machine learning
Monte-Carlo generators

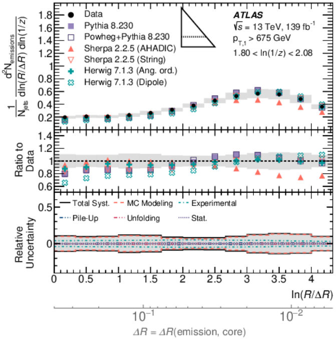
Substructure for MC development

Main idea: substructure observables probe QCD dynamics



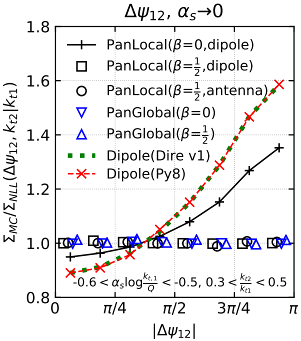
Main idea: substructure observables probe QCD dynamics

direct comparison between data and MC



[ATLAS,2004.03540]

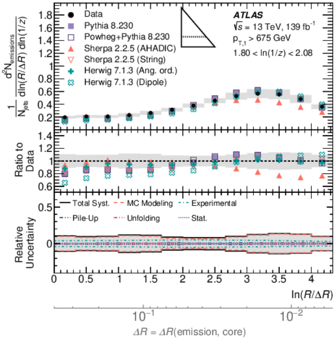
observables for accuracy tests/developments



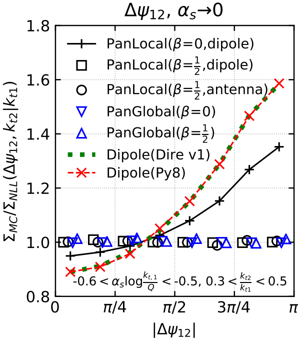
[M.Dasgupta,F.Dreyer,K.Hamilton, P.Monni,G.Salam,GS,2002.11114]

Main idea: substructure observables probe QCD dynamics

direct comparison between data and MC



observables for accuracy tests/developments



Better constraints ⇒ less modelling uncert. ⇒ improved searches

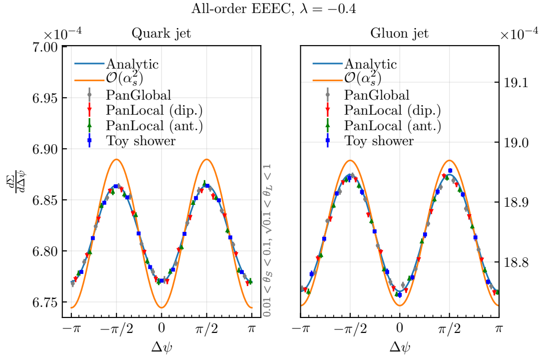
Versatile tool for future developments

Tweak the tool to match your needs

Connection with precision calculation

Example: Energy^3 Correlation functions sensitive to spin correlations at NLL

See Ian's talk



[A.Karlberg,G.Salam,L,Scyboz, R.Verheyen,2103.16526]

Conclusions & perspectives

Take-home messages

- **Substructure is now mainstream and is here to stay**
- **Wide range of applications (Taggers, pQCD, HI, MC, ML)**
- **Active exploration ground/laboratory for QCD, exploiting the large phase-space offered by the LHC**

Looking towards the future

- Expect more analyses with boosted jets
- Recent (and on-going) deep-learning revolution
- Need more calculations & (unfolded) substructure measurements
- Lots to do in substructure-based QGP studies
- Almost endless possibilities to test ideas/MCs/...

(Online) BOOST2021: https://indico.cern.ch/event/1037559/